

The Development and Evaluation of Hospital Pay-for-Performance in Lebanon

Casemix, Readmissions and Patient Perspectives

JADE KHALIFE

DEPARTMENT OF CLINICAL SCIENCES | FACULTY OF MEDICINE | LUND UNIVERSITY



The Development and Evaluation of Hospital Pay-for-Performance in Lebanon:
Casemix, Readmissions and Patient Perspectives

The Development and Evaluation of Hospital Pay-for-Performance in Lebanon

Casemix, Readmissions and Patient Perspectives

Jade Khalife



LUND
UNIVERSITY

Doctoral Dissertation

by due permission of the Faculty of Medicine, Lund University, Sweden.

To be publicly defended on 27th September at 13.00 in Agardh Hall,
Clinical Research Center, Malmö, Sweden.

Faculty opponent

Josephine Borghi, Professor,
Department of Global Health and Development,
London School of Hygiene & Tropical Medicine

Organization: Lund University

Document name: Doctoral Dissertation

Date of issue 2023-09-27

Author(s): Jade Khalife

Sponsoring organization:

Title and subtitle: The Development and Evaluation of Hospital Pay-for-Performance in Lebanon: Casemix, Readmissions and Patient Perspectives.

Abstract:

Background: Pay-for-performance (P4P) has been widely used in healthcare, but there are few experiences of hospital-based P4P at scale. The evidence of impact from these has been mixed, and there has been increased recognition of the importance of different contexts, designs, incentives and other factors. In 2014, the Lebanese Ministry of Public Health integrated a P4P model for determining hospital reimbursement tiers. In 2018, this model was updated to include a readmissions component, in addition to the preexisting components such as casemix and patient satisfaction. The impact of these interventions was previously undetermined. This also provided an opportunity to contribute to some of the known knowledge gaps regarding hospital P4P. The purpose of this thesis was to describe the development and evaluate the impact of hospital P4P in Lebanon, and ultimately to contribute more broadly to improved design and implementation of value-based healthcare, particularly in limited resource settings.

Methods: This thesis uses a mixed methods approach, combining quantitative and qualitative study designs, to conduct four research investigations. The first paper uses descriptive analysis to address how and why hospital P4P was developed in Lebanon. The second and third papers both use an interrupted time series design on data collected from the Ministry hospitalization database. The former uses Newey-OLS regression, and the latter uses Autoregressive Integrated Moving Average models. The second paper analyzes the impact of the 2014 P4P integration on casemix index, and the third paper analyzes the impact of the 2018 model update on readmissions. The fourth paper uses qualitative content analysis on data collected from eight focus groups discussions with patient participants.

Results: The Ministry developed hospital P4P after recognizing the limitations of the previous model that had been solely based on accreditation status. Casemix index was included in the P4P model, to improve the appropriateness and fairness of the Ministry-hospitals relation. The analysis of P4P integration impact on casemix included 1,353,025 hospitalizations between 2011 and 2016. This revealed an abrupt increase in casemix among short-stay cases, and a gradual increase in medium-stay cases. Code-level analysis suggested this was attributable to a decrease in unnecessary hospitalizations and improved coding practices. The analysis of P4P impact on readmissions included 1,333,691 hospitalizations across 2011-2019. An abrupt decrease of cholecystectomy and stroke readmissions was found, but not of general and pneumonia readmissions. Our qualitative investigation allowed us to identify six patient perspectives, including satisfaction, health status, perceptions on each of quality, access and health system, and valuing of health, all of central relevance to health systems performance.

Conclusion: Hospital P4P in Lebanon led to several positive impacts, including improving the relation between hospitals and the Ministry of Public Health, and providing a tool for continuous development of the health system. The 2014 and 2018 P4P interventions improved system effectiveness and related patient outcomes, by decreasing unnecessary hospitalizations and decreasing some types of readmissions. The Ministry should develop its P4P model to capture the entire spectrum of hospital visits. Using appropriate interrupted time series analysis on readily available data is a useful way to evaluate the effects of health system interventions in contexts with limited resources. Patients in Lebanon highly valued health and supported improving public hospitals and measures to counter the influence of personal connections and money. Health systems can more widely engage people for their perspectives, and patients can have a fundamental role in shaping the values and functions of a health system.

Key words: health system, hospital performance, pay-for-performance, evaluation, patient perspectives, casemix, readmissions, Lebanon.

Classification system and/or index terms (if any)

Supplementary bibliographical information

Language: English

ISSN and key title: 1652-8220

ISBN: 978-91-8021-446-9

Recipient's notes

Number of pages: 140

Price

Security classification

I, the undersigned, being the copyright owner of the abstract of the above-mentioned dissertation, hereby grant to all reference sources permission to publish and disseminate the abstract of the above-mentioned dissertation.

Signature:

Date 2023-08-15

The Development and Evaluation of Hospital Pay-for-Performance in Lebanon

Casemix, Readmissions and Patient Perspectives

Jade Khalife



LUND
UNIVERSITY

Coverphoto: “Accompanying the patient on their journey”, by Ramzi Semrani.

Copyright pp 1-140 Jade Khalife

Paper 1 © 2017 Taylor & Francis Group

Paper 2 © 2020 Khalife J et al. (CC BY 4.0)

Paper 3 © Khalife J et al. (manuscript unpublished)

Paper 4 © 2023 Khalife J et al. (CC BY 4.0)

Faculty of Medicine

Department of Clinical Sciences

Lund University, Faculty of Medicine Doctoral Dissertation Series 2023:105

ISBN 978-91-8021-446-9

ISSN 1652-8220

Printed in Sweden by Media-Tryck, Lund University

Lund 2023



Media-Tryck is a Nordic Swan Ecolabel certified provider of printed material. Read more about our environmental work at www.mediatryck.lu.se

MADE IN SWEDEN 

To Oliver & Nadine

From those before us, through us, and for those that follow.

Table of Contents

Abstract	11
Preface.....	13
List of Papers.....	17
Abbreviations	18
1 Purpose, Goals and Overview	19
1.1 Purpose.....	19
1.2 Goals	19
1.3 Overview	20
2 Introduction	27
2.1 Pay-for-performance	27
2.1.1 What is ‘performance’?	27
2.1.2 Theoretical underpinnings	29
2.1.3 Hospital pay-for-performance	30
2.2 Casemix.....	35
2.2.1 Why casemix?	35
2.2.2 Casemix approaches and calculation	36
2.3 Readmissions.....	37
2.3.1 Why readmissions?.....	37
2.3.2 Readmission approaches	38
2.3.3 Pay-for-performance impact on hospital readmissions	38
2.4 Patient perspectives	39
2.4.1 Why patient perspectives?.....	39
2.4.2 What are patient perspectives?	40
2.4.3 Patient perspectives and pay-for-performance	41
2.5 Knowledge gaps on pay-for-performance.....	43
3 Background and Interventions.....	45
3.1 Lebanon.....	45
3.2 Health system.....	46
3.3 Interventions using pay-for-performance.....	46
3.3.1 Hospital pay-for-performance model of 2014	46
3.3.2 Hospital pay-for-performance model of 2018/2019	47

4 Conceptual Framework	51
5 Theoretical Foundation for Study Design and Analysis	55
5.1 On causality.....	55
5.2 Interrupted time series analysis.....	59
5.2.1 The four validities.....	59
5.2.2 Impact models using interrupted time series design.....	62
5.2.3 Newey ordinary least-squares regression (Newey-OLS)	63
5.2.4 Autoregressive Integrated Moving Average (ARIMA).....	64
5.3 Content analysis and focus group discussions	66
5.3.1 Qualitative content analysis.....	66
5.3.2 Focus group discussions	67
6 Methods and Materials	69
6.1 Study design.....	69
6.2 Data collection and preparation.....	71
6.2.1 Context of pay-for-performance.....	71
6.2.2 Casemix and readmissions.....	72
6.2.3 Patient perspectives	74
6.3 Data analysis	76
6.3.1 Context of pay-for-performance: descriptive analysis	76
6.3.2 Casemix index: ITS and code-level analysis	76
6.3.3 Readmissions: ITS analysis	78
6.3.4 Patient perspectives: qualitative content analysis.....	79
6.4 Ethical considerations	80
7 Main results.....	81
7.1 Why and how was hospital pay-for-performance developed?	81
7.2 What was the impact of pay-for-performance integration on healthcare effectiveness?	83
7.3 What was the impact of pay-for-performance on hospital readmissions?	90
7.4 What are patient perspectives on hospital care?.....	93
8 Discussion	101
8.1 Discussion of findings and relation to the literature.....	101
8.2 Methodological considerations	110
8.2.1 Casemix and readmissions.....	110
8.2.2 Context of P4P and patient perspectives	113

9 Implications for policy, practice and research.....	115
9.1 For the Lebanese health system and the pay-for-performance model..	115
9.2 For health systems and pay-for-performance in other countries	117
9.3 For future research	118
10 Conclusion	119
Acknowledgements	121
Epilogue	124
References	125

Abstract

Background: Pay-for-performance (P4P) has been widely used in healthcare, but there are few experiences of hospital-based P4P at scale. The evidence of impact from these has been mixed, and there has been increased recognition of the importance of different contexts, designs, incentives and other factors. In 2014, the Lebanese Ministry of Public Health integrated a P4P model for determining hospital reimbursement tiers. In 2018, this model was updated to include a readmissions component, in addition to the preexisting components such as casemix and patient satisfaction. The impact of these interventions was previously undetermined. This also provided an opportunity to contribute to some of the known knowledge gaps regarding hospital P4P. The purpose of this thesis was to describe the development and evaluate the impact of hospital P4P in Lebanon, and ultimately to contribute more broadly to improved design and implementation of value-based healthcare, particularly in limited resource settings.

Methods: This thesis uses a mixed methods approach, combining quantitative and qualitative study designs, to conduct four research investigations. The first paper uses descriptive analysis to address how and why hospital P4P was developed in Lebanon. The second and third papers both use an interrupted time series design on data collected from the Ministry hospitalization database. The former uses Newey-OLS regression, and the latter uses Autoregressive Integrated Moving Average models. The second paper analyzes the impact of the 2014 P4P integration on casemix index, and the third paper analyzes the impact of the 2018 model update on readmissions. The fourth paper uses qualitative content analysis on data collected from eight focus groups discussions with patient participants.

Results: The Ministry developed hospital P4P after recognizing the limitations of the previous model that had been solely based on accreditation status. Casemix index was included in the P4P model, to improve the appropriateness and fairness of the Ministry-hospitals relation. The analysis of P4P integration impact on casemix included 1,353,025 hospitalizations between 2011 and 2016. This revealed an abrupt increase in casemix among short-stay cases, and a gradual increase in medium-stay cases. Code-level analysis suggested this was attributable to a decrease in unnecessary hospitalizations and improved coding practices. The analysis of P4P impact on readmissions included 1,333,691 hospitalizations across 2011-2019. An abrupt decrease of cholecystectomy and stroke readmissions was found, but not of general and pneumonia readmissions. Our qualitative investigation allowed us to identify six patient perspectives, including satisfaction, health status, perceptions on each of quality, access and health system, and valuing of health, all of central relevance to health systems performance.

Conclusion: Hospital P4P in Lebanon led to several positive impacts, including improving the relation between hospitals and the Ministry of Public Health, and providing a tool for continuous development of the health system. The 2014 and 2018 P4P interventions improved system effectiveness and related patient outcomes, by decreasing unnecessary hospitalizations and decreasing some types of readmissions. The Ministry should develop its P4P model to capture the entire spectrum of hospital visits. Using appropriate interrupted time series analysis on readily available data is a useful way to evaluate the effects of health system interventions in contexts with limited resources. Patients in Lebanon highly valued health and supported improving public hospitals and measures to counter the influence of personal connections and money. Health systems can more widely engage people for their perspectives, and patients can have a fundamental role in shaping the values and functions of a health system.

Preface

This thesis came about through a mix of factors, including my personal interests, the people I met, and chance. I have had an interest in population health since childhood, and specifically in how a country can improve its population's health. My perspective included a recognition of the limited scalability of private initiative, the wide differences in national health systems, and the role of historic circumstances. I was partly influenced through reflections on my grandparents' social initiatives in their towns, particularly my maternal grandfather, Ishak Sejaan. I was also influenced by the impact of the 1975-1990 war in Lebanon, and other conflicts elsewhere. I considered private initiatives to be important, but they could not replace the need for a strong public health system.

During my medical studies I became curious about healthcare design, and the gap between the medical and public health fields. My interests in public health impact took me towards non-communicable diseases (NCDs) and tobacco control. In 2008, I had the opportunity to observe a few people working at the Ministry of Public Health (MoPH). That so much was dependent on the dedication of so few was rather inspiring. The following year I got involved in some independent projects at the MoPH. I had initially turned down the opportunity to work on the health system reforms planned by the MoPH, as I had wanted to focus on NCDs, in Lebanon and elsewhere. I was encouraged to 'try it for a few months and see' by Walid Ammar, the then-Director-General, whose vision guided the reforms (out of which hospital pay-for-performance evolved). Months turned to years, as my interest in health systems grew. It was during this time that I met Björn Ekman, then senior health economist at the World Bank, who would later become my thesis supervisor at Lund University (LU).

During my time at the MoPH, I became more aware of a second interesting gap. Aristotle introduced three key terms in his works, particularly in "Nicomachean Ethics" and "Politics". These were *episteme*, *techne*, and *phronesis*¹. The first referred to theoretical knowledge, which was predominant during my university studies, and is generally the focus in academia. The second term referred to practical knowledge or technical expertise, which I came to witness more in the operational setting of the MoPH. The different focus of university and health authority was an issue that interested me. And it was particularly relevant in the joint collaborations I was involved in, which are described in the thesis Paper 1. The third term, *phronesis*, referred to practical wisdom, the type one could only gain with experience, and of which I saw much in the technical (but not political) leadership of the MoPH.

Some years later, I had partially relocated to Denmark to join my family there, but was regularly in Lebanon as well. I had not specifically intended to pursue a PhD,

but it seemed a logical path to interact with other health professionals, develop myself further and have a local network. At Björn's suggestion, I enrolled part-time with the PhD program at LU, to investigate the development and impact of the MoPH hospital pay-for-performance (P4P) of 2014. A year later, we had a team across Lebanon and Sweden successfully apply for funding from the UK Joint Health Systems Research Initiative (HSRI), to continue developing the MoPH work on P4P, specifically through an implementation-based research project. This involved the American University of Beirut, LU and the MoPH.

Between 2016-2019 much of the work was focused on the operational details of this project, while the later research outputs would include articles for inclusion in my thesis. I was able to devote most of my efforts towards the operational details, while also developing a solid base for subsequent research investigations. A substantial part of this involved algorithmic thinking and statistical review for developing the required data and methods. Of course, having been involved in both the development and evaluation of hospital P4P in Lebanon may place particular demands on my conduct and scientific approach. My perspective is that all researchers and practitioners have some form of bias or another, to various extents, and perhaps influenced by past experiences, current dependencies, career paths and other factors. Many system-determined incentives seem unhealthy. The best we can do to address this is to be transparent, adopt rigorous investigative approaches, and involve other professionals or stakeholders where relevant. In the research included in this thesis, my colleagues and I strove to meet all three of these actions. One of the strengths of this thesis, and the preceding development of P4P, was the involvement of both researchers and practitioners. In essence, this brought *episteme* and *techne* together, at the individual and group level.

The P4P initiative in Lebanon also had some setbacks, which affected the implementation of the thesis studies. Similar to many projects dealing with health policies, we had to contend with the changing political environment in Lebanon, and specifically for the then-health ministers to sign the approval for the modified P4P model. Despite a delay of over a year, the updated model was officially approved in 2019.

Two broader setbacks were the political-economic crisis in Lebanon in late 2019, and subsequently the COVID-19 pandemic. The first had major implications in limiting the future development and use of the P4P model, at least in the near-term. It has been four years since the last application of P4P was used to determine hospital reimbursement tiers. Much of the know-how can still be used, but the challenges of the current context require wider reforms, and even health system redesign, if population health is our priority. I think most of the participants in the focus group discussions documented in Paper 4 of this thesis would agree. From my perspective, this can only be achieved with the creation of a single Lebanese

National Health System. At the current time, the political priorities are unfortunately elsewhere.

The COVID-19 pandemic had major implications for humanity, but also on my professional and personal life. From my perspective, the pandemic revealed that most health systems are far from being patient or people-centered. This issue also happens to be a focus of Paper 4 of this thesis. There are several examples of declarations and agreements at various levels regarding patient-centeredness. A notable one is the Framework on Integrated People-centered Health Services, adopted by member states at the 69th World Health Assembly (2016). Despite this, when faced by COVID-19, most countries had approaches centered on hospital burdens, not their population's well-being (and ignoring long-term sequelae, including Long Covid). Maintaining hospital capacity was the priority, while limiting infection was only relevant in the context of hospital capacity. From my perspective, this represented a monumental failure of public health and ethics, and more broadly technical incompetence and political indifference, albeit with some important exceptions. 'To err is human' is a universal truth, but it was also highlighted in the US Institute of Medicine 2000 report on building a safer health system. As individuals, we will make mistakes. However, it is distinctly more grave to have systems that do not self-correct based on scientific evidence. Although not directly relevant to my thesis, the response to the pandemic was highly informational. I increasingly recognize the limitations of our systems, and the importance of putting people first, and developing self-correction mechanisms based on logic and evidence. Multi-disciplinary collaboration and engagement of the public are essential ingredients in this.

During my work on this thesis I learned much from the past work of others. Most of this has been through reading, engaging with others, and much reflection in between. This went across different contexts and disciplines. Many scientific articles have been useful in this process, but on some topics I think select books provide a considerably greater depth of understanding. One example that has been very relevant to this thesis is interrupted time series analysis. Although I had been rather critical of the use and miss-use of statistical approaches before this PhD, this increased after I had the opportunity to read some select books regarding time series and related issues. This enabled me to more easily identify limitations of different articles, and attempt to improve my own. I have encountered much material on P4P of variable validity, particularly regarding construct and statistical conclusion validities using time series design (see chapter 5). Given that one can never have expertise in all relevant disciplines (myself included, of course), collaboration is indispensable to ensure the validity and meaningful conclusions of our scientific investigations. That, coupled with the spirit expressed by the likes of Ibn al-Haytham and Richard Feynman (see Chapter 8).

This thesis was written with health practitioners and researchers in mind, but also in a manner to be more widely accessible to the public. An unavoidable trade-off is involved regarding readability and detail. Some of the theoretical and historical aspects may be skipped by those more familiar with this. Throughout this thesis, the pronoun ‘we’ refers to my contributions and perspective, unless otherwise specified.

I have been fortunate to work with various practitioners and researchers in Lebanon, Sweden and elsewhere. I am also very grateful to have engaged with patients before and during my thesis work, and to be surrounded by family and friends that have made all the difference in life. I acknowledge many of these in the last section of the thesis, although a more appropriate place would be the front cover.

If one were to sum up this thesis into a single question, it would be along the lines of “Does hospital pay-for-performance work?”. However, it is rather obvious that the wide diversity of contexts, designs, incentives, measures, and other features does not provide a useful answer to a question framed in this simple manner. From my perspective, I think that there are instances where P4P may be beneficial. Perhaps what is most relevant is to increase our understanding of the factors that make success more likely, and in parallel ensure that we avoid or decrease unintended consequences. We should have a reasonable chance to succeed, without risking someone’s health or well-being. More specifically, for my research I developed P4P as the common thread linking specific components such as casemix index, readmissions and patient perspectives, while also considering aspects of health systems, complexity, sustainability, robustness, and integration.

Overall, I think this thesis’ findings reveal some positive effects of hospital P4P in Lebanon, and some areas that should be further developed in the future, in Lebanon and beyond. In light of the ongoing economic crisis, P4P can continue to have an important role in the health system. But more useful than P4P itself may be some of the tools and findings developed in this process, particularly towards the recovery and redesign of the health system.

More widely, health systems should carefully design P4P initiatives using multi-disciplinary collaboration and principles of participatory governance, and more rigorously evaluate their impact. They also should improve engagement of patients and the public, both for P4P and for health system development.

Health systems are increasingly challenged in this new age of pandemics, climate catastrophe and political-economic upheavals. People highly value health, and we should strengthen our systems to reflect this.

List of Papers

Paper 1

Khalife J., Rafeh N., Makouk J., El-Jardali F., Ekman B., Kronfol N., Hamadeh G., Ammar W. (2017). Hospital Contracting Reforms: The Lebanese Ministry of Public Health Experience. *Health Systems & Reform*. 3(1):34–41.

Paper 2

Khalife J., Ammar W., Emmelin M., El-Jardali F., Ekman B. (2020). Hospital performance and payment: impact of integrating pay-for-performance on healthcare effectiveness in Lebanon. *Wellcome Open Research*. 5:95.

Paper 3

Khalife J., Ammar W., El-Jardali F., Emmelin M., Ekman B. Impact of pay-for-performance on hospital readmissions in Lebanon: An ARIMA-based intervention analysis using routine data. *In manuscript*.

Paper 4

Khalife J., Ekman B., Ammar W., El-Jardali F., Al Halabi A., Barakat E., Emmelin M. (2023). Exploring patient perspectives: A qualitative inquiry into healthcare perceptions, experiences and satisfaction in Lebanon. *PLOS ONE*. 18(8):e0280665.

Abbreviations

ARIMA	Autoregressive Integrated Moving Average
BPT	Best Practice Tariff (UK)
CMI	Casemix index
CMS	Centers for Medicare & Medicaid Services (US)
CPT	Common Procedural Terminology
CI	Confidence interval
COPD	Chronic obstructive pulmonary disease
DRGs	Diagnosis related groups
ESPISP-2	2 nd Emergency Social Protection Implementation Support Project
HACRP	Hospital-Acquired Condition Reduction Program (CMS)
HICs	High income countries
HRRP	Hospital Readmissions Reduction Program (CMS)
HSRI	Health Systems Research Initiative
ICD-10	International Classification of Diseases, 10th Revision
ITS	Interrupted time-series
LMICs	Low- and middle-income countries
MI	Myocardial infarction
MoPH	Ministry of Public Health
OECD	Organisation for Economic Co-operation and Development
OLS	Ordinary least squares
P4P	Pay-for-performance
WHA	World Health Assembly
WHO	World Health Organization
VBP	Value-based purchasing

1 Purpose, Goals and Overview

“All men can see these tactics whereby I conquer, but what none can see is the strategy out of which victory is evolved”
– Sun Tzu (544-496 BCE)

This thesis has similar elements to a strategy: a purpose, goals and means. This chapter presents the purpose of the thesis, and the goals which serve as intermediate steps. The means are elaborated in subsequent chapters. We further provide an overview of the linkage between the thesis goals and the research studies, and depict the main building blocks of each paper.

1.1 Purpose

The purpose of this thesis is to describe the development and evaluate the impact of hospital pay-for-performance (P4P) in Lebanon, and ultimately to contribute to improved design and implementation of value-based healthcare, particularly in limited resource settings

1.2 Goals

The goals of the thesis are to:

- a. Describe how and why hospital P4P was developed in Lebanon. (*Paper 1*)
- b. Analyze the impact of P4P integration on healthcare effectiveness. (*Paper 2*)
- c. Describe how routine data and casemix index may be used for hospital performance. (*Paper 2*)
- d. Analyze the impact of P4P on hospital readmissions in Lebanon. (*Paper 3*)
- e. Explore patient perspectives on hospital care in Lebanon, and contribute insights that may improve P4P design and effectiveness. (*Paper 4*)

1.3 Overview

The linkage between the thesis goals and the four research studies is depicted in figure 1. The goals are connected with the particular problems tackled in each paper. The figure also highlights the conclusions of each paper.

A research canvas approach is then used to depict each of the four studies, using a framework consisting of nine specific cells, developed by John Latham ². These provide a concise summary of the main building blocks of each paper, and help to clarify their alignment and coherence.

Each canvas is organized into two groups: the ‘T’ that depicts the foundation of the paper (problem, purpose, conceptual framework and research questions), and the ‘U’ for the methods (overall approach, literature review, data collection, analysis and conclusions). The conceptual framework for each paper notes its relation to the cross-cutting topic of pay-for-performance. The problem component notes the potential contribution of the paper to the context of Lebanon and LMICs, as well as to the broader field of P4P.

Conclusion

This chapter provided the thesis purpose and goals; the linkage between the goals and research studies; and the main building blocks of each paper.

In the next chapter we will provide an introduction to pay-for-performance, and the three concentration areas this thesis deals with: casemix, readmissions and patient perspectives.

The Development and Evaluation of Hospital Pay-for-Performance in Lebanon Casemix, Readmissions and Patient Perspectives

The purpose of this thesis is to describe the development and evaluate the impact of hospital pay-for-performance in Lebanon. The ultimate purpose is to contribute to the improved design and implementation of value-based healthcare, particularly in limited resource settings.

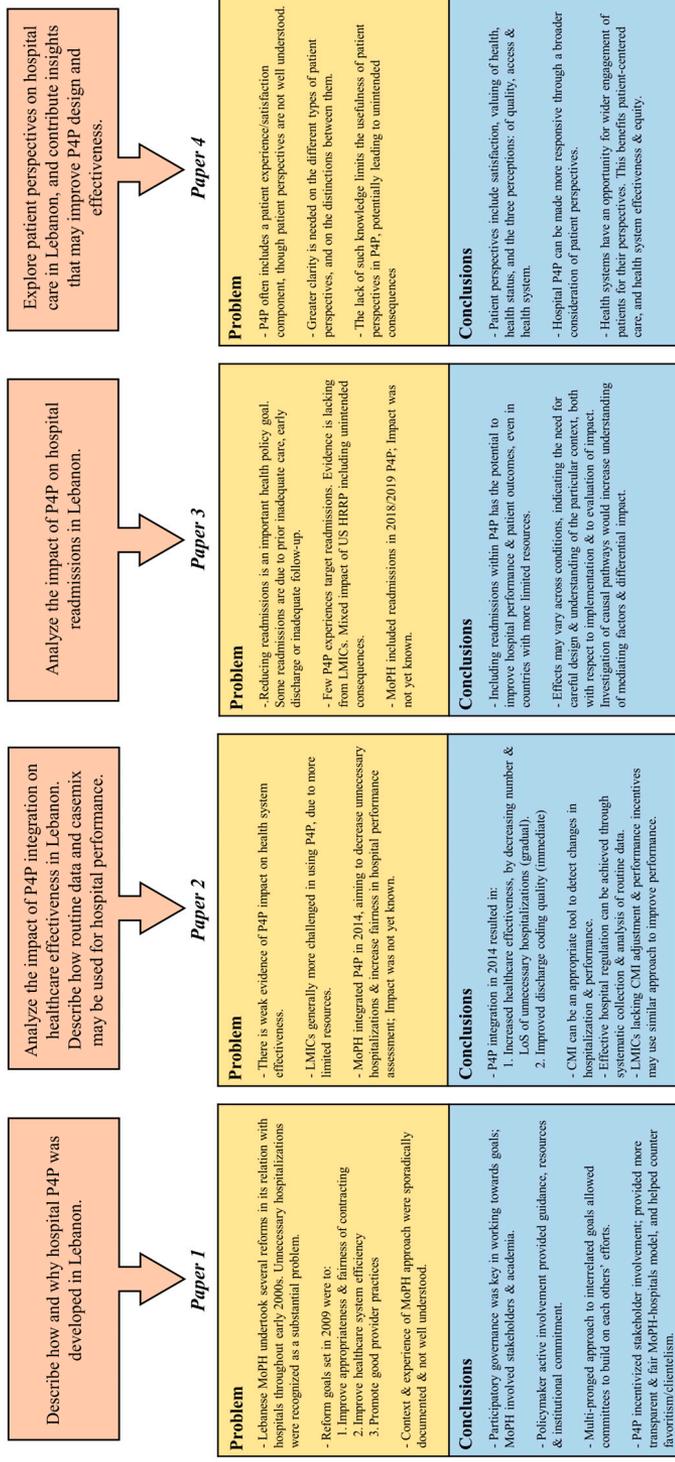


Figure 1: The linkage between the thesis goals and the four research studies.

Hospital Contracting Reforms: The Lebanese Ministry of Public Health Experience

Paper 1		
<p>Problem</p> <ul style="list-style-type: none"> - Lebanese MoPH undertook several reforms in its relation with hospitals throughout early 2000s. Unnecessary hospitalizations were recognized as a substantial problem. - Reform goals set in 2009 were to: <ol style="list-style-type: none"> 1. Improve appropriateness & fairness of contracting 2. Improve healthcare system efficiency 3. Promote good provider practices - Context & experience of MoPH approach were sporadically documented & not well understood. 	<p>Purpose</p> <ul style="list-style-type: none"> - To detail the experience of the MoPH 2009-2014 reforms, specifically the context & approach used to improve governance & regulation of hospitals, and counter favoritism & clientelism. - To investigate how & why the MoPH moved towards hospital pay-for-performance (P4P). 	<p>Research Questions</p> <ol style="list-style-type: none"> 1. What was the MoPH experience in pursuing the three goals set in 2009? 2. Why did the MoPH implement hospital P4P?
<p>Conclusions</p> <ul style="list-style-type: none"> - Participatory governance was key in working towards goals; MoPH involved stakeholders & academia. - Policymaker active involvement provided guidance, resources & institutional commitment. - Multi-pronged approach to interrelated goals allowed committees to build on each others' efforts. - P4P incentivized stakeholder involvement; provided more transparent & fair MoPH-hospitals model, and helped counter favoritism/clientelism. 	<p>Conceptual Framework</p> <ul style="list-style-type: none"> - Health reforms are influenced by various factors, including approach, stakeholder interests, and political & socioeconomic factors. - Alignment of interests provides a tool for the principal (payer) to incentivize the provider to improve outputs & outcomes. - A health system intervention linking measures to payment creates a financial incentive for hospitals to improve their performance vis-à-vis these measures. - Seen through the lens of principal-agent theory, P4P seeks to address recognized information problems within healthcare, particularly information asymmetry. 	<p>Literature Review</p> <ul style="list-style-type: none"> - The Lebanese health system is highly diverse, including a mix of public and private payers and providers. - Private hospitals dominate service provision, while the public sector is the major payer of hospital care. - In 2014 the MoPH contracted with 131 hospitals (105 private). - Hospital accreditation was linked to reimbursement tier in 2001. - Its inappropriateness as a sole determinant for reimbursement was soon apparent. - Stakeholders required a fairer contracting model, and the MoPH sought to incentivize performance.
<p>Data Analysis</p> <p>Descriptive analysis</p>	<p>Data Collection</p> <ul style="list-style-type: none"> - Review of reports & materials related to each of the Utilization Review, Performance Contracting & Admission Criteria committees. - Hospitalization database of the Ministry of Public Health. - Hospital performance scores & reimbursement tiers 	<p>Overall Approach</p> <p>Descriptive qualitative approach using project documentation materials, discussions with key personnel involved, and limited descriptive statistics.</p>

Figure 2: Summary of the main building blocks of Paper 1.

Hospital performance and payment: impact of integrating pay-for-performance on healthcare effectiveness in Lebanon

Paper 2

<p>Problem</p> <ul style="list-style-type: none"> - There is weak evidence of P4P impact on health system effectiveness. - LMCs generally more challenged in using P4P, due to more limited resources. - MoPH integrated P4P in 2014, aiming to decrease unnecessary hospitalizations & increase fairness in hospital performance assessment. Impact was not yet known. 	<p>Purpose</p> <ul style="list-style-type: none"> - To assess if P4P had an impact on healthcare effectiveness. <p>More specifically:</p> <ul style="list-style-type: none"> - To analyze if P4P affected the complexity of the average case hospitalized (CMI). - To quantify, & offer plausible explanations for changes, at the level of diagnoses & procedures. 	<p>Research Questions</p> <ol style="list-style-type: none"> 1. What was P4P impact on CMI level & trend, across public & private hospitals, by: <ol style="list-style-type: none"> a. Case type: medical, surgical & mixed b. LOS: short, medium & long-stays (medical) 2. What changes occurred in diagnoses & procedures by: <ol style="list-style-type: none"> a. Contribution to CMI changes b. Hospitalized cases
<p>Conclusions</p> <ul style="list-style-type: none"> - P4P integration in 2014 resulted in: <ol style="list-style-type: none"> 1. Increased healthcare effectiveness, by decreasing number & LOS of unnecessary hospitalizations (gradual). 2. Improved discharge coding quality (immediate) - CMI can be an appropriate tool to detect changes in hospitalization & performance. - Effective hospital regulation can be achieved through systematic collection & analysis of routine data. - LMCs lacking CMI adjustment & performance incentives may use similar approach to improve performance. 	<p>Conceptual Framework</p> <ul style="list-style-type: none"> - Alignment of interests provides a tool for the principal (payer) to incentivize the provider to improve outputs & outcomes. - A health system intervention linking measures to payment creates a financial incentive for hospitals to improve their performance vis-à-vis these measures. - Seen through the lens of principal-agent theory, P4P seeks to address recognized information problems within healthcare, particularly information asymmetry. 	<p>Literature Review</p> <ul style="list-style-type: none"> - P4P has been increasingly used in healthcare, but evidence on P4P impact is mixed. - P4P potential impact may be larger in LMCs, given relatively lower provider resources & more dynamic health reform context. - P4P usually used in distinct projects & not integrated within health system towards broad system goals; 'not seeing the forest for the trees'.
<p>Data Analysis</p> <ul style="list-style-type: none"> - Main impact measure was national CMI, calculated across 2011-2016, using medical & surgical procedure codes. - Single-group interrupted time series analysis model with Newey OLS regression was estimated. - Adjustment for seasonality & stratification by case type. - Code-level analysis used to attribute & explain CMI changes due to specific diagnoses/procedures. 	<p>Data Collection</p> <ul style="list-style-type: none"> - Hospitalization data of 1,353,025 cases under MoPH coverage, between January 2011 and December 2016. - Patient identifiers were anonymized. Fields included record number, case identifier, hospital code, admission date, discharge date, length of stay, total charge, medical code on discharge & surgical procedure code. - Algorithms developed to calculate monthly CMI, for medical, surgical & mixed cases; and for code-level analysis. 	<p>Overall Approach</p> <ul style="list-style-type: none"> - Intervention impact evaluation, using an interrupted time-series design. - Evaluates the impact of the 2014 Lebanese MoPH integration of hospital P4P on casemix (as a proxy indicator of unnecessary hospitalizations).

Figure 3: Summary of the main building blocks of Paper 2.

Impact of pay-for-performance on hospital readmissions in Lebanon: An ARIMA-based intervention analysis using routine data

Paper 3

<p>Problem</p> <ul style="list-style-type: none"> - Reducing readmissions is an important health policy goal. Some readmissions are due to prior inadequate care, early discharge or inadequate follow-up. - Few P4P experiences target readmissions. Evidence is lacking from LMICs. Mixed impact of US HRRP including unintended consequences. - MoPH included readmissions in 2018/2019 P4P; Impact was not yet known. 	<p>Purpose</p> <ul style="list-style-type: none"> - To estimate the impact of P4P on readmissions of general, pneumonia, stroke and cholecystectomy cases. More specifically: <ul style="list-style-type: none"> - To compare readmission rates before & after the announcement of P4P component on readmissions. - To analyze for heterogeneous changes across hospital sizes. 	<p>Research Questions</p> <ol style="list-style-type: none"> 1. What was the impact of the P4P components announcement on hospital readmissions? 2. Did impact vary across hospitals of different sizes?
<p>Conclusions</p> <ul style="list-style-type: none"> - The addition of a readmissions component to P4P resulted in a decrease in cholecystectomy and stroke readmissions. - No impact was found on general and pneumonia readmissions, also among small, medium and large hospitals. - Including readmissions within P4P has the potential to improve hospital performance and patient outcomes, but requires careful design and comprehensive understanding of context. 	<p>Conceptual Framework</p> <ul style="list-style-type: none"> - Alignment of interests provides a tool for the principal (payer) to incentivize the provider to improve outputs & outcomes. - A health system intervention linking measures to payment creates a financial incentive for hospitals to improve their performance vis-à-vis these measures. - Seen through the lens of principal-agent theory, P4P seeks to address recognized information problems within healthcare, particularly information asymmetry. 	<p>Literature Review</p> <ul style="list-style-type: none"> - US HRRP had mixed impact & unintended consequences. Included MI, heart failure, pneumonia, then COPD, CABG, hip & knee replacement. - Heart failure readmissions: up to 80% may be due to increased ER observation visits. Readmissions decreased, but mortality increased. - England BPT decreased hip fracture readmissions & mortality. - Challenges due to contextual changes, spill-over, measurement limitations, & overall design. - P4P anticipatory/short-term behaviors may differ from long-term behavior.
<p>Data Analysis</p> <ul style="list-style-type: none"> - Autoregressive Integrated Moving Average (ARIMA) models, with seasonality adjustment. - Stable medium-length time-series data. - Outcome measure was age-adjusted all-cause 30-day readmission rates, calculated at national level for each condition, and across hospital sizes for two conditions. 	<p>Data Collection</p> <ul style="list-style-type: none"> - Data extracted from MoPH hospitalization database, for 1,333,691 cases, between January 2011 and December 2019. - Case definitions developed for four conditions using literature, and algorithms developed for readmission identification & calculation using Stata v16. - Readmission definition: patient readmitted within 30 days of previous discharge. - 99 monthly data points were available (80 pre-intervention, 19 post-intervention). 	<p>Overall Approach</p> <ul style="list-style-type: none"> - Intervention impact evaluation, using an interrupted time-series design. - Evaluates impact of the 2018 MoPH announcement on the inclusion of a readmissions component to the ongoing P4P model.

Figure 4: Summary of the main building blocks of Paper 3.

Exploring patient perspectives: A qualitative inquiry into healthcare perceptions, experiences and satisfaction in Lebanon

Paper 4

<p>Problem</p> <ul style="list-style-type: none"> - P4P often includes a patient experience/satisfaction component, though patient perspectives are not well understood. - Greater clarity is needed on the different types of patient perspectives, and on the distinctions between them. - The lack of such knowledge limits the usefulness of patient perspectives in P4P, potentially leading to unintended consequences. 	<p>Purpose</p> <ul style="list-style-type: none"> - To explore how people with experience of being hospitalized perceive the healthcare, focusing on health perceptions, access to care, experiences of hospitalization & satisfaction of care. - To contribute insights that may improve MoPH P4P design & effectiveness. - To contribute to the knowledge on engaging patients towards person-centered health systems. 	<p>Research Questions</p> <ol style="list-style-type: none"> 1. What does health mean to people who have been hospitalized in Lebanon? 2. How do they perceive certain aspects of the health system? 3. How do they characterize their accessibility to healthcare? 4. How do they experience their hospital care? 5. Which factors do they identify as relevant to their satisfaction of hospital care?
<p>Conclusions</p> <ul style="list-style-type: none"> - Patient perspectives include satisfaction, valuing of health, health status, and the three perceptions: of quality, access & health system. - Hospital P4P can be made more responsive through a broader consideration of patient perspectives. - Health systems have an opportunity for wider engagement of patients for their perspectives. This benefits patient-centered care, and health system effectiveness & equity. 	<p>Conceptual Framework</p> <ul style="list-style-type: none"> - Explored what satisfaction meant for patients & what their experience & perceptions of care were. This was refined as the review & research unfolded into 'patient perspectives'. - Health system interventions such as P4P often claim to capture patient perspectives & intend to align provider interests through incentives. - Seen through the lens of principal-agent theory, P4P seeks to address recognized information problems within healthcare, particularly information asymmetry. 	<p>Literature Review</p> <ul style="list-style-type: none"> - Perspective of patients within health systems highlighted by WHA, US Institute of Medicine, OECD & others. - Patient-centeredness as "the doorway to all qualities". - Personal value pillar in value-based healthcare. - A return to Hippocratic medicine: accompanying the patient & meeting their individual goals. - Patient perspectives in P4P, e.g. Brazil, UK & US. - Patient experience & satisfaction recognized as multi-dimensional, but also ambiguous & under-theorized.
<p>Data Analysis</p> <p>Five overall themes were identified, which reflected the underlying meaning of the discussions:</p> <ol style="list-style-type: none"> 1. Health is everything 2. Being turned into second class citizens 3. Money & connections make all the difference 4. Wanting to be treated with dignity & respect 5. Tolerating letdown for the sake of right treatment <p>These themes were supported by 17 categories at manifest level of interpretation.</p>	<p>Data Collection</p> <ul style="list-style-type: none"> - Pilot, 8 FGDs & pile sorting exercise. - Random sample from hospitalization database, of persons discharged from hospital in preceding 3 months. - 42 participants (20 women, 22 men). - During July to September 2017. - Content analysis based on Granheim & Lundman (2004). 	<p>Overall Approach</p> <ul style="list-style-type: none"> - A qualitative approach was used, since it allows gathering information directly from participants, on a topic that is not well understood. - Focus group discussions were used, with persons who had been hospitalized under MoPH coverage. This was chosen due to our interest in a wide range of views & experiences, and to encourage discussion & explanation of issues.

Figure 5: Summary of the main building blocks of Paper 4.

2 Introduction

“Property which comes to a man from Zeus, both justly and cleanly, remains always steadfast. But if a man obtains it unjustly, improperly and with a profit-loving heart, or takes it contrary to justice through an oath, he seems at first to make some profit, but in the end he becomes poor again, and the design of the gods overpowers him”
– Theognidea 197-202 (6th century BCE)^{3p.57}

This chapter begins with a definition of health system ‘performance’, and the theoretical underpinnings for pay-for-performance. Having a shared understanding of this is necessary before proceeding with evaluating P4P. This is followed by the evidence on hospital-based pay-for-performance, with a focus on large-scale programs. We then move on to introduce three important areas addressed by this thesis: the casemix index as a measure of complexity, hospital readmissions, and patient perspectives. The last section presents a concise list of recognized knowledge gaps on pay-for-performance.

2.1 Pay-for-performance

2.1.1 What is ‘performance’?

Performance implies an action or achievement. In health systems policy development, performance is used broadly for a range of goals or expectations that health service providers (usually) are supposed to fulfill or achieve. Whereas quality of care is definable, albeit variably, performance is more subjective and less generalizable. Essentially, performance entails progress according to a set of measures. The linkage between these measures and improved healthcare is a necessary but secondary argument. Nevertheless, frameworks for health system performance are available, and it is useful to utilize or adapt these where relevant. Health systems share many similar features, but also important differences, for example across disease burdens, resources and scale.

Kruk and Freedman developed a framework based on performance indicators used in actual practice in LMICs, and adapted Donabedian's system of structures-processes-outcomes^{4 5}. This categorizes indicators under dimensions of effectiveness, equity and efficiency (see figure 6).

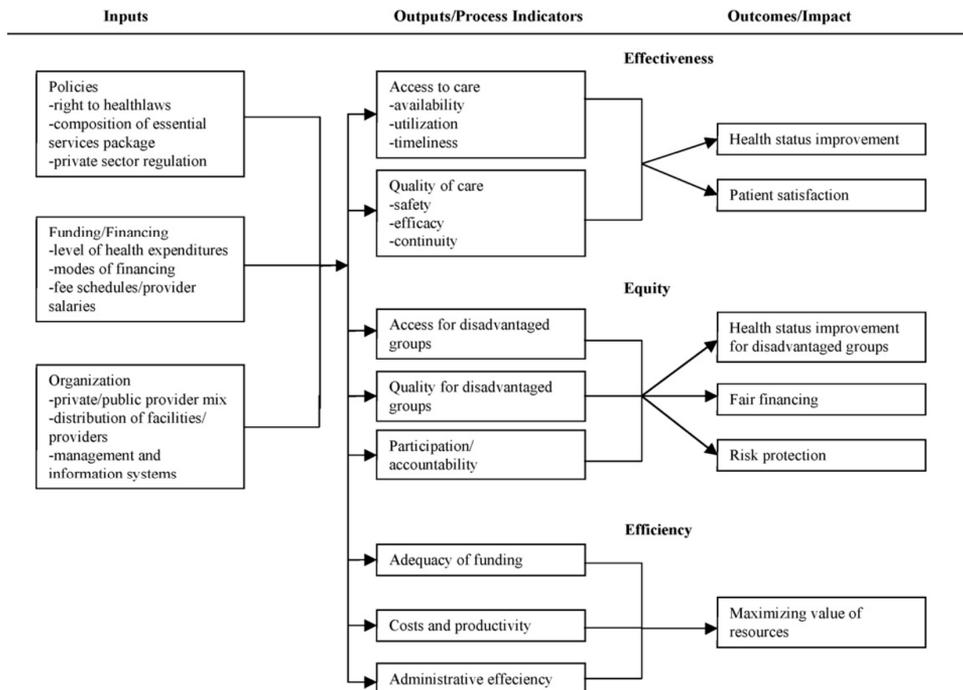


Figure 6: Kruk & Freedman framework for health systems performance measures⁴.

Structural measures have commonly been targeted in LMICs, as well as to a more limited extent process measures⁶. There has been increased attention towards outcome measures, although these are more challenging to target and track, particularly in countries with more limited resources⁷. Besides the necessary capabilities to calculate and monitor outcomes, the pathways, mediating factors and complexities of health present a formidable barrier to P4P in healthcare, relative to other fields such as engineering or other industry.

From a practical perspective, any P4P intervention requires a clear depiction of the measures used. Within hospital P4P, typical outcome measures have included the patient experience, readmissions and mortality, with ‘performance’ entailing changes across these measures.

2.1.2 Theoretical underpinnings

Contract theory provides a framework for understanding how different actors enter a formal (or binding) agreement, considering incentives and motivation. Unlike general equilibrium theory, which uses detailed contracts to reach efficient outcomes under ideal conditions, contract theory considers incomplete contracts and informational problems, such as information asymmetry. One type of contract is a principal-agent relation, whereby one actor (principal) delegates decisions or actions to a second actor (agent) to act on its (or some third party's) behalf.

Information asymmetry refers to an imbalance in the information available between actors. Such asymmetry may lead to problems such as moral hazard and adverse selection. Moral hazard refers to situations when an actor (agent) has an incentive to act contrary to the interests of a second actor (principal). Adverse selection occurs when one actor (principal) is less informed about another actor (agent), prior to entering into a contracting arrangement. In moral hazard, the relevant information is available before the contract is agreed, but not after. In adverse selection, the information is unavailable before the contract is signed.

Seen through a principal-agent lens, pay-for-performance is a tool to address the recognized information problems within healthcare, particularly information asymmetry⁸. This entails linking measures to payment, creating a financial incentive for an agent to perform vis-à-vis these measures. Within healthcare, an example of this would be an insurer or payer (e.g. health agency, ministry) contracting with a provider to deliver certain services. Thus, the first acts as principal, and the second as agent. The alignment of interests allows the principal to improve healthcare outputs and outcomes⁹. Within the context of this thesis, the principal is the Ministry of Public Health, and the agents are the hospitals with which it is contracted.

The recognition of informational asymmetry in healthcare was originally noted by Kenneth Arrow in 1963¹⁰. This included the roles of moral hazard, adverse selection, and more narrowly the roles of trust and qualifications within the doctor-patient relation. This emphasizes the significance of regulation both within the health profession and beyond.

It is relevant to recognize the several market failures that exist within the field of healthcare, which differentiate it from other markets¹⁰¹¹. Besides the prominent role of information asymmetry, this includes health as a public good, the roles of externalities, market power and equity. Health is also distinguishable due to its impact on catastrophic payments and human dignity. These further highlight the role of regulation, which also aligns with the free market of classical economists: a market that is free of land rent, bank usury and monopolies in private hands. The question is not whether to regulate, but what to regulate, by how much, by whom, through which means, and at which scale.

2.1.3 Hospital pay-for-performance

In its simplest sense, pay-for-performance may be thought of as a linkage of payment to performance. In practice, such alignment of interests between actors has occurred in various forms throughout history. The early 20th century included some large-scale applications within engineering and industry, such as automobile manufacturing. The use of P4P in healthcare, and more broadly performance-based financing, has spread over the past three decades, including a wide variety of programs and targets across various countries.

Many initiatives have focused on increasing finances directly towards service delivery and administration. More recently, there has been increasing emphasis towards a health system approach focusing on P4P integration and on health outcomes, although structure and process measures are predominant^{7 12 13}. Mixed findings have characterized P4P impact in healthcare, across high and low or middle-income countries¹⁴⁻¹⁷. Numerous lessons have been drawn from successes and failures of linking financial incentives to performance. The variation in results is generally attributed to factors such as different contexts, designs, implementation mechanisms and evaluation approaches¹⁸. Considering the complexity surrounding P4P, it may be more beneficial to engage in a realist approach examining how P4P affects outcomes and in what contexts¹⁴.

The experience of *hospital*-based P4P is more modest, with few large-scale experiences. One of the earliest examples was the sub-national P4P in the state of São Paulo, Brazil, which in 1998 included a performance component within a model used to set hospital budgets¹⁹. The component used patient satisfaction results and service volume targets to determine performance. At national level, the most prominent examples of hospital P4P were in the United States and the United Kingdom.

Advancing Quality Program, northwest England

This program targeted 30-day in-hospital mortality for pneumonia, heart failure and myocardial infarction, across 24 hospitals. Initial analysis of the program's first 18 months found a 6% mortality decrease among pneumonia patients only²⁰. However, subsequent analysis of the following 24 months found a greater decrease in mortality among hospitals not participating in the program²¹. Early improvements were therefore not sustained, and there was no difference among participating and control hospitals across the 42-month period. However, some findings were suggestive of a spill-over effect on non-targeted conditions at participating hospitals.

Premier Hospital Quality Incentive Demonstration (HQID), US

Launched in 2003 by Premier Inc. and the Centers for Medicare & Medicaid Services (CMS), the HQID program was the earliest large-scale hospital P4P,

including about 260 hospitals across 36 US states. This program targeted pneumonia, heart failure (HF), myocardial infarction (MI), coronary artery bypass graft (CABG), and hip and knee replacement. It included a mix of 34 process and outcome measures. Investigations into the program's impact at three years or later found no change in 30-day mortality for MI, HF, CABG and pneumonia, whether implicitly or explicitly linked to incentives^{22 23}. Similarly, no impact found on serious complications following CABG or replacement surgery²⁴. There was also no change among hospitals who were poorer performers at baseline, and weak evidence that receiving a bonus was associated with subsequent performance improvement^{25 26}.

Value-Based Purchasing, US

As a result of the 2010 Patient Protection and Affordable Care Act in the US, three programs were developed under CMS, dealing with value-based purchasing (VBP), the reduction of readmissions (HRRP), and the reduction of hospital acquired conditions (HACRP). The Value-Based Purchasing program was developed in 2012, using the infrastructure built since 2003 through the Hospital Inpatient Quality Reporting Program (which rewarded hospital reporting)⁷.

The VBP program initially used measures of clinical care processes measures and patient experience, to develop hospital total performance scores. A combination of achievement targets and improvement were used, with poor performance resulting in payments being withheld. The program later expanded to include measures on patient safety, efficiency and outcomes⁷.

By the first year of the program no improvement had been found, although some clinical processes improved just prior to VBP implementation²⁷. Mortality for both targeted (MI, HF, pneumonia) and non-targeted conditions was similar across 2,671 participating and 385 non-participating hospitals²⁸. Similar results were found for hospitals with a high share of Medicare patients (i.e. degree of exposure to program) compared with hospitals with a low share²⁹. The program was also found to disadvantage safety-net hospitals, which were more likely to be penalized by the program (scoring less on patient experience and process measures), as well as hospitals in socioeconomically disadvantaged areas^{30 31}.

A systematic review of six years of the VBP concluded that there were no meaningful changes in quality of care or patient outcomes³². The performance trajectory of hospitals across this period was also variable: a third of hospitals had mixed performance across six years, 24% improved, 14% maintained a good performance, 11% maintained a poor performance, and 18% declined³³. Other findings suggest that some hospitals may have been disadvantaged by metric changes during VBP's timeframe or methodological biases, had limited managerial influence on targeted conditions, or were not sufficiently incentivized³³⁻³⁶. A

comparison of hospital total performance scores found a large gap between the top 5% performing hospitals and others, with suggestions that the VBP rankings be redesigned to resemble the HRRP more, by rewarding improvement and incentivizing lower performing hospitals³⁶.

A comparison including 1189 US hospitals across P4P program early adopters (HQID) and late adopters (VBP) found no difference in process scores or mortality for non-targeted and targeted outcomes (pneumonia, HF and MI)³⁷. The complexity of the HQID and VBP programs (in contrast to HRRP, below) was suggested to have limited the meaningful engagement of hospitals, in addition to incentive design aspects³⁷.

Hospital Readmissions Reduction Program, US

The HRRP program initially targeted readmissions following myocardial infarction, heart failure and pneumonia. This was later expanded to include Chronic Obstructive Pulmonary Disease (COPD), hip and knee replacement surgery, and then coronary artery bypass graft surgery (CABG). Penalties were applied on hospitals with greater than average readmission rates, initially set at 1% of hospital reimbursement costs (in 2012), and subsequently up to 3%.

Early studies of the HRRP had encouraging findings, unlike those of the VBP. Following the announcement of the program (2010; prior to penalties), readmissions rates changed from having been stable throughout 2008-2010, to decreasing for all three targeted conditions, as well as for non-targeted conditions³⁸. Among hospitals that were subsequently penalized, the rates decreased more among targeted than non-targeted conditions, suggesting they acted to address their known performance (public reporting began in 2009)³⁸. The same was not found among non-penalized hospitals, suggesting less focused and wider system-approaches to decreasing readmissions³⁸.

Subsequent studies have suggested there may have been unintended consequences of the program. One concern had been the potential shifting of patients away from hospital admission and towards observation or emergency rooms (i.e. not hospitalized). Initially, no association was found between hospital visits and readmission rates, but later findings showed decreased readmissions were accompanied by increased emergency and observation room stays^{39,40}. Among heart failure cases, potentially up to 80% of the readmissions decrease may be accounted for by such shifting⁴¹.

Another concern was whether HRRP impacted the mortality rate of targeted conditions. No increased mortality was found among MI patients, however, studies arrived at opposite conclusions regarding patients with heart failure⁴¹⁻⁴⁷. Important methodological limitations of several studies, in addition to changes in HF epidemiology, precluded them from identifying a causal relation between HRRP

and mortality^{41 48}. This has led to considerable professional debate, and the impact remains uncertain^{41 49}. It has been suggested that HF patients tend to be older and more medically complex than other patients, and may therefore be more sensitive to outpatient longer-term care rather than inpatient care⁴².

Several recommendations for an improved HRRP have been proposed by professional societies, such as having a combined metric of readmission and mortality, and the inclusion of the entire spectrum of hospital interactions across observation, emergency and hospitalization⁴¹. Changes to incentive structure have also been suggested, noting that for heart failure the incentive for hospitals to reduce readmissions (through HRRP) was several times stronger than that to reduce mortality (through VBP)^{49 50}. This motivated the recommendation for combining the HRRP and VBP into a single comprehensive program⁴¹.

Hospital-Acquired Condition Reduction Program (HACRP), US

This program aimed to decrease hospital infection and increase patient safety across about 3,300 hospitals (in 2015). Three components formed the hospital HACRP score. Specific measures were used for central line-associated bloodstream infection and catheter-associated urinary tract infection. The third component was a composite patient safety score, initially including 8 measures (10 in 2021), such as rates of pressure ulcers, iatrogenic pneumothorax, in-hospital fall with hip fracture, and post-operative sepsis^{51 52}. A wider range of conditions was targeted through this program, compared to VBP and HRRP.

Considerable limitations of the composite score had been noted at the start of HACRP implementation, including issues with weighting, risk adjustment, surveillance bias and areas targeted⁵³. There is an overlap of measures used between HACRP and VBP, which is due to legal requirements, alignment efforts and the lack of alternative appropriate measures⁵¹. Unlike other programs, the HACRP penalized only the worst-performing quartile, with 1% of their total payment⁵¹. Thus, the incentive to improve primarily applied to hospitals within this lowest category.

Evaluations of the HACRP impact found weak or no change in hospital acquired conditions, both compared to historical trends as well as to non-targeted measures⁵⁴⁻⁵⁶. This highlighted concerns regarding the composite score, but also the ability to impact change on major harm or mortality⁵⁶. Hospitals serving more medically complex or socioeconomically deprived populations were more likely to be penalized⁵⁷⁻⁵⁹. Such hospitals were more likely to move from penalized to non-penalized status than others, however, when considering all hospitals, penalization did not result in improved patient safety^{55 60}.

A repeated finding regarding ‘Magnet’ hospitals, known for high nursing standards and quality improvement focus, had worse or similar performance in HACRP

measures ⁶¹⁻⁶³. Hospitals with more quality accreditations (e.g. Joint Commission, Commission on Cancer) or having better performance on non-HACRP processes/outcomes were also more likely to be penalized by the program ⁶⁴. This suggested the need not only for improvement of both HACRP and other non-CMS programs, but also for broader quality initiatives that are not limited to individual domains (e.g. nursing alone) ⁶¹⁻⁶³.

Altogether, these findings suggested important opportunities for improved program design and fairness, though several of these had been previously noted ⁵³. Foremost among these were for stratified hospital comparisons, and improvements to risk adjustment, monitoring systems, and measure selection and validity ^{53 65-67}.

While the findings from the VBP, HRRP and HACRP programs may be somewhat disappointing, given their limited successes, it is especially important to highlight the valuable lessons learned. Replicating their same course would be neither advisable nor ethical. However, the evidence that resulted from evaluations of these programs provides important opportunities for developing effective P4P initiatives.

Financial Incentive for Quality Improvement, France

The program ‘Incitation Financière à la Qualité’ (IFAQ) has been gradually developed since 2012, by several institutions and including hospital stakeholders ⁶⁸. Hospitals are brought together in comparison groups, and quality assessed using a set of indicators, with results used to rank hospitals. The comparison groups include acute care, rehabilitation care, home care and dialysis units. Groups are further subdivided based on volume and casemix. Several indicators are used, such as certification, patient-reported experience measures, pain assessment and quality of hospital discharge summary ⁶⁹. To improve ease of understanding, no more than 10 indicators are used within each comparison group, and an average success rate used to determine overall performance. Thresholds are set for remuneration, and the financial compensation is allocated based on their economic volume and success rate.

There is limited documented research regarding the impact of the IFAQ P4P in France. An evaluation of the 2012-2014 pilot phase of the program comparing 185 case hospitals and 192 control hospitals did not find a difference across nine process quality indicators ⁷⁰.

2.2 Casemix

2.2.1 Why casemix?

If two hospitals differed by cost-per-case or by outcomes such as mortality, one's initial thought would likely be that one receives sicker patients than the other, on average. Therefore, if we intend to be fair in comparing hospitals, we would have to measure and adjust for the severity of illness. While adjustment is a rather straightforward mathematical exercise, measurement is not; how do we compare the severity across thousands of medical conditions? The most widely used approach has been through resource consumption acting as a proxy for severity of illness. The 'casemix index' of a hospital is calculated based on this and reflects the aggregate risk of all persons hospitalized at that institution^{71 72}. The same can be applied at population levels, such as across regions or nationally by time.

Since the 1980s dozens of countries have developed their casemix systems, beginning in the US, then in some European countries and other regions. The applications of casemix are varied, and can be broadly categorized into three areas: policy, management and research⁷³. Within policy, casemix can be used for payment or reimbursement, planning of healthcare resources and program evaluation. Some uses within management include quality control and performance comparisons, planning and budgeting. The most important research application is for casemix as a risk adjustor when comparing utilization and outcomes.

In the context of performance, casemix is not typically a target in and of itself. A hospital's casemix is what it is; we cannot prescribe what it should be. However, being a proxy for illness severity, changes in casemix provide information on changes in hospital behavior. For example, this may include increased unnecessary or low-risk hospitalizations, or selective admission of specific diagnoses. A hospital's casemix may also strongly impact its performance across different measures. For example, readmissions are more frequent in hospitals receiving more severely ill patients (higher casemix index)^{74 75}. Similarly, hospitals with increased technical capability tend to have a higher casemix index, which is at least in part due to greater patient severity⁷⁶. Hospital ownership status has a more mixed relation with casemix, as it is always secondary to other factors (e.g. patient severity). In different contexts, a higher casemix may be found in public or private hospitals, which may vary by diagnostic category, and sometimes being a consequence of some private hospitals pursuing more profitable and less complex hospitalizations (sometimes referred to as 'cream-skimming')⁷⁷⁻⁷⁹.

2.2.2 Casemix approaches and calculation

The calculation of casemix is essentially the reduction of information from hundreds, thousands or more hospitalizations, into a single number. Such a process involves several parameters that can affect the casemix result. Sound statistical skills and knowledge of the local healthcare system are essential precursors for developing a fair and accurate measure of a hospital's casemix.

Diagnosis Related Groups (DRGs) are usually used to derive the casemix index. DRGs were developed at Yale University in the 1960s as a tool to facilitate hospital comparisons and evaluation, using administrative data⁸⁰. Similar to the concept of the casemix index, the function of DRGs is to reduce information into smaller blocks, which allow uses for various purposes, such as comparing performance or prospective payment. This typically involves grouping hospitalizations based on clinical data (diagnosis and procedures), age, sex, resource utilization (cost, length of stay) and severity. DRGs were first adopted by the US CMS in 1983, primarily as a cost-containment tool⁸¹. This supported the shift from a system based on fee-for-service reimbursement, to one using prospective payment based on DRGs⁸². Generally, the cost-containment goal was not successful, due to various provider responses⁸². Contrary to the US, the primary purpose in early European adopters was for increased fairness and efficiency, given their systems having per-diem payment and global budgets⁸¹. It is for such reasons that a DRG approach may be advantageous, although it also brings with it new risks and challenges that necessitate strong monitoring and information systems⁸³.

Many countries lack DRGs, which require considerable investment and maintenance by national health systems. In such contexts, casemix index can be calculated using diagnosis and procedures codes directly. Such an approach has been shown to be valid and feasible⁸⁴. The approach remains one of information reduction, and what differs is a question of the type and volume of grouping involved for the 'averaging' process. The strength of a well-designed DRG approach would allow a more accurate reflection of average patient risk, whereas the strength of a well-designed code-based approach would require fewer assumptions.

The standard generic formula for casemix index calculation as used by the US CMS is presented in equation 1^{85 86}. While variations exist, the underlying concept involves using a standard weight for each code defined based on the population average; a numerator and denominator of hospital and population cases, respectively; and a summation across all available codes.

$$CMI_h = \frac{\frac{\sum W_g N_{gh}}{\sum N_{gh}}}{\frac{\sum W_g N_{gn}}{\sum N_{gn}}}$$

Equation 1

Here, h is the hospital; W_g is the weight calculated for each ICD/CPT code; N_{gh} is the number of cases within each code in hospital h ; and N_{gn} is the number of cases within each code in the total population.

2.3 Readmissions

2.3.1 Why readmissions?

Hospital readmissions are frequent, costly and sometimes life-threatening⁸⁷. The reduction of readmissions is an important health system goal, with specific national policies pursued across several countries⁸⁸. This works firstly in the interests of patients themselves, and secondly for improved system effectiveness. Using the Kruk and Freedman framework, readmissions may be considered within quality of care, under the dimension of effectiveness^{4 89 90}. Pathways to impact may range across improving initial treatment to improving discharge instructions.

Two broad categories of readmissions may be defined: planned and unplanned. Planned readmissions are those intentionally scheduled as part of patient care, such as for chemotherapy, kidney dialysis or cardiac catheterization⁹¹. All other readmissions are considered unplanned, including acute illness or complications of previously received care. Only unplanned readmissions are a signal of impaired quality of care. Special methods involving algorithms are typically used to distinguish between the two readmission categories.

The risk of readmission is influenced by several factors, which can be identified by the extent to which they are modifiable. Non-modifiable factors include those at the individual level, such as patient age, comorbidities and the severity of the initial hospitalization⁹²⁻⁹⁶. Community-level factors are also an important factor, including socioeconomic status and poverty^{92 94 97}. Factors that to some extent are modifiable include hospital quality of care, as well as the discharge instructions given and patient compliance with these. Other relevant factors include hospital size, volume and geography, but these are less understood^{96 98-100}.

2.3.2 Readmission approaches

The most commonly used measure is 30-day all-cause readmissions, although 7-day, 60-day or 90-day readmissions have also been used. The rationale for having 30-days is because we expect the large majority of such readmissions to have been preventable, and therefore provide useful information for operational or policy decision-making. Public payers predominantly use all-cause readmissions, albeit with a few specific exclusion criteria (e.g. trauma, chemotherapy) to improve the validity of the performance measure. Readmissions can also be cause-specific, and these tend to be used by private insurance. The overall logic of using all-cause readmissions is that the body is complex, and at a population scale it is difficult to rule out a primary or contributing relation between most readmission and initial admission diagnoses. Also, using an all-cause measure restricts the potential for circumvention by some providers using coding changes. From the perspective of national health systems, it is important to get readmissions generally right, rather than specifically wrong. However, this also underlines the importance of having a sufficient denominator to minimize the role of miss-attribution or chance.

Readmission is generally calculated as the number of readmissions as a proportion of index cases, within a specific timeframe. Different systems may use variations, including for example a comparison of expected to actual readmissions and adjusted by national average. In all circumstances, the numerator is a subset of the denominator population, and all of the latter has the ‘opportunity’ to be within the numerator. This second point is relevant when considering data cut-offs at specific time-points. Hospital readmission rates are typically risk-adjusted, for example for casemix and age, to allow comparisons across hospitals.

Various prediction models for readmission risk have been developed, typically using administrative data to predict the likelihood of readmission of different patients. A direct application of these would be to inform interventions for reducing readmission. However, such models have been found to be poor at risk prediction, commonly with only 60-70% chance of a correct result ¹⁰¹.

2.3.3 Pay-for-performance impact on hospital readmissions

Financial incentives have been linked to the reduction of readmissions through pay-for-performance, including a few large-scale programs. We previously elaborated on the mixed impact and design limitations of the HRRP in the US (see 2.1.3). In contrast to this, the 2010 Hip Fracture Best Practice Tariff (BPT) in England focused specifically on hip fractures and had a simpler P4P design ^{102 103}.

The BPT scheme paid hospitals a supplement for each hospitalization that met six specific process-based criteria (e.g. assessments, time to surgery) ^{102 103}. Following implementation, mortality decreased among elderly patients with hip fracture, and

a previously increasing trend of readmissions was reversed ¹⁰². However, it is relevant to note the difficulty of disentangling the impact of the P4P itself, as the BPT was part of a broader initiative also involving a national clinical audit and database development ¹⁰².

2.4 Patient perspectives

“To cure sometimes, to relieve often, to comfort always.”
– aphorism adopted at the Saranac Lake sanatorium (19th century)

2.4.1 Why patient perspectives?

Listening to persons seeking care has been recognized by healers for thousands of years, across ancient civilizations including those in China, Egypt and India. A fundamental aspect of early Hippocratic medicine was accompanying the patient and meeting their individual goals. A modern age revival of the role of patients began in the 1950s, through the works of Michael and Enid Balint in the UK, and George Engel in the US ¹⁰⁴⁻¹⁰⁶. These early initiatives emphasized the need to understand and empathize with the patient, and recognize the social, mental and environmental context of patients.

Over the past four decades several landmark reports have underlined the inclusion of patients and their perspectives. These include Griffith’s Report (1983, UK), the Institute of Medicine (IOM; 2001, US), and the World Health Assembly, which in 2016 adopted the Framework on Integrated People-centered Health Services, stressing the patients’ role in defining their needs and co-producing health services reforms ¹⁰⁷⁻¹⁰⁹. In 2018, the IOM also called for an expansion towards person-centeredness, whereby the care provided is “respectful of and responsive to individual preferences, needs and values” ^{110 111}. Patient-centeredness has also been called “the doorway to all qualities”, a view that is also adopted by the World Health Organization, the World Bank and the Organisation for Economic Co-operation and Development (OECD) ^{111 112}. Physician-patient communication skills also became part of many national medical curricula. It is notable, however, that patient engagement is predominantly at the individual level regarding their treatment, and only rarely do health systems engage patients regarding system development or reform.

2.4.2 What are patient perspectives?

The earliest conceptualization of patient perspectives was through the concept of ‘patient satisfaction’. This has been extensively researched since the 1960s, and also gave rise to hospital patient satisfaction surveys in many countries¹¹³. As early as 1966, Avedis Donabedian emphasized that the effectiveness of care to achieve health and satisfaction “is the ultimate validator of the quality of care”⁵. But precisely what satisfaction meant was less clear, and the difficulty of its measurement was also recognized⁵.

Expectation theories were predominant in approaching patient satisfaction, considering patient expectations to play a central role in satisfaction. These included value expectancy model, fulfilment theory, discrepancy theory, disconfirmation theory and others¹¹⁴. However, the evidence has failed to support a major role for these theories, perhaps in part due to the complexity of expectation itself (different types and levels), and the influence of other factors on satisfaction¹¹⁴⁻¹¹⁶. Another critique is that patient satisfaction was derived from consumerist theories (portraying patients as customers), yet the receipt of healthcare was a more complex phenomenon than a consumer product. Most patients cannot be expected to ‘step out’ of their illness and take on the role of a ‘rational consumer’, nor have the technical knowledge to evaluate some aspects of care¹¹⁷. Also, in many cases patients cannot easily turn down ‘the product’ or seek an alternative¹¹⁷.

This is not to say that expectation does not play its role in satisfaction; it most likely does. Both expectation and satisfaction are multi-dimensional, but they are also ambiguous and under-theorized^{114 118}. That several decades of research had not considerably clarified patient expectation and satisfaction underscores the need for another approach regarding patient perspectives. It is relevant to note that satisfaction also closely involves emotions, and is of a generally subjective nature¹¹⁸. This poses additional challenges to understanding satisfaction. Altogether, although efforts to understand satisfaction should still be pursued, there remains a need for a wider approach which considers other patient perspectives as well.

Measures of patient perspectives that were later developed included those of the patient experience. These may be considered more focused on perceptions of quality of care (particularly structures and processes), compared to satisfaction measures. It is challenging to generalize across these measures, since different tools have incorporated different dimensions. Some notable examples of patient experience tools include the Ethiopian Patient Experiences with Inpatient Care (I-PAHC), Indian Patient Perceptions of Quality (PPQ), and the US Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS)¹¹⁹. The HCAHPS, for example, includes 29 questions, of which 10 are used by CMS for the VBP hospital scoring: two global satisfaction measures (overall rating of hospital, recommendation of hospital), three communication measures (doctors, nurses,

medications), three clinical process measures (staff responsiveness, care transition, discharge planning), and two environmental measures (cleanliness, quietness)^{120 121}. Despite the widespread use of tools for measuring patient perspectives, this remains a topic that is not well understood.

The evidence on the relation between patient experience and health outcomes is unclear. Various studies have found a relation between better experiences and improved outcomes such as mortality, readmission and complications, while others did not, or otherwise had mixed findings¹²²⁻¹²⁷. Overall, positive associations between patient experience and clinical outcomes are more commonly found than no associations¹²⁸. Mixed findings may be in part due to study designs and measure limitations, although it is notable that the dimensions of patient-doctor and patient-nurse communication tend to be often associated with better outcomes than other patient experience dimensions^{123 124 127}. Nevertheless, a better patient experience is an outcome in and of itself, or, as Donabedian noted ‘health *and* satisfaction’^{5 129}. It is also the one aspect of care that is evaluated directly by patients, not providers¹²⁹.

A common and major limitation of patient survey tools is their validity and reliability, which should be addressed to better understand patient perceptions and inform quality improvement¹³⁰⁻¹³². Also, the full potential of patient engagement is largely unexplored in some directions, and such tools also do not capture all the information that is available from engaging patients. Some have suggested that patients should be actively engaged to co-design health systems, with a focus on the actual experiences, and not only processes and outcomes in place^{133 134}.

2.4.3 Patient perspectives and pay-for-performance

Patient perspectives accounted for 25% of the total performance score within the VBP in the US. This formed a strong incentive for hospitals to improve the patient’s experience, at least nominally. The HCAHCS tool was used to evaluate patient experience, using random samples of patients discharged between 48 hours and six weeks. Investigation at years three, four and six after VBP announcement found no impact of the program on patient experiences^{36 121 135 136}. There was also no difference between safety-net and non-safety-net hospitals, noting that the former group were more likely to be penalized (further deteriorating resources)^{121 137}. These findings included the two global satisfaction measures, on 9 or 10 out of 10 rating and on recommending the hospital, whose long-term trends remained suggested a slowing down after VBP, but possibly attributable to other factors¹²¹.

Given the failure of VBP to improve patient experience, some suggestions have included the need for program redesign, measuring experience in non-hospital settings, and interventions beyond financial incentives or P4P^{121 135}. Other concerns have also been raised, such as instances of hospitals providing designer gowns or

valet parking, and of consultants coaching nurses' verbal communication ('teaching to test'), all to influence patient perceptions^{117 130}. In addition, caregiver behavior may also become more guarded, for example physicians failing to engage patients on smoking or obesity, or on non-favorable prognosis, for fear of antagonizing patients^{130 138}. Limitations to surveying patient experience also include the potential for over-prescription or unnecessary care (e.g. treating solely to 'satisfy the patient'), potentially harming patient outcomes and increasing costs¹³⁹. Such evidence provides increased awareness of the complexity involved in measuring patient perspectives, and the importance of incentive design.

One review of interventions to improve patient experiences (as measured by HCAHCPS) concluded there was minimal evidence to inform on what interventions work at scale, and in which contexts¹⁴⁰. A prominent limitation noted by the review was the abundance of pre-post designs which do not account for secular trends¹⁴⁰. In following the pathway from intervention to change in patient perceptions, it is worthwhile to revisit some of the assumptions involved. It is assumed that the intervention will change provider behavior (or not); that any change will be perceived by patients, and that the measurement tool would detect any change. Unlike the measures used for outcomes such as readmission or mortality, surveys on patient perspective have a lower precision. This is in part due to the nature of the data used (survey on a sample), but also non-response bias with, for example, sicker patients being less likely to respond^{141 142}. Generally, responses to patient surveys tend to be very positive, and can overstate the perceptions of patients¹⁴³. For this reason, it is typical to find satisfaction proportions of 80-90% across hospitals in different countries.

2.5 Knowledge gaps on pay-for-performance

There are several knowledge gaps regarding different aspects of P4P. A non-exhaustive list which is not specific to this thesis includes:

1. How can stakeholder buy-in and engagement be achieved?
2. What are the contextual factors, including health system, organizational culture and patient population?
3. What are appropriate measures or components?
4. Which measures or components reflects quality of care?
5. A limited understanding of impact on outcomes.
6. How can P4P be patient-centered?
7. To what extent can findings be generalized to other contexts?
8. How sustainable is P4P?
9. What is the optimal incentive design?
10. How can P4P be integrated with other payment models?
11. What are appropriate transparency and feedback mechanisms?
12. What are the unintended consequences, spillover effects, and impact on individual morale?
13. To what extent P4P may exacerbate inequalities.
14. How can P4P integrate social determinants to adjust for outcomes?
15. What is the long-term impact on hospitals and health system?
16. What are the perceptions of hospitals, health professionals and the public towards P4P?

This thesis primarily addresses gaps 1-6, and contributes to a more limited extent regarding gaps 8-11.

Conclusion

This chapter provided a definition of health system ‘performance’; the theoretical underpinnings for P4P; the evidence on hospital P4P; an introduction to each of casemix index, hospital readmissions, and patient perspectives; and a list of recognized P4P knowledge gaps.

In the next chapter we will provide an introduction to Lebanon, where this thesis work is located, and to the Lebanese health system, and the P4P models implemented by the MoPH in 2014 and 2018/2019.

3 Background and Interventions

“Pity the nation that wears a cloth it does not weave, eats a bread it does not harvest, and drinks a wine that flows not from its own wine-press.”
– Gibran Khalil Gibran (1883-1931)

This chapter begins with an introduction to Lebanon and its health system, and then continues with information about the P4P models of 2014 and 2018/2019, including their components.

3.1 Lebanon

Lebanon lies in West Asia, on the eastern coast of the Mediterranean Sea. Its location resulted in it being at the cross-roads of civilizations for several thousands of years. It has an area of about 10,452 km², and currently a population of about 6.8 million people, including almost 2 million refugees, most of whom have arrived after 2011 due to the conflict in neighboring Syria. It is among the top 20 countries by population density. The geography of Lebanon includes four regions: a coastal plain, two mountain ranges, and the valley between them. It has a length of about 217 km, and a width that varies between 48-89 km.

In the modern age, after gaining its independence in 1943 Lebanon has had alternating periods of stability and conflict, including the 1975-1990 war. Since 1990, Lebanon has struggled to develop in several sectors, due to a mix of factors including corruption and poor governance. The period starting 2005 has been marked with intermittent political instability, in parallel with conflict in other countries in West Asia.

More recent developments that have shaken the country include the economic crisis that began at the end of 2019, largely due to a national Ponzi scheme and poor governance. The impact of the crisis was further exacerbated in 2020 by the COVID-19 pandemic, and the explosion at the Beirut port in August of the same year. The consequences of the economic crisis included doubling to 82% of the population into multi-dimensional poverty¹⁴⁴.

3.2 Health system

The Lebanese health system is characterized by a mix of public and private payers and providers. Most payers are public, while most providers are privately owned. About four-fifths of citizens are under the coverage of the Ministry of Public Health or the National Social Security Fund (NSSF) ¹⁴⁵. The MoPH coverage extends to citizens who lack NSSF or other formal insurance and includes hospital care. The NSSF coverage is wider, additionally including subsidization of medication and outpatient care, and covers formal sector employees and their dependents.

The 1975-1990 period was expectedly a highly unregulated environment. Out of necessity, a large increase in private health facilities, given the decimation of the public infrastructure. By the early 1990s, there was a large number of private hospitals, with a widely variable quality of care. The stability of the 1991-2005 period saw increased initiatives to improve hospital care. This included a national hospital accreditation program, public hospital construction, and the development of a national network of primary care centers. Alongside other MoPH initiatives, this contributed to the improvement of outcomes such as infant and maternal mortality, life expectancy and out-of-pocket expenditure on healthcare.

In 2001, the MoPH linked hospital accreditation with reimbursement tiers. This financial incentive was intended to motivate hospitals to improve their quality of care, vis-à-vis achieving accreditation ¹⁴⁶. Within a few years, the limitations of this model had become more apparent. These included the absence of a measure for hospital case complexity and of any outcome measures. It was within this context that new model was initiated, namely hospital pay-for-performance.

3.3 Interventions using pay-for-performance

3.3.1 Hospital pay-for-performance model of 2014

In 2013, investigation of the long-term impact of the accreditation-reimbursement linkage had revealed that case complexity varied widely within and across reimbursement tiers ¹⁴⁷. This argued against the assumption that a higher tier implied more complex cases. Hospitals and other stakeholders were engaged regarding these findings during 2013-2014. In August 2014, the MoPH held an engagement event for hospital leadership and managers, announcing the adoption of a new model linking performance to reimbursement. This was legalized through Ministerial Decision #1980/1 (November 2014). The new model included a mix of weighted components, which were used to determine the Total Performance Score (TPS) of

each hospital (see table 1). The components were weighted based on their perceived relative importance by the MoPH, but also engaging the Syndicate of Private Hospitals.

Table 1: Components of the 2014 P4P

#	Component	Weight
1	Accreditation	40%
2	Casemix index	35%
3	Patient satisfaction	10%
4	ICU case proportion	5%
5	Surgical case proportion	5%
6	Deduction proportion	5%

The model retained accreditation status to maintain some direct incentive towards improving structures and processes. The TPS of hospitals was standardized using z-scores. All hospitals with an above-average TPS were categorized in the top reimbursement tier (T1); those within 0.5 standard deviations of the mean were categorized in the middle tier (T2); and those below 0.5 standard deviations were categorized in the lower tier (T3). The same model was applied among public and private hospitals, but standardized separately. This meant that public hospitals were compared with each other, and the same applied to private hospitals.

3.3.2 Hospital pay-for-performance model of 2018/2019

In January and March 2018, the MoPH held two events announcing the updated pay-for-performance model. These events included hospital leadership, quality managers and financial managers. The first event was centered on the policy level, while the second event was largely at the technical level. The new model benefited from updated literature and the lessons learned following the 2014 model, including its impact on different measures. The new model was legalized through Ministerial Decision #925/1 (May 2019), which also disseminated the component weights (see table 2). The period between announcement and legalization was due to the intervening parliamentary elections and new government formation.

Six components were included, three had been retained but differently weighted (accreditation, casemix index, patient satisfaction), two were new components (readmissions, elderly cases proportion), and one had been modified (ICU cases & beds). Component weights had been in part determined within the MoPH technical

team, using an analytic hierarchical process and further refined using an iterative process with pre-defined rules for balancing components.

Table 2: Components of the 2018/2019 P4P

#	Component	Weight
1	Accreditation	30%
2	Casemix index	45%
3	Patient satisfaction	20%
4	Readmissions	2%
5	ICU cases & beds	2%
6	Elderly cases proportion	1%

Hospitals that had accreditation status received a complete score on this component, while those that had not, received a partial score. This differed from the 2014 process, which had penalized hospitals with a ‘caution’ note on their most recent accreditation report.

The casemix index was calculated separately for medical, surgical and mixed cases, and the first was further segregated into short, medium and long-stay medical cases. Casemix results were weighted by case type volume to calculate the ‘all-stay’ casemix index.

The results of the most recent patient satisfaction survey were used for the third component. This included a composite score based on patient responses on eight items regarding the admission desk, doctor and nurse interactions, pain relief, dignity and respect, cleanliness, recommendation of hospital to others, and receiving a receipt after payment at the hospital.

The readmission component included four conditions: general cases, pneumonia, cholecystectomy and stroke. Their selection was based on review of other initiatives, and their incidence, trends and utility in using these within the Lebanese context. Case definitions were developed, specifying the calculation, and inclusion and exclusion criteria. Within each readmission measure, hospitals within +0.5 standard deviations received the full score (0.5%), while those above received none. This design took into account variation around the mean, while also providing incentives for the reduction of readmissions.

The elderly cases proportion measured adults aged 64 years and older, among all adults (18 years and older). The intention behind this component was two-fold. Firstly, this would compensate for the exclusion of age-adjustment in casemix

calculation. Secondly, it was intended to discourage cherry-picking by some hospitals which may avoid hospitalization of elderly persons.

The ICU component had been calculated in 2014 based on the proportion of ICU cases among all hospitalizations. This was revised in 2018 into two separate subcomponents: the proportion of ICU to total cases; and the proportion of ICU to total beds. This was motivated by the lack of a consensus in the literature regarding which approach is more effective at incentivizing hospitals to increase their ICU capacity. The scores of components #3, #5, and #6 were standardized and capped within two standard deviations of the mean.

A similar approach was used for determining the hospital TPS as in 2014. However, a -0.75 standard deviation was used as the cut-off between T2 and T3 reimbursement tiers. Figure 7 illustrates an example of the performance card disseminated to each hospital in 2019.

Four domains have been used to describe the structure of US Medicare programs: program scope (broad or narrow), performance (absolute or relative), awards (achievement, improvement or both), and incentive (reward, penalty or both) ¹⁴⁸. Using a similar approach, the Lebanese P4P model may be described as having a mixed scope, and rewarding achievement based on relative hospital performance. The incentive for hospitals is to increase (or maintain) their reimbursement tier. Since the P4P was integrated within the MoPH mechanism for determining reimbursement tiers, and not relying on separate funds, it was by design sustainable. The Lebanese P4P shared certain aspects with the US and French P4P models, such as readmissions and certification/accreditation, respectively.

Conclusion

This chapter provided an introduction to Lebanon and its health system, and information about the MoPH P4P models of 2014 and 2018/2019.

In the next chapter we will provide the conceptual framework for the P4P model developed at the MoPH, and where the four papers included in this thesis fit within this framework.

GENERAL INFORMATION

Hospital Name: **Example Hospital**
Performance Report Year: **2019**

Period of dataset used: 2017 (CMI, ICU, elderly); 2018 (patient satisfaction); 2015 (accreditation)

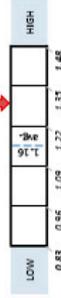
KEY PERFORMANCE INDICATORS

A. CASE-MIX INDEX (CMI)

- All cases



- Medical cases



- Surgical cases



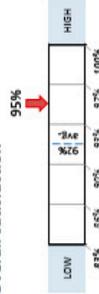
- Mixed cases



B. PATIENT SATISFACTION

Category	Hospital	National Average
Admission desk	90%	96%
Doctor interaction	94%	94%
Nurse interaction	94%	95%
Pain relief	93%	95%
Dignity and respect	88%	93%
Cleanliness	100%	93%
Recommendation	85%	90%
Receipt	96%	77%

- Overall satisfaction



C. ACCREDITATION

- Status at last accreditation round: **Accredited**

D. 30-DAY READMISSION RATE

Condition	Hospital	National Average (-2sd)	Low	High (+2sd)
Cholecystectomy	2.8%	3.8%	0%	8.4%
Stroke	5.1%	7.6%	0%	15.9%
Pneumonia	6.3%	5.1%	0%	11.3%
General (all)	5.1%	5.6%	2.2%	9.0%

E. ICU PROPORTION

- Case proportion (ICU/total cases): **12.7%**
- Beds proportion (ICU/total beds): **18.6%**

F. ELDERLY ADULT PROPORTION

- Elderly proportion of adult cases: **39.4%**

SCORING RESULT

Hospital score distance from national average (z-score)

0.763

New reimbursement category

T1

Figure 7: Hospital performance card example disseminated to each hospital in 2019.

4 Conceptual Framework

“Be kind to people, be ruthless to systems”

– Michael Brooks (1983-2020)

This chapter provides the conceptual framework for the P4P model developed at the MoPH, and where the four papers of this thesis fit within this framework.

Recognizing the complexity of pay-for-performance in health systems, we developed a conceptual framework for the model developed at the MoPH (see figure 8). Since the goals of this research are based on the MoPH P4P, we sought to clarify the various aspects involved. This framework includes several dimensions: the intervention, contextual and mediating factors, response expected and organization.

The intervention is the P4P model, which includes several characteristics that are relevant for impact. These include the size of the incentive and the weight of individual components, both of which would be expected to influence if and to what extent hospitals respond to the P4P. A model that is easier for hospitals to understand is also more likely to result in a response, as would one that has a more frequent evaluation (e.g. annually) compared to one that is less frequent. Transparency would be expected to increase trust among hospitals regarding the fairness of the model, and in other instances this involves having performance measures determined not by the payer, but by a third party. Sustainability is also relevant, as it signals to hospitals whether the P4P model is a transient initiative or a longer-term one. Because the total performance score of hospitals is a composite score, is also affected by the variation within each score. For example, if two components were equally weighed, but the first varied between 80-90%, and the second between 60-90%, then the second will have a greater effect on the TPS.

Internal factors include the hospital-perceived cost-benefit, resources available to them, how knowledgeable they are about the model and ways to respond effectively. The reimbursement tier may also be a factor: hospitals who perceive themselves near the top of the highest tier, or the bottom of the lowest tier, may not be as incentivized to respond to P4P as those who are closer to the tier boundaries. The alignment between the model and hospital values may also affect their response.

External factors include the political, financial and socioeconomic context. These may be relevant in different ways. For example, outputs or outcomes may be less

modifiable by hospitals serving a more socioeconomically disadvantaged population. Changes in healthcare regulations or other initiatives may also affect hospital P4P response.

Organizational factors include the types of services a hospital provides, and the relevance of the P4P components to these. The volume of cases a hospital receives may also be relevant, particularly in terms of patients covered by the MoPH. For example, if MoPH patients represent a small proportion of total hospital cases, a hospital may be less incentivized to respond to P4P. In another manner, due to the ‘law of large numbers’, we may expect more variability in outputs and outcomes among low-volume hospitals. Additional organizational factors relate to the location and ownership of the hospital (public/private).

The response expected and outputs/outcomes are based on the components of the 2018/2019 P4P, but include most of the 2014 components as well. We expect that hospitals would respond to the model components to different extents, possibly targeting all components or a selection of them. The mechanism by which hospitals would do this is labelled ‘black box’. This label is used because this represents one of the main gaps in the general evidence base on P4P, and the limited MoPH insight regarding such mechanisms. We also acknowledge the important role that unintended consequences may have in interventions such as P4P.

The reasons leading to the development of P4P, including the overall context and MoPH-hospitals relation, are the subject of Paper 1 (see figure 9). Decreasing unnecessary hospitalizations is measured vis-à-vis casemix changes and represents improved effectiveness and patient outcomes. This is the subject of Paper 2. Decreasing readmissions also represents improved effectiveness and patient outcomes and are the subject of Paper 3. Improved patient perspectives are another outcome; in Paper 4 we explore what patient perspectives are and how to relate these to P4P. We do not measure changes in patient perspectives, due to construct validity limitations regarding this topic, and other issues noted in 2.4.3.

Conclusion

This chapter provided the conceptual framework for the P4P model developed at the MoPH, and where the four papers of this thesis fit within this framework.

In the next chapter we will provide the theoretical foundation for the study design and analysis. This includes a discussion on causality, interrupted time series design, and the Newey-OLS and ARIMA mechanisms. We end with an introduction to the analytical approach and data collection method used in our qualitative investigation.

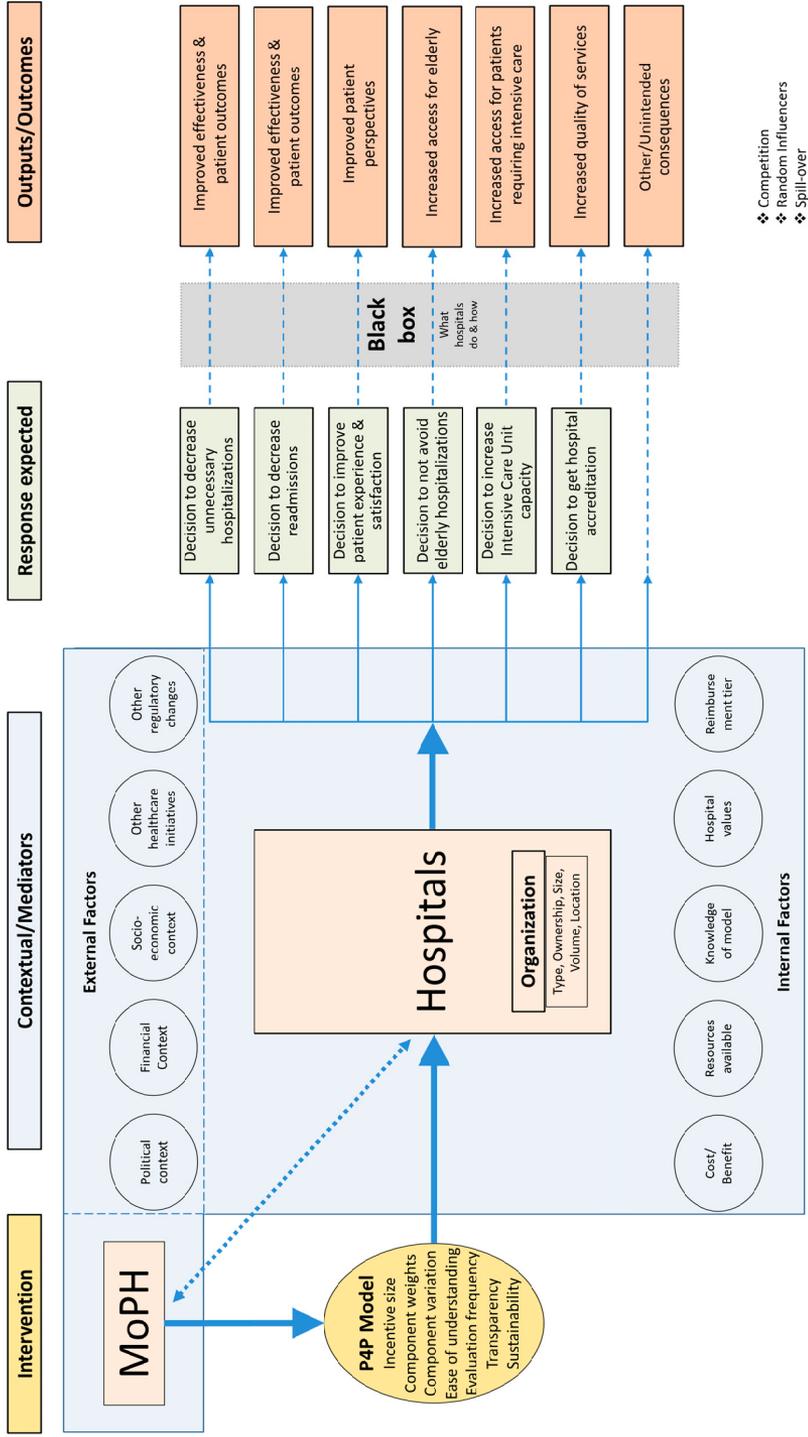


Figure 8: Conceptual framework for the P4P model developed at the MoPH .

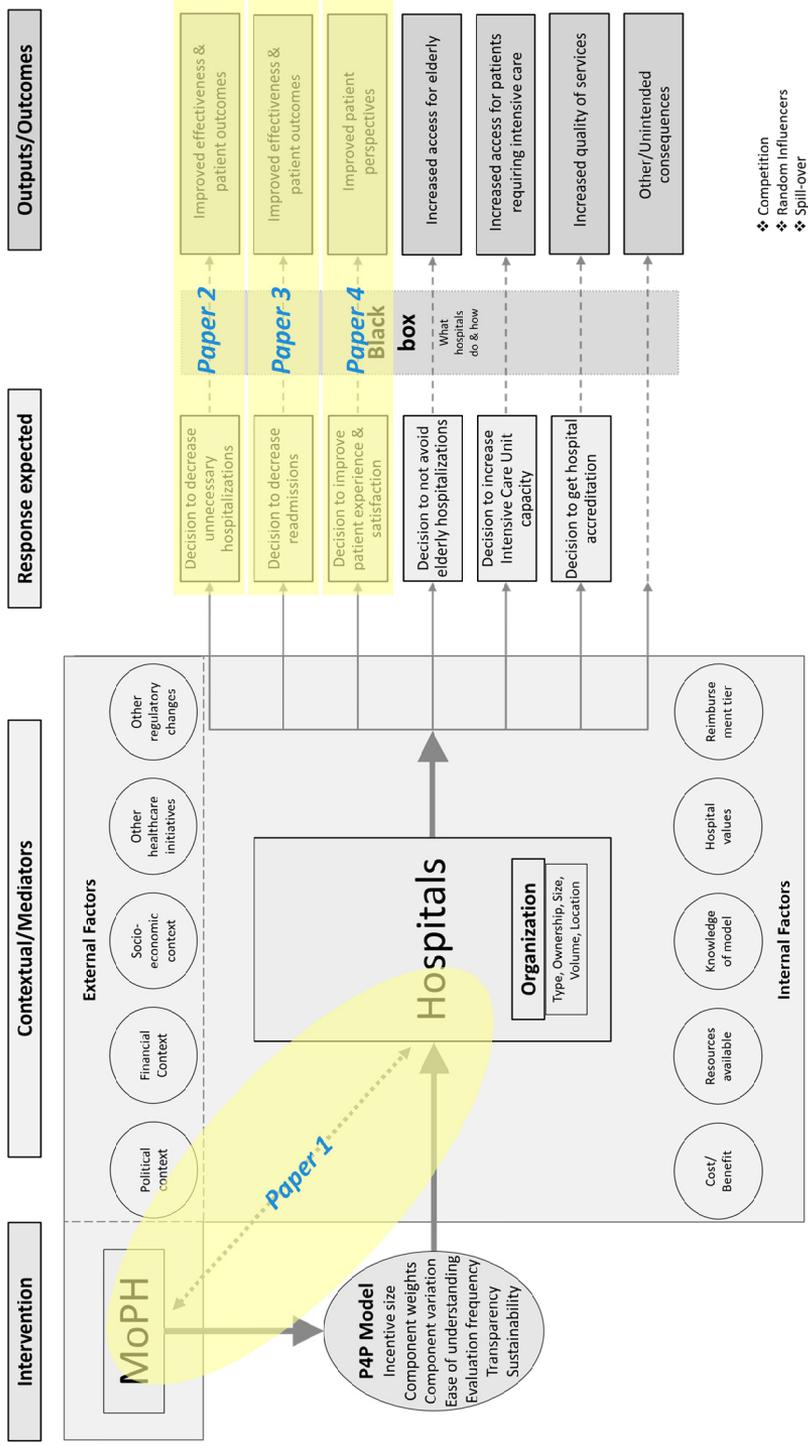


Figure 9: Highlighting the four papers of this thesis within the conceptual framework for the P4P model.

5 Theoretical Foundation for Study Design and Analysis

“Since the extensive use of randomized experiments is limited to the last half century, and in fact is not used in much scientific investigation today, one is led to the conclusion that most scientific ‘truths’ have been established without using randomized experiments.”

– Rubin, 1974, p.688, as cited in McCleary et al. (2017) ¹⁴⁹

This chapter begins with an introduction to the history of causality, how it is understood within some disciplines, and the role of experiments and quasi-experiments. We establish that quasi-experiments have the potential to provide causal inference. This is followed by an introduction to interrupted time series (ITS) design, its accompanying four validities, and different types of impact that can be investigated using ITS. We then describe the mechanism of using Newey-OLS and ARIMA for ITS. The last section presents the qualitative approach (content analysis) and the data collection method (focus group discussions; FGD) chosen for exploring patient perspectives.

5.1 On causality

From a health or epidemiological perspective, we pursue knowledge on causality to better understand how to prevent disease and improve well-being. But the history of thinking on causality is diverse, stemming from different roots, and contributing to different meanings of causality, depending on one’s discipline.

Aristotle defined four types of causes (material, formal, efficient, final), and was interested in the deeper philosophical purpose of objects or processes. David Hume developed a subjective and psychological construct of causality (as a response to the ‘induction problem’). John Stuart Mill approached this differently from Hume, arguing that causality is based on our empirical observations and is reliable in understanding the external world, despite limitations. These three examples represented the philosophical, natural and empirical traditions on causality, with the

third having the greater influence on modern scientific understanding of causality
150p.133-134

Numerous important contributions were made in the 20th century, including those of Ronald Fisher (randomization, confounding), Jerzy Neyman (counterfactuals) and Donald Rubin (potential outcomes framework for estimating causal effects). All three developed methodologies and statistical tools for drawing causal inferences using observational data. On experimental designs in particular, Donald Campbell developed the quasi-experimental framework for deriving causal inferences in real-world contexts or when randomization was not possible.

It is not uncommon for epidemiologists and other scientists to avoid using the term ‘causal’, since our approaches allow us to infer causality to varying degrees of confidence, but not prove it as a universal truth. The stronger the ability to rule out alternative explanations, the closer our approximation of the truth is. From a health perspective, for example, we cannot be fully certain that a specific exposure caused the disease in a person, because we do not know with certainty what would have happened had the exposure not occurred (i.e. the counterfactual).

It is useful to briefly illustrate the difference regarding causality in disciplines such as physics and chemistry, as opposed to epidemiology and health sciences. The first are mainly concerned with understanding fundamental laws and mechanisms at the microscopic level, using reductionist approaches and deterministic models. Contrary to this, the complexity of the systems studied in epidemiology and health sciences does not allow similar reductionist approaches, due to the variable factors involved (e.g. biological heterogeneity, ethics, social and environmental determinants). Therefore, making causal inferences in this latter category is more challenging. Physics and chemistry usually produce mechanistic evidence, while epidemiology and health sciences usually produce probabilistic evidence, though noting important exceptions (e.g. in quantum mechanics, in biology).

An instructive example of the different ‘causalities’ across disciplines is with the COVID-19 pandemic. The different interpretations contributed to the failure of many health institutions to recognize that the airborne spread of SARS-CoV-2 could be prevented using tools such as masking and ventilation. Such tools are applications of the physical and chemical sciences, and typically have national agencies that maintain regulated standards since decades ^{151 152}. Epidemiologic approaches are not equipped to disprove such scientific findings established using the physical and chemical sciences ¹⁵³⁻¹⁵⁷. Rather, their role can be to increase knowledge on how such tools can be scaled up and adopted in the complex real-world environment.

In epidemiology and health sciences

The modern approach and reasoning on causality within the health sciences has been built on important contributions of persons such as Ignaz Semmelweis, Louis Pasteur, Robert Koch and Jakob Henle. We say approach and reasoning, because there is no algorithm or checklist for inferring causality. However, there are some very useful guidelines. The most prominent of these has been the considerations developed by Austin Bradford Hill in 1965. These originally included strength of association, consistency, specificity, temporality, biological gradient, plausibility, coherence, experiment and analogy¹⁵⁸. Over the subsequent decades these have been debated and revised in some instances^{150p.167 159p.382-383}. However, it is important to note that Bradford Hill never intended these as ‘criteria’, but rather as considerations to guide decision-making. The consequence of action based on evidence may also play a role in determining the level of evidence that is sufficient¹⁵⁸.

“What I do not believe - and this has been suggested - is that we can usefully lay down some hard-and-fast rules of evidence that must be obeyed before we accept cause and effect. None of my nine viewpoints can bring indisputable evidence for or against the cause-and-effect hypothesis and none can be required as a sine qua non. What they can do, with greater or less strength, is to help us to make up our minds on the fundamental question - is there any other way of explaining the set of facts before us, is there any other answer equally, or more, likely than cause and effect?”

– Austin Bradford Hill (1965)¹⁵⁸

Causal inference and study design are interlinked but distinct issues^{150p.175}. In and of themselves, there is no design that confirms causality or cannot add to the evidence, but experimental designs tend to provide greater support^{150p.175}.

Many individuals have contributed to the development of our understanding of an ‘experiment’ vis-à-vis the scientific method. These include Epicurus, Ibn al-Haytham, Ibn Sina, Francis Bacon, Galileo and many others. Our modern definition of an experimental design draws heavily from the works of Ronald Fisher, particularly his emphasis on randomization, control and replication, which remain features of what is considered an experiment. A ‘quasi-experiment’ differs from this in that “the investigator lacks full control over the allocation and/or timing of intervention. [The] inability to allocate subjects randomly is a common situation that may be best described as a quasi-experiment”¹⁶⁰. This is not to be confused with a ‘natural experiment’, where investigators lack control of the independent variable, and may lack a control group.

There has been much emphasis on randomization in research, and specifically the role of randomized control trials (RCTs). Although these are sometimes considered the ‘gold standard’ for evidence, they are not. All study designs have their pitfalls and limitations^{150p.175}. Randomization is generally the best approach to strengthen the internal validity of an investigation, by minimizing confounding and selection bias, whether through participant self-selection or by researchers^{159p.383 161p.47-49}. As such, it intends to produce groups that are similar in baseline characteristics. Randomized designs are therefore powerful when studying phenomenon at the level of individual human body. However, there are many instances when randomization is not possible, not desirable, or both.

Firstly, the strengthening of internal validity comes as a trade-off with external validity (generalization). This limits the utility of such approaches at the population level. It is also used to study specific interventions in tightly controlled environments; such control is usually not possible in complex environments or real-world settings. Secondly, there are several ethical issues that should be considered before a randomized design is used. There should be a situation of equipoise, whereby there is uncertainty as to whether one intervention or the other is more beneficial to participants^{161p.50 162p.157}. Such uncertainty should be genuine and widespread, and not limited to the researchers alone. History records numerous instances when equipoise was absent, in unethical human experimentation such as the Tuskegee study (syphilis) and the Vipeholm study (dental caries). Thirdly, some interventions cannot be studied using randomized designs due to legal constraints and equity. For example, to study the impact of a law, we cannot randomize some people to be subject to the law, and others not^{161p.47-50}. RCTs also require considerable resources and are therefore not often feasible. Important threats to RCT validity also include the Hawthorne effect and Nocebo effect.

“Experiment: Occasionally it is possible to appeal to experimental, or semi-experimental, evidence. For example, because of an observed association some preventive action is taken. Does it in fact prevent? [...] Here the support for causation hypothesis may be revealed.”

– Austin Bradford Hill (1965)¹⁵⁸

Quantifying a cause-and-effect relation is not limited to randomized designs alone. Firstly, this is because causal inference remains an informed judgement, regardless of the tool used (even randomization). Secondly, focusing only on one methodological aspect risks ignoring other challenges to inferring causality^{163p.11}. The general approach is one that considers various uncertainties, and aims to exclude all alternative explanations, to finally express the varying level of confidence regarding the cause-effect being investigated^{163p.11}. The harms of tobacco smoking are illustrative of a widely established cause-and-effect across numerous diseases, which has been entirely based on observational studies, and

further complemented by laboratory findings. Although this is seen as settled by most people today, even if it were feasible to conduct an RCT on smoking today, it would be considered unethical ^{164p.58}.

Both Campbell and Rubin recognized that quasi-experimental designs can be used for making causal inferences. In such designs, unanticipated events can approximate the effects of randomization; when approximation is strong, the internal validity is strong ¹⁴⁹.

5.2 Interrupted time series analysis

Time series refers to a collection of data points ordered by time. If an intervention is included, it may be seen as an extended pre-post-test. However, the availability of many data points both before and after the intervention may allow, in some instances, for a cause-and-effect investigation.

We may define three types of time series designs: descriptive, correlational and (quasi-)experimental. Descriptive designs explore trends and cycles, but cannot be used for causal inference, because even entirely random processes may generate non-random patterns of change ^{149p.4}. Correlational designs can investigate causal relations between two series, using their covariance. However, these require a strong theoretical reasoning explaining the relation between the two series. Quasi-experimental time series, sometimes called ‘time series experiment’, refers to designs that involve data being ‘interrupted’ by an intervention, and are therefore named interrupted time series. This usually involves non-randomized designs, but can be used in randomized ones as well ^{161p.128}. Causal inference primarily relies on abrupt interruptions not being a feature of natural time series processes. Hence, an interruption that coincides with an intervention may be interpreted as evidence of cause-and-effect, provided that validities are maintained and alternative explanations are not plausible ^{149p.8}. In a broad manner, this approach has been used since the start of the 20th century ^{149p.7}.

5.2.1 The four validities

We may define four relevant types of validities, all of which should be met, to allow causal inference from a time series design. This section draws generally on the work of McCleary et al. (2017) ¹⁴⁹.

1. Internal validity

This refers to the extent to which a study’s design and implementation minimize alternative explanations, other than those that may be attributed to the manipulated

variable. In time series designs, the five plausible threats to internal validity are history, instrumentation, maturation, regression and selection. History concerns the possibility of another event that coincides in timing with the actual intervention. This highlights the importance of understanding the context of the underlying process, and excluding the presence of alternative explanations for any impact. Instrumentation refers to any change in measurement, whether in the form of tools or processes (e.g. changing standards in coding diseases).

Maturation refers to any underlying process, that may otherwise be responsible for observed changes. Failing to control for such trends may result in causality being wrongfully attributed to the intervention. Regression to the mean may be a threat to internal validity when an intervention is applied at a point when the underlying process is at either a considerably high level or low level. Since interventions are often applied as a response to deviation from desired levels, regression should be considered as an alternative explanation in such instances. Having stable and non-short time series data is usually sufficient to avoid this. Finally, selection is relevant in designs with an intervention and a control group, whereby they may differ on some theoretical grounds.

2. External validity

This refers to the extent to which results can be generalized, whether to other populations, contexts, treatments or outcomes. The threats to external validity are due to either variation in subjects and situations, or variation in timeframes. The main threat to external validity is how participants react to being observed (reactivity). Experimental trials almost always use blinding (e.g. with placebo) to address the threat of reactivity. However, in other designs this is not often possible. Another approach is making use of secondary data, which was not originally collected for the purpose of the research. This is one advantage of secondary over primary data, in that we expect reactivity to be minimal or absent. The influence of reactivity may be considerable, also in healthcare. When data is used to evaluate people generating or collecting it, the data is subject to be altered ('Campbell's law of data corruption')¹⁶⁵.

Temporal drift is a threat that may occur when social or behavioral features of a population changes over periods of time. Semantic drift is one form, which may occur when the meaning of specific words or concepts changes over long periods of time. This threat is not usually encountered within the timeframes of most time series designs. Another form may be from short-term temporal variation, for example, weekday to weekend variation in speeding fines.

It would be preferable to control threats to external validity using study design. However, this is not possible in many investigations. The remaining alternative is replications, which is costly, but also necessary.

3. Statistical conclusion validity

This refers to soundness of the statistical analysis and interpretation. For time series design, the first threat to this validity is from low statistical power. Statistical power is a function of N , which is usually large in time series data, provided it is well-designed. There is no standard reference for what qualifies as too few. Some refer to at least six data points in each of pre and post-intervention periods. However, it is reasonable to assume we need several data points in both periods, and still more if there are seasonal or cyclical patterns in the underlying data. Overall, if the underlying process presents a generally stable trend, we can expect that at least 30 data points would be sufficient for many causal inferences^{149p.259}. With less stable time series, a greater number would be required. More than 100 data points may be considered a long time series; however, series that are too long may be detrimental (see construct validity).

Besides time series length, two other factors to consider are forecasting abuses and missing data. From a practical perspective, forecasting approaches may be appealing, however, they are not more advantageous than other approaches, and have a difficulty in distinguishing abrupt-permanent changes from others^{149p.261}. Time series often contain missing data, however, this may be imputed or analyzed with no disadvantage, provided that this data belongs to a single time series, is contiguous and not near the intervention^{149p.262}.

The second threat is from violated assumptions of statistical tests. We have to ensure the normality of the time series data used, since this is an integral assumption of the statistical testing involved. If normality is unmet, this may be addressed using transformations, otherwise this would lead to biased estimates and incorrect inferences. Outliers may also compromise normality and therefore should be assessed, although we may note that outliers are a phenomenon when using sample data, but not population data^{149p.258}.

4. Construct validity

This refers to whether the measures used truly capture the theoretical construct intended, i.e. they measure what they are supposed to measure. The timeframe for the experiment should be carefully considered. Although a stable time series allows us to incrementally increase statistical power with each new observation, beyond a certain point this would come at the expense of construct validity. That point is when the underlying process is under a different regime or treatment. Therefore, a time series that is too long would include factors we are unaware of or did not intend to include, and this would threaten any inference regarding the intervention.

A second threat is periodicity. The data in a time series may be aggregated at different levels, for example, daily, weekly, monthly or quarterly. Although this depends on the underlying process, using daily or weekly periods may result in a non-normal time series. Non-normality can be resolved using transformation, but

normal processes are preferable. Quarterly periods may result in normality, but also decrease the number of data points available. Monthly data points are usually more likely to provide both normality and sufficient periods, though it may be debatable if 28-day months are better than calendar months.

Two additional threats are ‘fuzzy’ onsets and ‘fuzzy’ responses. An example of the former is when an intervention is implemented at slightly different times among two or more subgroups of the target population. The latter may involve an intervention that alternates between active and inactive, or high and low activity. Such threats may be controlled by carefully understanding the underlying theory regarding the intervention and its impact. A final threat is alternative worldviews, relating to different motives or interests of parties collecting the same data (e.g. national health authority and private insurance).

To allow causal inference from a time series design, all four validities presented above should be met. Particularly internal validity may be the most consequential. This is because it refers to the confidence with which we can infer a causal relation, or the extent to which the observed effects are attributable to the intervention rather than other factors. What distinguishes time series designs is the reliance on a statistical model to control against some of the common threats to internal validity^{149p.10}. Therefore, these threats play a smaller role than those related to the three other validities, provided that the study is well-designed.

5.2.2 Impact models using interrupted time series design

Different types of impact may be encountered when using interrupted time series designs. Before analysis is conducted, it is important to clearly define the relation of the intervention to the outcome, as well as what type of impact is expected¹⁶⁶. An a priori theoretical basis for both these aspects strengthens the validity of our investigation. The type of impact may be defined using the literature, past experience or logical argument based on the mechanism of change¹⁶⁶. If none of these are available, exploratory analysis of alternative data may be used¹⁶⁶. Generally, the impact may be abrupt or gradual, and temporary or permanent. Different variations of simple and complex impact may be encountered.

Figure 10 illustrates some of the more common types of impact, with changes in level and/or slope.

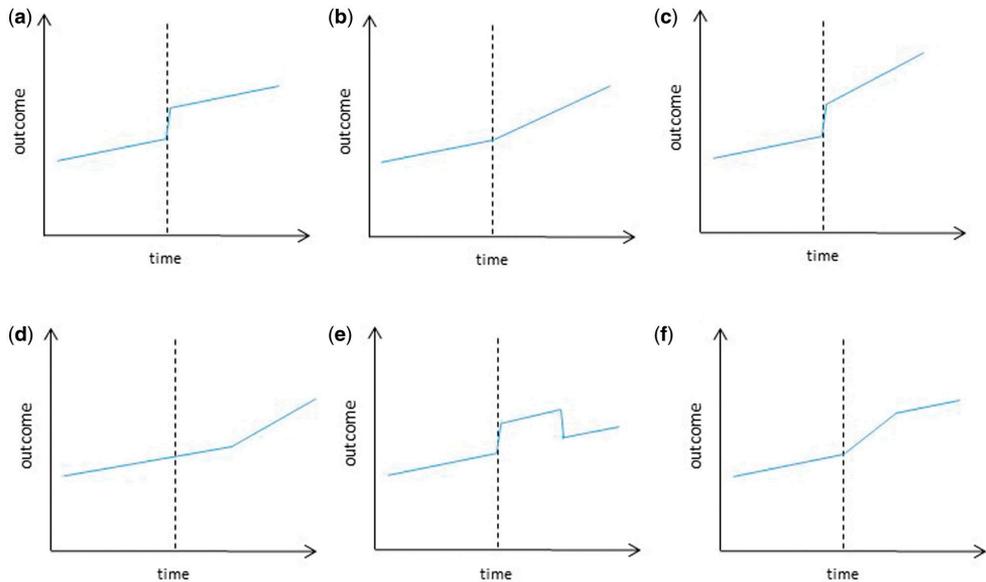


Figure 10: Examples of impact models used in ITS ¹⁶⁶

(a) Level change (b) Slope change (c) Level and slope change (d) Slope change following a lag (e) Temporary level change or pulse change (f) Temporary slope change leading to a level change.

5.2.3 Newey ordinary least-squares regression (Newey-OLS)

The method of Ordinary Least Squares (OLS) is a widely used approach in linear regression models. It aims to find the line of best fit between data points, based on which gives the smallest sum of squared differences between the observed and predicted values (i.e. sum of squared ‘residuals’). Newey-OLS is a type of OLS that uses the Newey-West estimator, which corrects for autocorrelation and heteroscedasticity, to improve the OLS model. Newey-OLS uses heteroscedasticity and autocorrelation consistent (HAC) estimation, using weighted averages of the residuals, which decrease over time.

Autocorrelation refers to when values in a time series at a particular point are related to values at a previous point. Addressing autocorrelation in time series analysis is important because its presence violates assumptions of independence among observations, which is used in different statistical tests. Heteroscedasticity refers to when residuals in a regression model have a variable variance (spread), as opposed to having the same variance (homoscedasticity). Left unaddressed, heteroscedasticity may result in regression models having biased coefficients of the

standard errors in the regression model, and thereby any inferences made. The use of Newey-OLS rather than classic OLS therefore provides less biased and more accurate results.

5.2.4 Autoregressive Integrated Moving Average (ARIMA)

The underlying mechanism of ARIMA models is to predict future values in a time series, based on past values, changes between past values and discrepancies between observed and predicted values in the past. The latter two components denote differenced values and random error terms, respectively.

George Box and Gwilym Jenkins are credited with the development of ARIMA models in the 1970s, building on previous work by Peter Whittle (1951). There has been increasing use of such models over the past few decades, with some minor development of the analytical approaches and increasing emphasis on model usefulness and interpretability. There has also been increasing use of ARIMA models in public health over the past several years, alongside other ITS models¹⁶⁷.

Three components comprise the ARIMA model. An AutoRegressive (*AR*) component represents the relation of an outcome variable to its own past values and a stochastic term (randomness). A Moving Average (*MA*) component represents the influence of past error terms on the current error term (not to a moving average of the outcome variable). An Integrated (*I*) component indicates a transformation (integration) function which is only used if we need to normalize the time series. In essence, an *AR* process remembers past realizations, while an *MA* process remembers past shocks. If the time series data has a cyclical or seasonal pattern, this may be incorporated to develop a seasonal ARIMA model (also called SARIMA). Values at the defined seasonal lags would therefore be included in the process. Such a seasonal component results in a powerful model which is able to capture seasonal dynamics and relations. We will use an example of monthly rainfall as a simple illustration of this process.

We would like to use previously recorded rainfall data spanning several years to forecast future rainfall, using an ARIMA model. The *AR* component would look at how the rainfall in the past affects the rainfall today. If it was rainy last month, it would be more likely to be rainy this month. From a first look at the data, if we detect that a long-term or seasonal trend was occurring (e.g. increasing rainfall), then we can use the *I* component to filter out this trend, and allow us to focus on the short-term changes instead. The *MA* component will look at the average rainfall and how it relates to previous miss-matches between our predicted and actual rainfall; if we had been over or underestimating rainfall, it would help adjust our model. Of course, real weather prediction is intrinsically far more complex, and uses other specialized models.

The basic equation for a seasonal ARIMA model may be expressed as follows.

$$Y_t = c + \sum_{i=1}^p \phi_i Y_{t-i} + \sum_{j=1}^q \theta_j e_{t-j} + \underbrace{\sum_{s=1}^S \Phi_s Y_{t-s} + \sum_{s=1}^S \Theta_s e_{t-s}}_{\text{Seasonal component}} + e_t$$

Equation 2

Here, Y_t represents the time series data at time t ; c is a constant term for the intercept; ϕ_i are the *AR* coefficients for non-seasonal lags $i=1$ to p ; Y_{t-i} is the value of the time series at lag i ; θ_j are the *MA* coefficients for non-seasonal lags $j=1$ to q ; e_{t-j} is the error term at lag j (difference between predicted and observed values); S represents the seasonality period (e.g. 12 months); Φ are the seasonal *AR* coefficients at seasonal lags $s=1$ to S ; Y_{t-s} is the value of the time series at seasonal lag s ; Θ_s are the seasonal *MA* coefficients for seasonal lags $s=1$ to S ; e_{t-s} is the error term at seasonal lag s ; and e_t is the error term at the current time t .

A pre-condition for running ARIMA models is that the time series data be stationary. Stationarity denotes that the mean, variance and co-variance of the values do not change over time. This is because the analytical mechanism requires that a time series process operate identically in the future as it has in the past. If the data is non-stationary, then we can use the *I* component to difference the data until it is stationary. Otherwise, using non-stationary data would threaten statistical conclusion validity and therefore any inferences drawn from our analysis.

ARIMA models should also be examined for autocorrelation and partial autocorrelation. Partial autocorrelation denotes correlation between a point and its previous values (at specific lags), excluding the influence of intermediate values. The autocorrelation function is used to determine the required *MA* order, and the partial autocorrelation function to determine the required *AR* order.

5.3 Content analysis and focus group discussions

5.3.1 Qualitative content analysis

“In qualitative content analysis interpretation involves a balancing act. On one hand, it is impossible and undesirable for the researcher not to add a particular perspective to the phenomena under study. On the other hand, the researcher must ‘let the text talk’ and not impute meaning that is not there.”

- Graneheim and Lundman (2004) ¹⁶⁸

Content analysis is an analytical approach which involves an investigation of text, images or symbols, not necessarily from the author’s or reader’s perspective ^{169p.10}. In some regions, such systematic analysis dates back at least as early as the 17th century and church investigations of content in written newspapers ^{169p.10}. It is probable that the first well-documented content analysis was in 18th century Sweden ^{169p.10}. With the advent of mass media in the late 19th and early 20th century, much of the focus was on quantitative analysis of newspapers. It was also used in political and propaganda initiatives during the Second World War. However, some limitations of quantitative content analysis became increasingly apparent, for example the potential to provide shallow or insensitive information ^{169p.16}. This drove the interest in developing qualitative approaches that could be systematic, valid and reliable ^{169p.16}. In subsequent decades, qualitative content analysis was applied by researchers in different disciplines, including psychology, history and anthropology. Various approaches have developed in both quantitative and qualitative content analysis, and these are widely used today for different applications. Examples within the latter include discourse analysis, rhetorical analysis, social constructivist analysis and conversation analysis.

The early emphasis of content analysis was on establishing meaning at the manifest (or evident) level. This was at least in part motivated by the desire to ensure validity and reliability, through using an objective and systematic approach and avoiding ‘reading between the lines’ ^{169p.25}. Later developments saw content analysis also including the latent (or underlying) meaning, which denotes a greater depth of understanding and level of abstraction ¹⁷⁰. Some of the reasons for this change were recognition that the content need not have single meanings, is influenced by the reader’s own understanding, and communicates something beyond the actual content alone ^{169p.27-29}. One may argue that increased inference by the researcher threatens the external validity of findings. However, it is important to recall that qualitative investigation pursues having a nuanced, contextualized and in-depth understanding, rather than generalization to other contexts. It is rooted in a

constructivist view of science, which views knowledge as something that is constructed through the interaction between participants and researchers.

There is a diversity of approaches to content analysis by social scientists today. This reflects different historical perspectives and researchers' diverse perspectives on the nature of reality¹⁶⁸. Graneheim and Lundman (2004) present one such approach, which has been widely utilized in health research. This approach adopts an underlying assumption that "[...] reality can be interpreted in various ways and the understanding is dependent on subjective interpretation"¹⁶⁸. In this approach, the unit of analysis is the entire interview or transcript, and the meaning unit is words, sentences or paragraphs which are connected through their content and context. Following a shortening process of the content (condensation), grouping under higher order headings is done, creating codes, categories and themes at different levels.

As suggested by Graneheim and Lundman (2004), codes are used as a heuristic to support the thinking process. Content areas are assigned to reflect explicit areas of the content, generally at a manifest level. Categories aggregates content in a manner that may be described as internally homogeneous and externally heterogeneous, although this may not always be possible when dealing with experiences^{168 171}. These are also considered as an expression of the manifest level. Themes are then developed, reflecting the latent level, which may be seen as threads of an underlying meaning, running through meaning units, codes and categories¹⁶⁸.

Qualitative content analysis differs from other commonly used approaches such as phenomenology and grounded theory. Phenomenology focuses on understanding the essence of the lived experiences of participants, without using preconceived categories, and is more distanced from the researcher than content analysis^{172p.18}. Grounded theory is an inductive approach which aims at generating theories or conceptual frameworks. It also avoids using preconceived frameworks, and is 'grounded' in the data, allowing theories to be developed from observations^{173p.4}.

5.3.2 Focus group discussions

Focus group discussions are a specific form of data collection and have a long history in research. In the late 1930s and 1940s social scientists were investigating approaches for nondirective interviews, to reduce the role of the interviewer, and shift more attention towards the participant. During the second world war, increased interviewing in group formats was done, mainly to increase military morale^{174p.7}. This subsequently got popularized over the following three decades particularly in market research, which sought to understand the thinking of consumers^{174p.8}. In the 1980s there was a resurgence in the use of focus group discussions among social scientists, which become more prominent in the 1990s^{175p.5}.

Participating in group discussions plays a role in forming our attitudes and perceptions across various issues. A focus group discussion resembles everyday social interactions far more than interviews. Creating a permissive and non-threatening group environment may allow participants to share information that would not be accessible through questioning. The aim is to encourage self-disclosure among focus group participants. In the selection of participants, a greater emphasis is placed on commonality rather than diversity ^{174p.14}. The interest is not to drive at a consensus or agreement on certain points, but rather to gather a range of perceptions or attitudes, regarding an event, experience or topic. Within such a format, it is expected that participants would also influence each other and motivate further discussion among the group ^{175p.6}. The discussions are carried out in a generally predetermined manner and using open-ended questions.

“I submit to you that there is no unanimity of goals or practice in these groups. Nor is there a uniform similarity in educational background among the moderators of these groups. The reason is that the focus group is to qualitative research what analysis of variance is to quantitative research. The technique is robust, hardy, and can be twisted a bit and still yield useful and significant results.”

– Gerald Linda (1982), as cited in Krueger (1994) ^{174p.20}

Conclusion

This chapter provided an introduction to the history of causality, how it is understood within some disciplines, and the role of experiments and quasi-experiments. This chapter included an introduction to interrupted time series design and the four most relevant validities, and the mechanisms of Newey-OLS and ARIMA for ITS. We also included an introduction to content analysis and focus group discussions, which were used in our qualitative investigation on patient perspectives.

In the next chapter we will provide the methods and materials used in this thesis, including study design, data collection and preparation, data analysis and ethical considerations.

6 Methods and Materials

“All scientific work is incomplete – whether it be observational or experimental. All scientific work is liable to be upset or modified by advancing knowledge. That does not confer upon us a freedom to ignore the knowledge we already have, or to postpone the action that it appears to demand at a given time. Who knows, asked Robert Browning, but the world may end tonight? True, but on available evidence most of us make ready to commute on the 8.30 next day.”

– Austin Bradford Hill (1965)¹⁵⁸

This chapter provides the overall approach (mixed methods) used in this thesis, including both qualitative and quantitative study designs. It also includes details on data collection and preparation, data analysis and ethical considerations. We follow a breakdown by paper within each section.

6.1 Study design

This thesis uses a mixed methods approach, combining quantitative and qualitative traditions to research¹⁷⁶. Recognizing the complexity of P4P development, implementation and impact, we choose this approach to enable us to address research questions on different aspects of P4P. Mixed methods approaches are problem-centered, with the research problem strongly guiding the research design^{176p.165}. This ‘mixed methods way of thinking’ does not rely on one particular theoretical framework or philosophy, although a pragmatist position is common^{176p.168}.

The main aspects of pragmatism are:^{176p.169 177}

1. Recognition of both the natural/physical and social/psychological worlds.
2. Knowledge is both understood and based on the reality of the world we experience and live in.
3. Theories are valuable as tools, and are true to the extent they apply in particular circumstances.
4. Action is emphasized over philosophizing.

Table 3: Overview of the methods used in the papers of this thesis.

#	Goals	Design	Data collection	Participants	Period	Main analysis
1	Describe how and why hospital P4P was developed in Lebanon.	Observational and primarily qualitative.	Project documents, discussions with key personnel	Key personnel involved in ESPISP-2 project	2009-2014	Descriptive analysis
2	Analyze the impact of P4P integration on healthcare effectiveness in Lebanon; <i>And</i> Describe how routine data and casemix may be used for hospital performance.	Quasi-experimental, retrospective cohort, ITS.	MoPH hospitalization database	1,353,025 hospitalized cases	2011-2016	ITS analysis using Newey-OLS regression
3	Analyze the impact of P4P on hospital readmissions in Lebanon.	Quasi-experimental, retrospective cohort, ITS.	MoPH hospitalization database	1,333,691 hospitalized cases	2011-2019	ITS analysis using ARIMA
4	Explore patient perspectives on hospital care in Lebanon, and contribute insights that may improve P4P design and effectiveness.	Qualitative, cross-sectional.	Eight focus group discussions	42 persons previously hospitalized during the preceding 3 months.	2017	Qualitative content analysis

The first paper sought to describe how and why hospital P4P was developed in Lebanon. This paper documented the rationale and process behind the P4P initiative, including the overall goals that the MoPH had pursued during 2009-2014. This paper involved an observational approach with a primarily qualitative focus, and some limited quantitative information.

The second paper assessed the impact of the 2014 P4P, specifically on the complexity of the average hospitalized case (casemix). It also describes how routine data and casemix can be used for hospital performance in limited resource settings. The approach used a quasi-experimental interrupted time series design, including a retrospective cohort of 1,353,025 hospitalizations during 2011-2016. Newey-OLS regression was used within an intervention impact model.

The third paper sought to analyze the impact of the 2018 events held by the MoPH, which announced the inclusion of readmissions into the new P4P model. This used a quasi-experimental interrupted time series design, with a retrospective cohort of 1,333,691 hospitalizations during 2011-2019. An autoregressive integrate moving average model was used to assess the impact across four types of readmissions.

The fourth paper explored patient perspectives on hospital care in Lebanon. The purpose included gaining insight that may help improve P4P design and effectiveness, as well as to contribute to the knowledge on engaging patients towards person or people-centered health systems. A qualitative approach was used, involving 42 participants in eight focus group discussions held in 2017. A pile sorting exercise was also included, to increase our understanding of how patients prioritized different factors regarding their hospitalization. Content analysis based on Graneheim and Lundman (2004) was used to analyze the FGD data ¹⁶⁸.

The four papers address different research questions regarding the P4P initiative in Lebanon, which figure 9 broadly illustrates. Neither a concurrent nor sequential approach was planned for the four studies involved. However, the first paper documented the P4P context and contributed to the framing of the subsequent papers. The second and third papers both used an ITS design, but with considerably different analytical tools. The ITS design was necessitated by the nature of the P4P interventions, and the legal constraints on the MoPH to contract with all hospitals with the same framework. While the first paper described P4P from the perspective of key personnel involved in this process, the fourth paper explored patient perspectives, with the intention to subsequently contribute to P4P and the health system in general.

6.2 Data collection and preparation

6.2.1 Context of pay-for-performance

The idea of this paper arose as an opportunity to review and document the experience of the MoPH reforms throughout 2009-2014, a period during which the Ministry ran the Second Emergency Social Protection Implementation Support Project (ESPISP-2), as a grant from the World Bank. The project had formally ended in 2014, however, the activities were largely sustained by the MoPH with support of the WHO Lebanon country office. The MoPH was embarking on a new phase of its hospital performance initiative, having been awarded a grant from the Health Systems Research Initiative (HSRI). Documentation of the MoPH experience had been spread across numerous documents and reports, and none sufficiently captured the underlying rationale and process.

We first collected from project personnel all documentation at the MoPH regarding the ESPISP-2 project and related activities. These were reviewed and relevant information extracted and summarized. An extract was used from the MoPH hospitalization database, containing anonymized administrative data on all medical hospitalizations, but not those that were surgical or mixed hospitalizations.

Discussions were conducted with project participants including the chairs of all three committees (utilization review, performance contracting, admission criteria), the MoPH lead on hospital quality and reimbursement, and the Ministry Director-General.

6.2.2 Casemix and readmissions

The second and third papers both rely on data extracted from the MoPH hospitalization database. The database is maintained by the Information Technologies (IT) department at the Ministry. Extracts including administrative data of all medical, surgical and mixed hospitalizations were requested and received from the IT department, with patient identifier anonymized. The variables included unique admission record number, case identifier, hospital code, admission and discharge dates, length of stay, total charge, medical code on admission and on discharge (ICD-10), and surgical procedure code (CPT). For Paper 2, the dataset was comprised of medical and surgical cases between January 2011 and December 2016, and mixed cases between January 2012 and December 2016. For Paper 3, the dataset included medical and surgical cases between January 2011 and December 2019.

Both datasets were cleaned from cases with likely data entry errors, which totaled to about 10-20 per year (age>120 years, missing ICD-10 codes, or missing total charge). They were also subject to specific exclusion criteria, which differed between the two datasets. For Paper 2, we excluded all cases with chemotherapy ICD-10 codes (Z51.1 and Z51.2). Chemotherapy involves low-cost, short-stays, and in high volume. In the Paper 2 dataset, chemotherapy cases formed about 4.4% of total cases. Many hospitals had miss-coded chemotherapy into other cancer-related codes, and the MoPH had engaged with hospitals to improve this coding practice. Therefore, we anticipated chemotherapy codes would increase, and other cancer-related codes would decrease. Retaining chemotherapy codes would have had the effect of artificially decreasing casemix index. These were excluded, but we retained other cancer-related codes.

The Paper 3 dataset excluded the following conditions, which are highly unlikely to be related to the cause of readmission: cardiac catheterization, lithotripsy, renal dialysis, chemotherapy, radiotherapy, malignancy, obstetric cases, motor vehicle accidents, blood transfusion, palliative care and transfer/same-day cases.

Casemix index calculation

The casemix index calculation approach of the MoPH involves using ICD-10 and CPT codes directly, due to the absence of national DRGs (see 2.2.2) ¹⁴⁷. Among medical cases, which are fee-for-service, the weight-setting uses average costs per

code. We used five-year averages to set code weights. Low-volume medical conditions (<20 cases in 5 years) had their weights set at the overall medical case average. Weight-setting process was separated across short-stay (<2 days), medium-stay (2–15 days), and long-stay cases (>15 days)¹⁴⁷.

Since surgical CPT codes are capitated (fixed), the weights were set without using averaging. Eleven procedure codes were capped at a weight of 10.00, to limit the effect of outliers. The 2013 inflation update of procedure costs was controlled by using the same weights based on this update for the entire period (2011-2016). Both primary (main) and secondary procedure codes were included.

The casemix index was calculated using the formula below. This is a modification of the standard generic formula (see 2.2.2), in that the denominator adjustment component is not used. This is because we were interested in the casemix index across the total population, and not that of individual hospitals.

$$\text{CMI} = \frac{\sum(W_g N_{gn})}{\sum_g N_{gn}}$$

Equation 3

Here, W_g is the weight calculated for each ICD or CPT code; N_{gn} is the number of cases within each code in the total population.

Algorithms were developed using Stata software package (v.11) for calculating the monthly CMI at the national level, for each of the three medical case stays, surgical cases, and mixed cases. The discharge date was used to categorize records into calendar months. An all-stay medical CMI was developed by combining short, medium and long-stay CMI, using a volume-weighted approach. Medical and surgical CMI were calculated.

Readmissions calculation

Case definitions were developed for each of the four types of readmissions: general, pneumonia, cholecystectomy and stroke. This process was informed by a review of the literature and organizational resources such as from the US CMS, with consideration for the preferred attributes of publicly reported outcomes^{178 179}.

A readmission was identified as a patient having been readmitted within 30 days from previous hospitalization discharge. This was regardless of readmission cause, unless otherwise specified by exclusion criteria (see 5.5.2), and thus is considered an all-cause readmission approach (see 2.3.2). General cases included all medical and surgical hospitalizations, but not mixed cases. Patients with multiple readmissions per year had only the first readmission counted as such, to limit the effect of outliers due to patients with high comorbidities.

Algorithms were developed using Stata software package (v.16), to calculate monthly age-adjusted readmissions rates across the four readmission measures. For general and pneumonia readmissions, this was also done separately across hospital sizes, with small (<50 beds), medium (50-100 beds) and large hospitals (>100 beds). Since the data was provided with an annual timeframe by the MoPH, admissions in December had been artificially limited to an end-of-year discharge date. The calculation of monthly readmissions was therefore made for 11 months per year (January to November), including December cases only for identifying the readmissions, but not the index admission (i.e. in the numerator, not the denominator). This allowed the time opportunity for November cases to be readmitted in December, and in line with preferred attributes¹⁷⁹. This resulted in 99 monthly data points, comprised of 80 in the pre-intervention period and 19 in the post-intervention period (after March 2018). To improve interpretability, direct linear interpolation was used to populate the values for the month of December between 2011 and 2018.

To account for the potential confounding effect of age changes in the population hospitalized under MoPH coverage, we applied direct adjustment on the calculated crude readmission rates. The 2015 denominator population was used as the standard reference for readmission data across 2011-2019. Each month of 2015 was used as the standard population for the corresponding month of other years. We used this approach since we expect the proportion of ages to vary across different months according to seasonal disease patterns. Six age groups were used: 0-5 years; 6-20 years; 21-40 years; 41-60 years; 61-80 years; and ≥ 81 years. In effect, the monthly age-adjusted rate was the weighted average of the age-specific (crude) rates.

6.2.3 Patient perspectives

The main motivation for pursuing this paper was to increase our understanding of patient perspectives and use this knowledge to improve how hospital P4P was designed and implemented. We say patient ‘perspectives’, but this was a term selected after our preliminary review on the literature on patient ‘satisfaction’ and ‘experience’ in 2016. An early attempt at a conceptual framework for patient satisfaction was also completed. However, this alongside the focus on ‘satisfaction’ was abandoned, in favor of the more comprehensive patient ‘perspectives’. This was largely motivated by the findings elaborated in section 2.4.2, particularly the lack of conceptual clarity on ‘satisfaction’ and the limitations of expectation theories.

Based on our reading of the literature and further discussions within the research team, we first developed the themes we were interested in exploring using focus group discussions. These themes were: the meaning of health; description of local healthcare; characteristics of services received; description of ‘good’ and of ‘bad’ hospital stays; information needed upon hospital admission; and factors that would

make a person revisit a hospital. We used these to develop open-ended questions within a discussion guide. The guide was piloted in one FGD with eight participants (men). Based on the pilot, we revised some questions and terms to increase their clarity. We had considered conducting a second pilot (women), however, based on our experience with the first pilot, we deemed this to be redundant.

Participants were drawn at random from the MoPH hospitalization database, among persons discharged within the preceding three months, for both the pilot and subsequent eight FGDs. We limited pilot participants to the Beirut region, out of consideration for participants' travel time from other regions. We had used a random sample since we lacked specific knowledge about participants and considered that such an approach would allow a greater variation among participants. Research assistants contacted potential participants by telephone, and recruitment continued until we had a broad variation of men and women across a range of ages, residing in different regions, to participate in the eight FGDs.

Our final sample included 42 participants (22 men, 20 women), with a median age of 49 years (range: 25-65 years). About 21% of contacted persons had agreed and participated in the FGDs. Five discussions were held in mid-July 2017, and three in early September of that year. The main reasons for declining participation were vacationing or having other engagements. We had intended to have 6-8 participants in each FGD, however, group sizes varied due to some late cancellations and rescheduling by participants. Participants were offered two-way complimentary taxi transport and snacks, but no material or financial compensation.

All discussions were held in a private room at the MoPH headquarters in Beirut. The facilitator (author) introduced himself as being involved in the research project on P4P development and evaluation, as a collaboration between Lund University, the American University of Beirut and the MoPH. Two research assistants (Abeer Al Halabi and Elise Barakat) noted the group layout and interactions, but did not participate in the discussions. The facilitator and research assistants met after each FGD to exchange feedback and notes, which were subsequently used to inform the next discussion and subsequent analysis. The median duration of FGDs was 62 minutes (range: 37-82 minutes).

After discussions ended, a pile sorting exercise was conducted. This is a tool used since the 1960s to investigate how people classify items and may be considered a combination of quantitative and qualitative techniques¹⁸⁰⁻¹⁸². This may include asking participants to sort statements into two piles of 'more' or 'less' important, which we used in this exercise¹⁸³. We had developed 16 statements, based on our assumptions of factors related to patient satisfaction and experience, and on topics addressed in existing literature and survey tools, including the US HCAHPS^{184 185}. Statements dealt with topics such as importance of regular contact with personnel, hospital organization, cleanliness, communication, pain and privacy.

6.3 Data analysis

6.3.1 Context of pay-for-performance: descriptive analysis

We used descriptive analysis to detail the experience of the MoPH reforms between 2009 and 2014, as well as why and how the MoPH moved towards hospital P4P. Using the documentation available, we developed a concise description of the history of MoPH hospital contracting reforms and the challenges faced by the health system. We elaborated on the goals of the MoPH through the ESPISP-2 project, which was framed under the themes of the three committees: utilization review, performance contracting and admission criteria. Iterative discussions with key project personnel were used to confirm and where necessary to correct findings, as well as to increase understanding regarding the overall rationale and the approach of the MoPH in engaging stakeholders in this process. Descriptive statistics were drawn on hospitalization frequency, length of stay and cost data, using a dataset drawn from the MoPH hospitalization database on the period from June 2011 to June 2015.

6.3.2 Casemix index: ITS and code-level analysis

The analytical approach of this paper was two-fold. The first was to estimate the impact on casemix index of different hospitalization types across public and private hospitals. The second was to explain these changes in CMI, by investigating changes in medical diagnoses and surgical procedures.

In the first part of our approach, we applied an interrupted time series analysis with historical control and seasonality adjustment, using Newey-OLS regression on the monthly casemix index between 2011 and 2016. This was done separately for surgical cases and then medical cases of different stays: short, medium, long and all-stay. The same was applied starting from 2012 for the medical and surgical components of mixed cases, and both combined. This was undertaken for all hospitals, and repeated among public and private hospitals separately, for a total of 24 ITS models. We had a total of 72 monthly data points for medical and surgical CMI, and 60 data points for mixed CMI. Based on the data points available, this may be qualified as a medium-length time series.

The intervention point was given a two-month lag for expected response, which was set at October 2014. The response time was chosen based on discussions with MoPH personnel most familiar with hospital response times regarding hospitalization and coding (Dr. Jihad Makouk). Sensitivity analysis was undertaken by varying the intervention-to-response lag between zero and four months. Statistical significance was set at $p < 0.05$ for all tests. Cumby-Huizinga test was used to assess

autocorrelation, to ensure adequacy of our ITS models. Where autocorrelation was detected, it had been accounted for by the set lag period.

The second part of our approach sought to explain the changes detected in the aforementioned ITS analysis, among medical and surgical cases. We used a pre-post comparison of changes at code level, but in terms of how much these changes explained CMI findings from ITS analysis. Since this involved 14,469 different codes, we used an annual approach to divide the pre-intervention period as 2013-2014, and post-intervention period as 2015-2016. An algorithm was used to calculate the change in CMI, and attribute changes to specific codes. This involved iterative process, with subtraction of a code and CMI calculation, and this process being repeated for each code. For each case type, we choose the top twenty codes that had the greatest effect on CMI in any direction (positive or negative).

The codes which had the greatest impact on overall CMI were identified as a function of code weight in relation to the overall CMI ‘average’, the code weight itself, and frequency. The codes with the greatest change in terms of their share of CMI are identified as a function of code weight and frequency only.

$$WN_0 = (W_g N_{gn})^{2013} + (W_g N_{gn})^{2014} \quad \text{Equation 4}$$

$$WN_1 = (W_g N_{gn})^{2015} + (W_g N_{gn})^{2016} \quad \text{Equation 5}$$

$$\text{Code count effect (CCE)} = ((WN_1 - WN_0) * (W_g - CMI_{ref}))^2 \quad \text{Equation 6}$$

$$\text{Code attributable change (\%)} = \frac{CCE_g}{\sum CCE} * 100 \quad \text{Equation 7}$$

Where W_g is the weight calculated for each ICD or CPT, and N_{gn} is the number of cases within each ICD in the total population.

CMI *share* change formula:

$$\text{Code share change} = \frac{WN_1 - WN_0}{WN_0} \quad \text{Equation 8}$$

6.3.3 Readmissions: ITS analysis

We applied an interrupted time series analysis on monthly readmissions, using seasonal Autoregressive Integrated Moving Average models, across 2011-2019. Ten ARIMA models were developed: one for each readmission type, and for general and pneumonia readmissions an additional three models each, by hospital size. The time series was medium-length, based on the 99 non-interpolated data points.

The expected response point was set at April 2018, following the two events in January and March 2018. Sensitivity analysis was undertaken by using February and March as response points. Statistical significance was set at $p < 0.05$ for all tests. An iterative identify-estimate-diagnose process was used, which involved evaluating several models before narrowing down on the final model^{149p.19}. The analytical process is detailed in table 1 in Paper 3 (see appendix). Although we anticipated an immediate impact of the intervention, resulting in a level change in readmission rates within weeks, we also investigated for a slope and/or pulse change. No other policies that may have affected readmission rates were identified.

We visualized the data by plotting monthly readmission rates and reviewed for potential trends and outliers. Using the pre-intervention data points, we assessed the stationarity of the series (heteroscedasticity), using the Breusch-Pagan/Cook-Weisberg test, and then the Dickey-Fuller test for unit root (non-stationarity). In the event of non-stationarity, differencing was used until the series was stationary. We then used autocorrelation and partial autocorrelation functions (ACF and PACF), to assess autocorrelation and stationarity, and to select the *AR* and/or *MA* terms of our final model. Seasonal ARIMA models were developed, using 12 months. Models were developed in an iterative manner, aiming for the most parsimonious model, through removing non-significant parameters and using the Bayesian information criterion (BIC). The most practically useful model was selected. Following this, intervention variables were generated for level, slope and pulse changes, and the data was visualized to support interpretation.

Model diagnostics were then run, to confirm that the assumptions necessary for our analytical approach were maintained. This involved comparing residuals to white noise, and tests of normality and independence. We plotted the residuals to assess for ACF, PACF and non-heteroscedasticity. Kernel density plot and standardized normal probability plots were used, followed by the Kolmogorov-Smirnov test. A scatter plot of residuals by time was used, followed by the Ljung-Box test. The model was considered statistically adequate once normality and independence criteria were confirmed. We did not use a forecasted-to-observed difference, to avoid the potential errors associated with this approach^{149p.167}. We chose to compare our analysis, using single-group ITS with Newey-OLS regression and seasonality adjustment (as used in Paper 2).

6.3.4 Patient perspectives: qualitative content analysis

Qualitative content analysis was chosen as our analytical tool, because we were interested in the content that would be gained from engaging patients, and to interpret these at the manifest and latent levels. Specific research questions had been developed, though we were open to other issues that may arise during the discussions. Our intent was not to focus on generating new theories regarding patient perspectives.

The recordings from the eight focus group discussions were concurrently transcribed and translated from Lebanese Arabic to English verbatim, by the two research assistants. The transcriptions were checked for accuracy and sense-making by the facilitator and research assistants, with corrections made where necessary.

We analyzed the transcripts based on the approach developed by Graneheim and Lundman (2004) ¹⁶⁸. All transcripts were read several times before coding was begun. Statements were not condensed, and were directly coded using NVivo software v.12.0. Statements were analyzed in relation to the specific research questions, which also formed the basis for the content areas. Related codes were used to construct categories based on the manifest and latent meaning. The latter was used to develop themes. Table 4 provides an example of this process.

Table 4: Example from the analytical process, moving from text to code and category.

Text	Code	Category
You have nothing even if you don't have health, even if you own the whole world. (FGD1-P5)	Without health we have nothing	More important than money or wealth
Briefly, health is the whole life, who does not have good health, has nothing because the sick person is always depressed. (FGD3-P1)	Without health we have nothing	
Health is everything, if you have all the money in the world but you have poor health, it means you are poor and you own nothing. (FGD2-P3)	Health is more important than money	
At the end you give priority to health over other needs, this is how I think, for example I buy anything cheap, but I don't buy a cheap medicine to save money, and same for the doctor [...] a person should be frugal on everything except on his health, this is how I think [...]. (FGD3-P3)	Being frugal except with your health	

6.4 Ethical considerations

Ethical approval was sought and granted from the Institutional Review Board (IRB) at the American University of Beirut (ID: FHS.FE.21), for the research protocols of Papers 2 to 4. Paper 1 involved a retrospective review of project documents and discussions with project personnel (which included the author), and did not require ethical approval. Papers 2 and 3 had the requirement for patient consent waived by the IRB, as the datasets provided by the MoPH IT department to the project team were in a format with patient identifiers anonymized.

Oral patient consent was provided by all focus group discussion participants regarding Paper 4, in compliance with IRB requirements. All participants were initially contacted by telephone and provided with oral information in the invitation to participate, and this information was repeated on-site prior to the start of each FGD. This included the study's purpose, fully voluntary nature of participation, the right to refuse, and that the decision to participate and any information shared would not be associated with or affect their MoPH coverage. They were also assured that all material would be handled confidentially and no results would be presented that may be used to identify participants. Participants were given the choice of whether or not to allow the researchers to use a digital audio-recorder (all accepted). They were also asked to respect that what is said in the group stays in the group, as researchers could only ensure confidentiality on behalf of the research team.

All research team members had to undergo the Collaborative Institutional Training Certificate training program for 15 modules regarding Social and Behavioral Research curriculum, and obtain the score necessary to pass certification (80%)¹⁸⁶. The research included in this thesis was consistent with the Belmont Report and the Declaration of Helsinki.

Conclusion

This chapter presented the mixed methods approach used in this thesis, with details on quantitative and qualitative study designs, data collection and preparation, data analysis and ethical considerations.

In the next chapter we will present the main results of our analyses.

7 Main results

“During World War II rescue workers, digging in the ruins of an apartment house blown up in the London blitz, found an old man lying naked in a bathtub, fully conscious. He said to his rescuers: ‘You know, that was the most amazing experience I ever had. When I pulled the plug and the water started down the drain, the whole house blew up.’”
– Fred Ederer (1975)¹⁸⁷

This chapter provides the main results of this thesis, with a breakdown of one section for each paper. It presents the reasons and approach for P4P development, the impact of the 2014 and 2018 P4P interventions on casemix index and readmissions, respectively, and the explored patient perspectives.

7.1 Why and how was hospital pay-for-performance developed?

The MoPH contracted with 26 public hospitals and 105 private hospitals to provide hospitalizations for more than 200,000 cases per year. In most cases, patients had a co-payment of 5% (later 10%) at public hospitals, and 15% at private hospitals, with the MoPH paying hospitals the remainder. For each hospital, the MoPH set global budgets and one of three reimbursement tiers. The financial amount generally varied between 10-30% across reimbursement tiers, depending on surgical procedure and base-rate. Between 2001 and 2013 the reimbursement tiers were solely determined by accreditation results (see figure 11).

The linkage between accreditation and reimbursement was instrumental in incentivizing hospitals to improve their structures and processes. However, by 2009, both the MoPH and hospitals had recognized several limitations of the accreditation-reimbursement linkage. These included the heterogeneity among hospitals, particularly the severity of cases admitted and the associated complexity of care delivered. Another factor was that the standards included in hospital accreditation had become numerous, and challenged the reduction of information into a single

‘final score’ to determine reimbursement tier. This context was favorable for undertaking new reforms.

The 2009-2014 MoPH reforms had three main goals:

1. To improve the appropriateness and fairness of MoPH-hospitals contracting.
2. To improve the efficiency of MoPH spending.
3. To promote good provider practices and discourage misuse/abuse of services.



Figure 11: Timeline of MoPH hospital contracting reforms.

The MoPH developed three committees to pursue these interrelated goals. These were focused on utilization review, admission criteria, and performance contracting. Committee members included medical, public health and information technology professionals, with affiliations including several universities and hospitals, and the MoPH. Relevant information was communicated across committees, allowing them to build over each other’s work.

The performance contracting committee (P4P development) used information developed by the utilization review committee to review approaches for evaluating case complexity and performance indicators. The utilization review committee developed information using hospitalization data, which were used by another committee to determine which frequent and expensive conditions to develop hospital admission criteria for. These processes were supported by annual inter-committee meetings, and having one overlapping member across committees to facilitate coordination.

Among the evidence developed were research findings that casemix varied widely across and within hospital reimbursement tiers, suggesting that using accreditation as a sole determinant of reimbursement was inappropriate and unfair (see 3.3.1)¹⁴⁷. Hospitals were engaged regarding these findings, and a new ‘mixed-model’ was developed, with casemix, patient satisfaction, accreditation and other minor components (see 3.3.1).

An evaluation of the early impact (one year) of the new model revealed an increase in the average casemix index among private hospitals from 1.14 to 1.18. Also, following the implementation of the new model, numerous hospitals were reported to have sought to improve their coding quality, through trainings held at private universities.

Table 5: Tier-level changes at group level, as a result of the 2014 P4P.

Hospital tier	Before P4P		After P4P	
High	44	34%	38	29%
Medium	58	45%	51	40%
Low	28	22%	40	31%
Total	130	100%	129	100%

7.2 What was the impact of pay-for-performance integration on healthcare effectiveness?

The study population totaled to 1,353,025 hospitalizations between 2011 and 2016 (see table 6). This was composed of 55% medical cases, 43% surgical cases and 2% mixed cases. Almost eight out of ten medical cases were medium-stay, with the majority of the remaining being short-stay. There was limited variation across the years in terms of total admissions and case proportions. Public hospitals admitted an increasing share of patients throughout 2012-2016, increasing from 31% to 37%.

Before the intervention in 2014, the monthly casemix index coefficient was 0.975 for medical cases, 1.284 for surgical cases and 1.783 for mixed cases (see table 7). Among short-stay medical cases monthly casemix index was lowest for short-stay cases (0.352) and highest for long stay cases (3.326).

Following the intervention, we had a reversal of a previously decreasing CMI trend among medical cases in public and private hospitals (see figure 12). We also had a large level change in CMI among short-stay but not medium-stay cases (see tables 7 and 8). Overall, for medical cases the largest changes were a level change among short-stays, and a trend change among medium-stays. Among surgical cases, the only change was an increasing trend in CMI in public hospitals. Among mixed cases, the medical component had an increasing trend in CMI in private hospitals only.

Seasonality was observed in medical CMI, and in particular medium-stay cases with a November–December major peak, among other minor peaks and troughs. For

surgical CMI seasonality was limited to private hospitals only, particularly with an April–June major peak. Mixed casemix seasonality had March and October peaks.

Code-level analysis allowed us to attribute casemix changes back to specific diagnoses and procedures. The most impactful changes on CMI were from 2,970 fewer cases of abdominal and pelvic pain, 698 fewer cases of intestinal infectious diseases, 1,001 fewer cases of fever of unknown origin, and 783 fewer cases of essential hypertension (see table 9). Altogether these changes represent about 4.3% of all annual medical cases. The greatest change on medium-stay CMI was due to a decrease of 2,237 cases of diarrhea and gastroenteritis.

Two types of shifting were also seen among several codes: from medium to short-stays (e.g. A09, I10, I20-I25.9), and from three-digit to four-digit ICD10 codes (e.g. J18, J44, P22). The greatest change on short-stay CMI was due to the improved coding of chemotherapy cases under the relevant code (Z51.1) instead of under neoplasms (C00-D49).

Separate investigation was made regarding the large change in vaginal deliveries, which revealed that private hospitals had 36.9% less vaginal deliveries in 2015–2016 than in 2013–2014, while public hospitals increased by less than 1%. Concurrently, cesarean deliveries decreased at private hospitals by 7.5% and increased at public hospitals by 9.2% (increasing surgical CMI at public hospitals). We identified the start of this trend back to early 2014, prior to the P4P intervention.

Change was also found in the share of CMI explained by different codes. The largest of these were for percutaneous transluminal coronary angioplasty (PTCA) and vaginal delivery (surgical); malignant neoplasm of breast and acute lymphoblastic leukemia (short-stays); bronchopneumonia, diarrhea and gastroenteritis (medium-stays).

Table 6: Hospitalization cases under the MoPH coverage at public and private hospitals in Lebanon, 2011-2016.

Case type	2011		2012		2013		2014		2015		2016		All years	
	n	%	n	%	n	%	n	%	n	%	n	%	n	%
All hospitals														
Medical	19,574	10.4	26,044	11.5	27,119	11.5	26,720	11.3	24,578	10.4	24,220	10.5	148,310	11.0
Short-stay	79,264	42.3	94,523	41.8	101,132	43.0	101,524	42.9	107,379	45.3	101,772	44.2	585,809	43.3
Medium-stay	2,156	1.1	2,851	1.3	2,793	1.2	2,770	1.2	2,674	1.1	2,494	1.1	15,744	1.2
Long-stay	100,994	53.8	123,418	54.6	112,044	55.8	131,014	55.4	134,631	56.8	128,486	55.7	749,863	55.4
All medical cases	100,994	53.8	123,418	54.6	112,044	55.8	131,014	55.4	134,631	56.8	128,486	55.7	749,863	55.4
Surgical	93,516	-	111,355	-	115,523	-	114,021	-	110,168	-	109,083	-	650,666	-
Cases	83,025	44.3	98,396	43.6	99,767	42.5	100,934	42.7	97,037	40.9	95,602	41.5	574,975	42.5
Mixed	5,247	-	6,358	-	6,164	-	6,668	-	8,241	-	8,932	-	41,610	-
Cases	3,584	1.9	4,120	1.8	4,114	1.8	4,446	1.9	5,517	2.3	6,396	2.8	28,187	2.1
Total cases	187,603	100	225,934	100	234,325	100	236,394	100	237,185	100	230,484	100	1,353,025	100
Public														
Medical	6,301	12.7	9,263	13.3	9,178	13.0	9,541	12.8	10,051	12.8	10,634	12.4	55,033	12.8
Short-stay	20,968	42.2	29,229	41.8	30,672	43.3	33,680	45.3	37,492	47.8	40,832	47.7	193,094	45.0
Medium-stay	501	1.0	747	1.1	633	0.9	630	0.8	669	0.9	704	0.8	3,889	0.9
Long-stay	27,770	56.0	39,239	56.2	40,483	57.2	43,851	59.0	48,212	61.5	52,170	60.9	252,015	58.7
All medical cases	24,049	-	33,463	-	33,031	-	33,055	-	32,871	-	36,471	-	192,940	-
Surgical	20,906	42.1	29,167	41.8	28,845	40.7	28,896	38.9	28,485	36.3	31,491	36.8	167,990	39.1
Cases	1,476	-	2,430	-	2,275	-	2,367	-	2,423	-	2,726	-	13,697	-
Mixed	957	1.9	1,453	2.1	1,492	2.1	1,593	2.1	1,678	2.1	1,997	2.3	9,180	2.1
Total cases	49,633	100.0	69,859	100.0	70,820	100.0	74,340	100.0	78,375	100.0	85,658	100.0	429,185	100.0
Private														
Medical	13,273	9.6	16,781	10.8	17,941	10.9	17,179	10.6	14,527	9.1	13,586	9.4	93,287	10.1
Short-stay	58,296	42.3	65,294	41.8	70,460	42.9	67,844	41.9	69,887	44.0	60,940	42.2	392,721	42.5
Medium-stay	1,655	1.2	2,104	1.3	2,160	1.3	2,140	1.3	2,005	1.3	1,790	1.2	11,854	1.3
Long-stay	73,224	53.1	84,179	53.9	90,561	55.2	87,163	53.8	86,419	54.4	76,316	52.8	497,862	53.9
All medical cases	69,467	-	77,892	-	79,492	-	80,966	-	77,297	-	72,612	-	457,726	-
Surgical	62,119	45.0	69,229	44.4	70,922	43.2	72,038	44.5	68,552	43.2	64,111	44.4	406,971	44.1
Cases	3,771	-	3,928	-	3,889	-	4,301	-	5,818	-	6,206	-	27,913	-
Mixed	2,627	1.9	2,667	1.7	2,622	1.6	2,853	1.8	3,839	2.4	4,058	2.8	18,666	2.0
Total cases	137,970	100.0	156,075	100.0	164,105	100.0	162,054	100.0	158,810	100.0	144,485	100.0	923,499	100.0

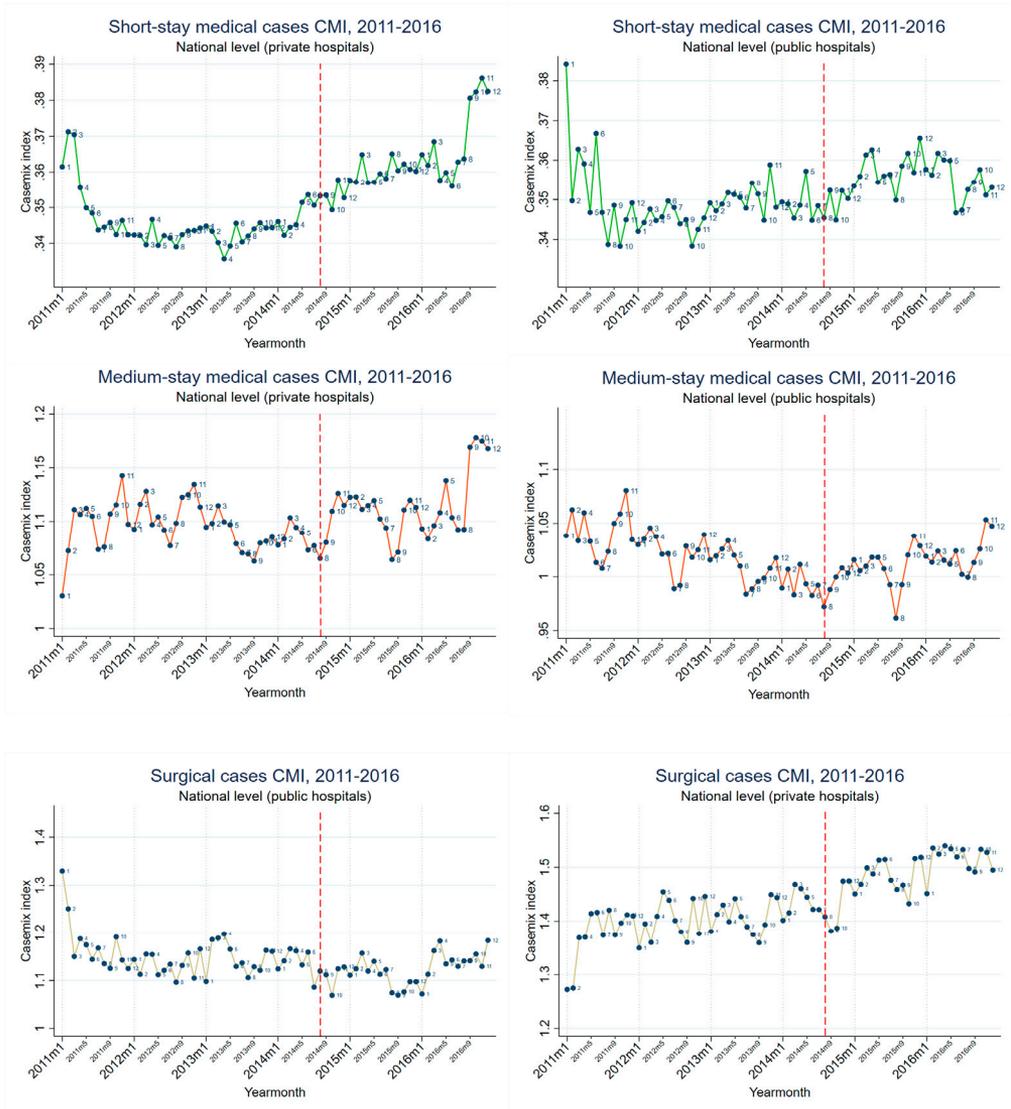


Figure 12: Medical and surgical hospitalization monthly casemix index, at public and private hospitals, 2011-2016. A dashed line represents the August 2014 hospital engagement event.

Table 7: Results of the ITS analysis on casemix index, adjusted for seasonality, 2011-2016, with the intervention point of August 2014 (two-month effect lag).

Case type	Hospitals	Before intervention			After intervention			
		Monthly CMI coefficient	TREND		TREND	LEVEL		
			% (CI)	Explained by		% (CI)	Explained by	
Medical	All	0.975	↘ 0.10% (0.06 - 0.13%)	-	↗ 0.11% (0.02 - 0.21%)	Medium-stay cases	↕ 2.25% (0.51 - 3.98%)	Short-stay cases
	Public	0.941	↘ 0.17% (0.11 - 0.23%)	Medium-stay cases	↗ 0.15% (0.06 - 0.22%)	-	-	-
	Private	0.989	↘ 0.06% (0.01 - 0.11%)	-	↗ 0.19% (0.06 - 0.32%)	Short-stay cases	↕ 2.70% (0.15 - 5.24%)	Short-stay cases
Surgical	All	1.284	↗ 0.05% (0.01 - 0.10%)	-	↗ 0.14% (0.06 - 0.21%) ¹	-	-	-
	Public	1.179	-	No trend	↗ 0.13% (0.02 - 0.24%)	-	-	-
	Private	1.326	↗ 0.12% (0.03 - 0.21%)	-	↗ 0.24% (0.13 - 0.35%) ²	-	-	-
Mixed	All	1.783	-	-	-	No trend	-	-
	Public	1.964	-	-	-	No trend	-	-
	Private	1.689	-	-	↗ 0.35% (0.10 - 0.60%) ³	-	-	-

¹p=0.06, ²p=0.11, ³p=0.33 ; no significant change between pre and post-intervention

Table 8: Relative changes, using respective monthly casemix index starting level, percentages.

Hospitals	ITSA aspect	MEDICAL			SURGICAL			MIXED		
		All stays	Short-stay	Medium-stay	Long-stay	All components	Medical component	Surgical component		
All	Starting level	0.975	0.352	1.078	3.326	1.284	1.783	1.530	1.991	
	Level at 2 months post-intervention	2.25%	2.61%	-	12.55%	-	-	-	-	
	Monthly trend, pre-intervention	-0.10%	-	-0.06%	-	0.05%	-	-	-	
	Monthly trend, post-intervention	0.11%	0.14%	0.09%	-	0.14%	-	-	-	
	Monthly trend, change	0.22%	0.17%	0.16%	-	-	-	-	-	
Public	Starting level	0.941	0.356	1.049	2.697	1.179	1.964	1.708	2.200	
	Level at 2 months post-intervention	-	2.05%	-	-	-	-	-	-	
	Monthly trend, pre-intervention	-0.17%	-	-0.12%	-	-	-	-	-	
	Monthly trend, post-intervention	0.15%	-	0.11%	-	0.13%	-	-0.81%	-	
	Monthly trend, change	0.31%	-	0.24%	-	0.22%	-	-	-	
Private	Starting level	0.989	0.352	1.089	3.498	1.326	1.689	1.421	1.906	
	Level at 2 months post-intervention	2.70%	2.56%	-	13.98%	-	-	-	-	
	Monthly trend, pre-intervention	-0.06%	-	-	-	0.12%	-	-	-	
	Monthly trend, post-intervention	0.19%	0.23%	0.15%	-	0.24%	0.35%	-	0.41%	
	Monthly trend, change	0.25%	0.28%	0.18%	-	-	-	-	-	

Table 9: Diagnoses and procedures with the greatest change effect on casemix index, pre- versus post-intervention.

Description	ICD/CPT code	Major effects	Notes
Neoplasms	C00-D49	Increased ss-CMI (87%)	Greatest change on ss-CMI ↓10,179 cases, net Mainly due to malignant neoplasm of breast and acute lymphoblastic leukemia Concurrent with increase in chemotherapy Z51.1 ↑11,666 cases
Intestinal infectious diseases (category)	A00-A09		
Diarrhea and gastroenteritis of presumed infectious origin	A09	Increased ms-CMI (25%)	↑961 ms-cases; ↑263 ss-cases
Unspecified non-infective gastroenteritis and colitis	K52.9	Decreased ms-CMI (2%)	Greatest change on ms-CMI ↓2,237 ms-cases; ↑179 ss-cases ↑745 ms-cases; ↑108 ss-cases
Abdominal and pelvic pain (category)	R10-R10.4		
Abdominal and pelvic pain, other/unspecified abdominal pain	R10, R10.4	Increased ms- and ss-CMI	↓2,970 ms-cases, net
Influenza and pneumonia	J09-J18	Decreased ms-CMI (4%)	↓1,975 ms-cases; ↑174 ss-cases
Pneumonia, non-specific	J18		↑3,909 ms-cases; ↑298 ss-cases
Pneumonia, specific	J18.0, J18.9		↓1,456 ms-cases ↑4,692 ms-cases
COPD	J44-J44.9	Increased ms-CMI (5%)	↑1,306 ms-cases, net
COPD with acute exacerbation	J44.1		↑625 ms-cases
COPD, non-specific	J44		↓234 ms-cases
Acute bronchitis	J20-J20.9	Decreased ms-CMI (3%)	↑1,145 ms-cases ↑747 ms-cases
Essential hypertension	I10		↓957 ms-cases; ↑174 ss-cases
Ischemic heart diseases	I20-I25.9	Decreased ms-CMI (7%)	↓1,100 ms-cases; ↑275 ss-cases
Fever of unknown origin	R50	Increased ms-CMI (3%)	Mainly due to angina pectoris and acute myocardial infarction ↑989 ms-cases; ↓12 ss-cases
Stroke	I64	Decreased ms-CMI (2%)	↓383 ms-cases; ↑12 ss-cases
Respiratory distress of newborn, non-specific	P22		↑334 ms-cases
Respiratory distress of newborn, specific	P22.0		↑287 ms-cases
Vaginal delivery	F9410L1	Increased surgical CMI (43%)	↑3,939 cases
Percutaneous Transluminal Coronary Angioplasty (PTCA)	X2983/6	Increased surgical CMI (36%)	Greatest change in absolute and in CMI share among all ICD/CPT codes ↑778 cases

ss: short-stay, ms: medium-stay, COPD: chronic obstructive pulmonary disease.

7.3 What was the impact of pay-for-performance on hospital readmissions?

The study population included 1,333,691 hospitalizations across 2011-2019. The mean monthly readmissions varied across the four conditions between 2.42% (cholecystectomy) and 6.48% (stroke). Index hospitalizations and readmissions were considerably greater for general cases and pneumonia than for cholecystectomy and stroke (see table 10). The time series of all four conditions were found to be stationary and seasonal (see figure 13). We applied a first-order seasonal difference for each of pneumonia, cholecystectomy and stroke models, as this was the most practically useful model, considering our analytical approach ^{149 p83}.

For estimating the impact of the intervention, we anticipated a level change in readmissions. However, given the uncertainty on the type of change, we included all three changes (level, ramp, pulse) in our first iteration of the models. The final models included a level change only. We ran diagnostic tests and plots to confirm that the criteria for independence and normality were met.

We found that following the intervention there was a level change in both cholecystectomy and stroke readmissions following the intervention. Mean monthly cholecystectomy readmissions decreased by about 24.8% (5.1%-44.5%) and stroke readmissions decreased by about 8.5% (1.5%-15.5%) (see table 11). There was no evidence of impact on general and pneumonia readmissions, neither at the level of all hospitals, nor separately among small, medium and large hospitals. Validation using Newey-OLS regression confirmed the decreased level change for cholecystectomy readmissions (by 52.7%, CI: 6.3%-99.0%, $p=0.026$), but no change was found among stroke, pneumonia or general readmissions.

Table 10: Descriptive statistics of 30-day readmissions for four condtions, 2011-2019

		General cases	Pneumonia	Cholecystectomy	Stroke
Admissions, total		1,333,691	70,585	26,820	13,370
Readmissions, total		80,080	3,569	681	876
Monthly readmission rate	Mean	5.91%	4.81%	2.42%	6.48%
	S.D.	0.51%	0.90%	1.02%	2.31%
	Min.	4.65%	3.13%	0.48%	1.71%
	Max.	7.29%	7.65%	4.89%	14.21%

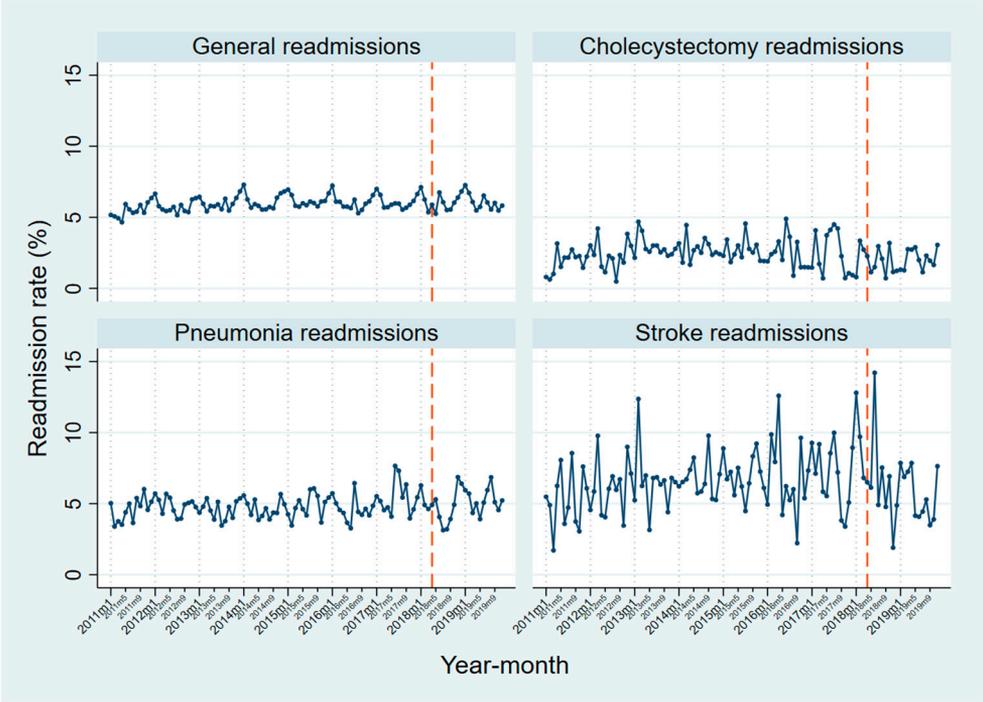


Figure 13: 30-day readmission rates for general readmissions, pneumonia, cholecystectomy and stroke, 2011-2019. A dashed line represents the anticipated impact time point, lagging after the hospital engagement events.

Table 11: Final ARIMA models and results across four readmission types, 2011-2019.

Model	General cases		Pneumonia		Cholecystectomy		Stroke	
	(1,0,0)	(1,0,0) ₁₂	(1,0,0)	(1,1,0) ₁₂	(1,0,0)	(1,1,0) ₁₂	(0,0,1)	(0,1,1) ₁₂
BIC		79.1		215.9		252.2		364.4
Level coeff., p. (95%CI)	0.256	0.075 (-0.026 to 0.537)	-0.154	0.658 (-0.837 to 0.528)	-0.714	0.048 (-1.420 to -0.008)	-1.637	0.012 (-2.907 to -0.367)
Constant	5.825	<0.001 (5.571 to 6.080)	0.081	0.520 (-0.166 to 0.328)	0.084	0.450 (-0.134 to 0.302)	0.274	0.011 (0.063 to 0.485)
AR	0.348	<0.001 (0.192 to 0.503)	0.270	0.037 (0.017 to 0.523)	0.071	0.585 (-0.184 to 0.326)	-	-
SAR	0.664	<0.001 (0.510 to 0.817)	-0.597	<0.001 (-0.774 to -0.420)	-0.502	<0.001 (-0.708 to -0.296)	-	-
MA	-	-	-	-	-	-	-0.056	0.693 (-0.331 to 0.220)
MAR	-	-	-	-	-	-	-0.870	<0.001 (-1.190 to -0.550)
Sigma	0.344	<0.001 (0.299 to 0.388)	0.973	<0.001 (0.866 to 1.080)	1.188	<0.001 (1.018 to 1.358)	2.281	<0.001 (1.924 to 2.638)
Log likelihood	-41.1	-	-134.9	-	-152.9	-	-221.1	-
Kolmogorov-Smirnov test	-	0.480	0.899	-	-	0.950	-	0.389
Ljung-Box test	-	0.806	0.739	-	-	0.949	-	0.900

BIC: Bayesian information criterion, (S)AR: (seasonal) autoregressive term, (S)MA: (seasonal) moving average term.

7.4 What are patient perspectives on hospital care?

Using qualitative content analysis, we developed five themes and 17 categories, which reflected the underlying and manifest meaning of the discussions, respectively. These are illustrated in figure 14.

Theme 1: Health is everything.

This theme reflected the importance that participants attached to health. It was portrayed as more important than money or wealth, and valued above other goods and services. It was necessary as a means of survival to be able to work and provide for yourself and those dependent on you. This included not only physical health, but also psychological and emotional well-being.

“Health is everything, I am a carpenter and I am paid on a daily basis, I have stopped working since a month and a half. Health is the basis of our existence, if we are not healthy we cannot work or do anything else.” (FGD4 Men-P3)

Theme 2: Being turned into 2nd class citizens.

Perceptions of the health system were captured by this theme. Participants recognized that “some hospitals are not for us”, which was reinforced by interactions with hospitals. They also felt neglected by the state, with citizens under the coverage of payers other than the MoPH having better healthcare access, services and respect from hospitals. Participants wanted the Ministry to be stronger in standing up for the rights of the poor and “make us feel that we are human beings”. Excessive bureaucracy was also a problem, such as needing to travel to different locations for approvals of some surgical procedures.

“The hospitals in Lebanon are classified into classes, if you tell someone you are going to [well-known hospital X], they tell you ‘this is not for you’; this is the way they reply.” (FGD1 Men-P1)

Participants reported seeing public hospitals neglected and their potential ignored, though they have a major role in supporting the poor. Participants often avoided primary care centers, seeing these as under-staffed, with less qualified or underpaid personnel. They also found ‘outsiders’ being favored with faster and free services, in reference to refugees covered by international non-governmental organizations and agencies.

“If you want to benefit this country you need to think of cutting down on the processes; the current way is very tiring.” (FGD4 Men-P4)

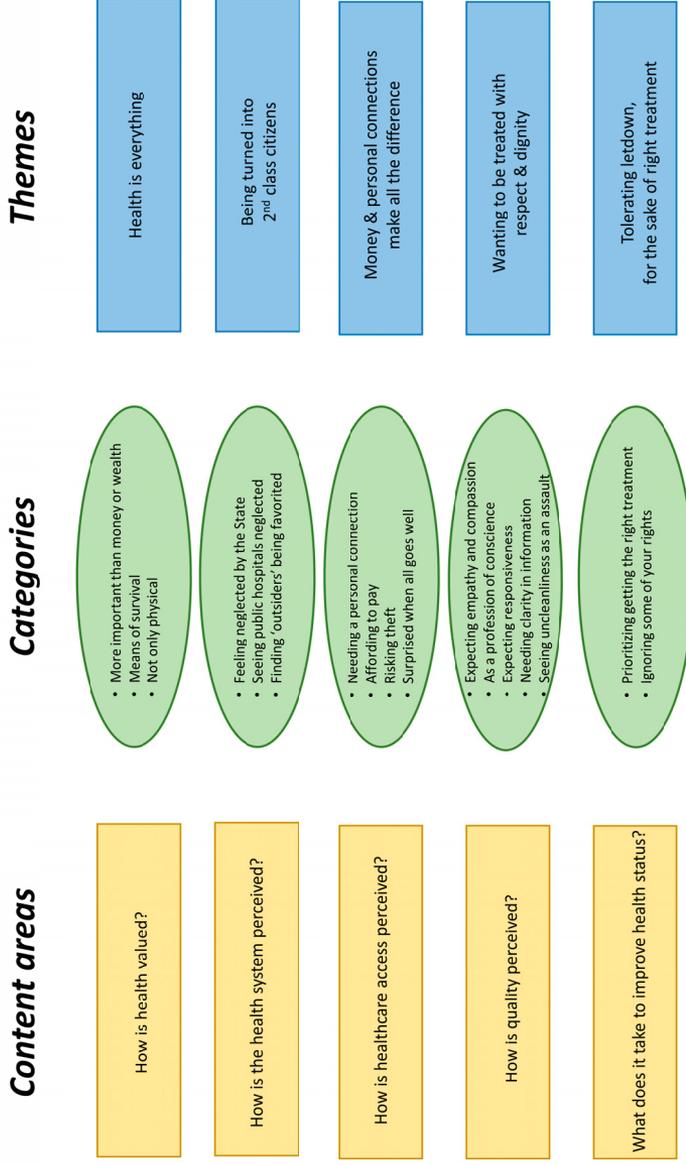


Figure 14: Overview of the main findings including content areas, categories and themes.

Theme 3: Money and personal connections (*wasta*) make all the difference.

This theme reflected participants' perceptions on access to healthcare. Needing a personal connection or having enough money were widely considered to improve access. They also recognized a mutually beneficial relation between some institutions and figures or authorities. For example, they reported that healthcare institutions would be protected from accountability, while political or religious figures/authorities would use their influence to facilitate services for some patients. Personal connections could also be used to decrease hospital bills or resolve perceived injustice or theft. Personal connections can sometimes be positive. For example, regular patients who become familiar to staff may be allowed to 'pay later'. But participants also considered that the health system would be better off without personal connections.

"As long as a person wants to be admitted through 'wasta', the hospital will not work properly." (FGD3 Women-P3)

The financial cost of health services was a major concern for participants, specifically affording to pay. This also affected their perception and behavior towards healthcare. Money was seen as a solution to any problem encountered at hospitals, especially if one lacked a personal connection. Participants noted that sometimes hospitals would claim that no hospital beds were available, to deny admission for those covered by the MoPH. They suggested this was a deceptive practice to allow hospitals to retain more profitable patients covered by other payers, or to compel patients to pay out-of-pocket.

"I told the nurse my mother is not the daughter of a minister or a president; I cannot pay [out-of-pocket]. Then we took her to another hospital." (FGD3 Women-P4)

Unaffordability led some patients to early hospital discharge or to forego medical tests. Some participants recounted a family member being worried about the bill and pretending to be better, to be discharged earlier. Participants recognized this was harmful to their health, but they had no alternative.

"When I had breast cancer [...] I couldn't do regular tests for checkup. I went through very hard times to do the tests and get the treatment. I sold my wedding ring [to get treatment]. The Ministry couldn't cover all the expenses; I reached a very difficult situation." (FGD7 Women-P1)

Participants recounted examples of borrowing money from relatives, the vulnerabilities when one has no income, and the difficulties in paying for recurring

costs such as chronic medications. They emphasized that they have a right to health, but they do not know how to realize these rights.

Risking theft when seeking healthcare was emphasized by participants. A common complaint from participants was of a healthcare personnel misinforming on procedures covered by the MoPH. Participants noted that it was common to be asked by hospitals or physicians for payment above the pre-defined co-payment amount, but were unaware this was illegal. They also recognized that not all doctors were the same, and some went far to support patients and their rights.

“I didn’t have any problem with the process at the Ministry, but they wanted me to pay 3,500 USD at the hospital. We disagreed with them, even the physician disagreed with such payment, then they decreased the amount to 2,500 USD [after the doctor spoke to the hospital].” (FGD4 Men-P3)

Insufficient information regarding payment meant that patients had a weaker role in their interaction with hospitals. Participants considered themselves the weakest of the three parties, including MoPH and hospitals. They also recognized some of the complex relations between hospitals and the MoPH.

“Every patient admitted under the Ministry’s coverage doesn’t know how much they are expected to pay [...]. The Ministry delays its payments to hospitals, so [hospitals] want to benefit from another source.” (FGD4 Men-P3)

Many participants also reported positive interactions with healthcare, whether with the Ministry or hospitals. They were surprised when all goes well. Such as when the admission approval went smoothly, upon receiving coverage by the MoPH (85%) for the hospital bill, or during hospitalization. Such interactions affected the perception and trust of participants towards the MoPH.

Theme 4: Wanting to be treated with dignity and respect.

This theme reflected how patients perceived the quality of care at hospitals, expressed as wanting to be treated with dignity and respect, implicitly and explicitly. It also illustrates how patients view the health profession in itself. Reflecting on their past experiences, participants acknowledged positive and negative interactions. These were not necessarily tied to the bio-physical outcome of treatment, but they did have an impact on how participants perceived hospitals.

Expecting empathy and compassion during their interactions with hospital personnel was important. This included the manner of communication between health professionals and patients, as well as with administrative personnel. Negative interactions had strong impressions on patients and their feelings of self-worth.

They also perceived that patients were treated with dignity and respect at private hospitals, but not public hospitals.

When asked what factors result in a positive experience during a hospital stay participants referred to experiences of compassionate personnel, positive attitude and care of nurses, and doctors striving far to support them.

“[The doctor said I needed surgery] and he visited me in the morning, and noon, and night. [...] He doesn’t take a Lira from me. He knew I am poor and suffering. I told him I’m a farmer [...] He said ‘my brother, this is helping someone in need.’” (FGD8 Men-P4)

Seeing health practitioners belonging to a profession of conscience was also an expression of patients’ desire to be treated with dignity and respect. They considered that being humane was the most important attribute. They also considered it a fundamental human right to receive care, regardless of ability to pay.

*“Humanity is the most important thing to be found at hospitals”
(FGD2 Women-P1)*

Participants were also expecting responsiveness from personnel. They empathized with the challenging conditions and long work hours of hospital personnel. Responsiveness was particularly important when one was in pain, or needed to use the toilet. Lack of responsiveness in such situations led to strongly negative patient experiences.

“When the nurse is in this profession, he must be expecting what he will face, he must not get annoyed and he must be patient. When the patient is at the hospital, he isn’t going to be faking it, he will really be in pain. This is why he will be nagging; because no one nags for no reason.” (FGD7 Women-P5)

Participants highly valued the time personnel devoted to them. This included having enough time with the doctor, to get a clear explanation of their condition and treatment options.

Moderator: “How would you differentiate between a good and humane doctor, and a bad one?”

P4: “When he provides you with information, as I told you. Because my doctor’s clinic is so busy, if I want to ask him a question he says ‘there is no need to know about these things, I know about them’; this annoyed me.” [...]

P5: “The doctor is good when he gives you from his time, even though sometimes he is in a rush, but he has to make you relaxed, to explain your condition to you.” (FGD7 Women)

It was also important to have a doctor you are comfortable with and could rely on. ‘Following the doctor’ was how participants largely explained their decision to visit or re-visit a hospital. They also acknowledged that chance also plays a role in what doctor you find, and the importance of the recommendation of friends or relatives before choosing a doctor.

Needing clarity in the information provided was highlighted, both from hospitals and the MoPH. This included information on treatment options, length of stay, and co-payment amount. Participants noted that information clarity may be even more important for health services, than in other types of services, since patients are more vulnerable and dependent on others.

More broadly, participants thought it was important to know which hospitals were better performers; the location and medications accessible from medication dispensaries and primary care centers; and the cost (or co-payment) of surgical procedures under MoPH coverage. Many were unaware of the difference between a deposit receipt and their hospitalization bill, or of their right to have a detailed bill. They reported minimal bill information was given at discharge, and this was usually verbal, not written.

Participants saw uncleanliness as an assault and as a danger to themselves, especially regarding the toilet and bedsheets. They noted cues they would use to assess cleanliness, such as the frequency and timing of cleaning. Cleanliness left a strong impression of hospitals among participants; “you see and sense cleanliness” (FGD4 Men-P3).

Theme 5: Tolerating letdown, for the sake of right treatment.

This theme reflected what participants see it takes to improve one’s health status. They underscored that their purpose in being in a hospital was to get the right treatment, which they generally prioritized. Some would accept being in a less-favored hospital, if it meant they could ‘follow their doctor’ and get appropriately treated. Getting the right treatment may involve ignoring some of your rights. Participants would sometimes be compelled to not to complain to personnel, because they did not want to compromise on their treatment outcomes.

“I ignore lots of things, you can say that I ignore 40-50% of my rights, the most important thing is to get the treatment.” (FGD4 Men-P4)

The results of the sorting exercise reflected the wide range of factors that patients consider important and affects their satisfaction (see appendix, Paper 4, table 3). This provided ‘patient satisfaction’ as a sixth patient perspective. The most prioritized statements related to issues that were commonly raised by participants during the FGDs. Specifically, this applied across statements a to k. Having clear

instructions at discharge (b) was an exception though, which suggests a missed opportunity to engage on this during the FGDs.

Conclusion

This chapter provided the main results of this thesis. We found that the 2014 P4P resulted in a decrease in unnecessary hospitalizations and improved coding, with the former being reflected vis-à-vis increased casemix index. The 2018 P4P intervention resulted in a decrease in stroke and cholecystectomy readmissions, but not general and pneumonia readmissions (regardless of hospital size). The MoPH developed P4P, including several components, to increase the appropriateness and fairness of the relation with hospitals. We identified six patient perspectives, reflecting how patients value health, being turned into a second-class citizen, the influence of money and connections, the importance of being treated with dignity and respect, tolerating letdown for the sake of right treatment, and patient satisfaction.

In the next chapter we will discuss the findings of each paper and the thesis as a whole, and then move on to discuss the methodological considerations.

8 Discussion

“It does not matter who you are, or how smart you are, or what title you have, or how many of you there are, and certainly not how many papers your side has published, if your prediction is wrong then your hypothesis is wrong. Period.”

– Richard Feynman

This chapter provides a discussion of the main findings of this thesis, which is related with the literature where relevant, and organized under general statements. We then move on to the methodological considerations, divided into those relating to the two P4P interventions (interrupted time series design), and the context of P4P and patient perspectives (content analysis).

8.1 Discussion of findings and relation to the literature

This thesis described the development and investigated the impact of hospital pay-for-performance in Lebanon. We have detailed how the Ministry of Public Health used participatory governance in developing P4P, which was intended for improving fairness and transparency in the relation between hospitals and the MoPH, as well as to improve effectiveness.

We have estimated the impact of the 2014 P4P integration, which resulted in improved effectiveness by reducing unnecessary hospitalizations, and improved coding quality. We also demonstrated how casemix index and routine data can be used to improve hospital performance in limited resource settings. We also estimated the impact of the 2018 P4P on hospital readmissions, which resulted in decreased cholecystectomy and stroke readmissions, but not general and pneumonia readmissions. However, readmission components within P4P require careful design and understanding of the specific context.

Our exploration of patient perspectives showed that these include satisfaction, valuing of health, health status, and perceptions of quality, access and the health system. We also described how pay-for-performance can be made more responsive to the patient population, through a broader consideration of their perspectives.

Using participatory governance, pay-for-performance was developed to improve fairness and transparency in the relation between hospitals and the MoPH, and to improve effectiveness.

The MoPH had set several interrelated goals for its reforms throughout 2009-2014, which it approached using features of participatory governance, developing interdisciplinary committees of a mix of professionals across hospitals and the MoPH. The Ministry was actively involved in this process in providing guidance, resources and institutional commitment. The multi-pronged approach to interrelated goals enabled the three committees to build on each other's work.

A major product of this period was the development of hospital P4P, which was integrated within the MoPH, with the mechanism used to determine hospital reimbursement tiers. Casemix and patient satisfaction were among the components of the P4P model, replacing the previous model which had been solely based on accreditation.

A possibly unique aspect of the P4P design was the inclusion of casemix as a performance measure. This is likely not used elsewhere. More developed systems tend to have had casemix incorporated before the advent of P4P in healthcare, and less developed systems have not attempted such a combination. Another important aspect was the use of ICD-10 and CPT codes to directly determine casemix index, overcoming the challenges due to the absence of national DRGs.

The determination of global hospital budgets is often the purpose of casemix in several countries, whereby casemix and historical volume is used to prospectively set budgets. However, the political support for such use of casemix in Lebanon was absent.

The P4P model provided a more transparent and fair relation between the MoPH and hospitals, particularly through accounting for the variable complexity of cases admitted in different hospitals. Another underlying result of the increased transparency is that it helped counter favoritism and clientelism, whereby hospitals with better political or sectarian 'connections' used to be more able to influence reimbursement tier determination, due to the limitations of the former model based solely on accreditation.

The integration of pay-for-performance in 2014 resulted in improved healthcare effectiveness by reducing unnecessary hospitalizations, and improved coding quality.

The 2014 integration of P4P into the MoPH-hospitals payer-provider relation resulted in a decrease of unnecessary hospitalizations, and improved coding quality. These changes were reflected in the casemix index, particularly among medium and short-stay cases. The in-depth analysis used in Paper 2 allowed us to attribute casemix change to specific codes, and understand the underlying changes. Without this, we would have been limited to our statistical finding of increased casemix, suggesting improved system efficiency. Uncovering that these changes included decreased unnecessary hospitalizations allows us to categorize this under quality of care, within the effectiveness dimension, vis-à-vis the Kruk and Freedman framework for health systems performance measures ⁴. Further downstream, we may expect this to be reflected in outcomes on patient health status (effectiveness), and on maximizing value of resources (efficiency).

The greatest effect on medium-stay casemix was from decreased hospitalizations for diarrhea and gastroenteritis. We expect that such cases are likely to be responsible for more unnecessary hospitalizations than any other condition, as has been previously suggested ^{188 189}. Acute gastroenteritis is typically self-limiting, not requiring hospitalization and accompanied with diarrhea. The most common reason for hospitalization is dehydration, particularly among infants or elderly. However, acute gastroenteritis is a good candidate for ‘cream-skimming’ by hospitals, since it is less complex, with lower expected costs (see 2.2.1) ⁷⁸.

Prominent decreases were also found in hospitalizations for abdominal and pelvic pain, intestinal infectious diseases, fever of unknown origin, and essential hypertension. These also present opportunities for unnecessary hospitalization ¹⁸⁹. We do not expect such changes to be due to changed disease burden, since the decrease was at least one order of magnitude greater than any change since 2011. For some of these conditions there was also a shift from medium-stays to short-stays, which suggests that not only did hospitalizations decrease, but there was also a decrease in their length of stay (LoS). Generally, changes in LoS are difficult to link to quality of care. However, for these specific conditions it likely reflects improved hospital practices.

The 2014 P4P integration did not influence the annual MoPH setting of global hospital budgets. The P4P incentive was linked to hospital reimbursement tiers. Given the aforementioned changes, one may reason that resources that would have been spent on unnecessary hospitalizations may have facilitated hospitals admitting more complex conditions, such as pneumonia, acute bronchitis and COPD (which increased).

The improvement of coding practices following the intervention was primarily among chemotherapy hospitalizations, with more breast cancer and leukemia treatment being correctly coded. Changes in coding practices following casemix introduction have been reported in other contexts^{190 191}. While such situations can be an example of ‘Campbell’s law’ (see 5.2.1), in this instance coding improvement was confirmed since diagnostic hospitalizations are considerably more costly typical chemotherapy sessions.

Seasonality and other changes in diagnoses and procedures were also detected in our investigation. A prominent example was in vaginal and cesarean deliveries. We were able to disentangle their effect, attributing this to pre-intervention changes.

Casemix index and routine data can be used to improve hospital performance in limited resource settings.

The MoPH did not have to develop new information systems or data collection for its P4P initiative beyond what had already existed at the Ministry. The data used is entered at the hospital level, and centrally stored in the hospitalization database. While this limits P4P to a certain range of measures, it does not involve additional costs to maintain. This demonstrates how effective hospital regulation can be achieved through systematic collection and analysis of routine data. Implementation costs of any P4P initiative are important to consider, and in some contexts these may even exceed the cost of incentives themselves¹⁹². The P4P integration contributes more broadly to the health system, using an approach that has been increasingly suggested^{13 193 194}.

This thesis demonstrates how casemix may be applied as a performance measure, provided there is an underlying reason. In this instance, the reason was the recognized problem of unnecessary hospitalizations and the potential for ‘cream-skimming’. Countries with such contextual challenges may use a similar approach to improve performance, by linking casemix to incentives. In other another context lacking such reasons, using casemix as a performance measure may not be helpful. A casemix component in the Lebanese P4P may be beneficial only up to a point in time when other mechanisms to control unnecessary hospitalizations are strengthened, or when casemix can instead be used for setting global hospital budgets.

Including readmissions within P4P resulted in decreased cholecystectomy and stroke readmissions, but not general and pneumonia readmissions.

The 2018 MoPH announcement that readmissions were to be added as a new P4P component resulted in decreased cholecystectomy and stroke readmissions. These readmission types had generally stable pre-intervention trends, which strengthens the validity of our finding. However, no impact was found on general and pneumonia readmissions, even after aggregating by hospital size.

The diverse findings by readmission type are not unusual, and several factors may plausibly be involved, including case volume and complexity. General and pneumonia readmissions were considerably greater in scale than cholecystectomy and stroke readmissions, and more widely distributed across about twice as many hospitals. A dilution effect may have occurred, whereby improvement across a few hospitals may be more easily reflected at the aggregate level, when the scale is smaller. This is especially noticeable in general readmissions, which did not reflect the decreases in cholecystectomy and stroke readmissions, despite the inclusion of their cases within general readmissions (by definition).

A second factor is case complexity, which also is related to risk of readmission^{74 75}. Pneumonia patients tend to be older and more medically complex than other hospitalizations. Such unmodifiable patient characteristics may mean that a large proportion of pneumonia readmissions may be unavoidable. A similar reasoning has been suggested regarding heart failure readmissions in the US HRRP⁴². However, this factor alone would not explain the lack of evidence of impact, since some proportion of pneumonia cases is expected to be avoidable, and has been shown to respond to incentives elsewhere (see 2.1.3)³⁸.

It is important to consider that the incentive may not have been sufficiently strong for hospitals to address some readmission types, particularly if perceived as high hanging fruit. Each of the readmission types had an equal weight within the readmissions component. Hospitals may have selected to address those that they consider more modifiable or more focused. The P4P inclusion of general readmissions was intended to incentivize broader improvement by hospitals, in the same logic of broader rather than narrower measures¹⁷. However, hospitals may have considered this too demanding, and perhaps understandably.

Acknowledging the mixed relation of hospital size and readmissions, we had investigated but found no evidence of impact on general and pneumonia readmissions when aggregating by hospital size^{97 195 196}. This further suggests that hospitals were either unwilling or unable to address these readmission types.

Including readmissions within P4P requires careful design and comprehensive understanding of context.

Understanding the context, potential pathways to impact and mediating factors is important for all P4P initiatives¹⁴. This may be particularly stressed when including readmission components. Considerable evidence has emerged highlighting unintended consequences with regards to readmissions in P4P, particularly from the US HRRP experience (see 2.1.3). Specifically, this involved the shifting of hospital visits from admission to emergency or observation room stays^{40 41}. Similar to the HRRP, the Lebanese P4P did not incorporate emergency and observation stays, and such data is not gathered by the MoPH. However, we do not expect this to have had a major impact, since such stays play a small role in the Lebanese context, and change was found in two of the four readmission types. Nevertheless, this is a weakness of the MoPH P4P, and should be addressed by incorporating all hospital visits. The possibility of hospital refusals to re-hospitalize should also be considered, and this likely would require an innovative approach regarding data collection, whether in Lebanon or elsewhere.

Another consideration for readmissions in P4P is to include mortality measures. These may be incorporated into P4P, or used for monitoring unintended consequences. In the US, some findings from the HRRP suggested it may have been associated with an increase in heart failure mortality, though a causal relation has been difficult to infer, due to methodological limitations and background epidemiology (see 2.1.3)^{47 50}. Nevertheless, it is logically plausible that mortality may be related, and given the gravity of such outcomes, P4P designs should incorporate an approach addressing mortality. In the Lebanese P4P, mortality was not linked to hospitalization data, due to technical and policy-related reasons. While the lack of mortality monitoring is a weakness, deaths were maintained in the denominator for readmission calculation (unlike standard metrics), and in effect functioned as a form of competing risk model¹⁹⁷. Such a feature has been suggested as a solution to some unintended consequences, particularly for measures regarding high-risk conditions/procedures and subgroups¹⁹⁷.

Patient perspectives include satisfaction, valuing of health, health status, and perceptions of quality, access and the health system.

Our qualitative investigation of patient perspectives in Lebanon contributes to the limited evidence base on patient perspectives and how patients may be engaged to evaluate their care¹⁹⁸. A key finding was that patients appear to have a clear idea of how their healthcare experience can be improved, and this extended beyond the boundaries of their hospital stay.

Patients valued health highly, and above other common goods or services. This applied to health itself, as well as health as a prerequisite for daily functioning and work. Patients felt the State neglected them, and neglected public hospitals, which

they perceive as having a major role in supporting the poor. As persons under the MoPH coverage, they felt like second-class citizens compared to refugees and to citizens covered by other payers.

Patients consider that personal connections (*wasta*) and money play a large role in determining access to healthcare. The health system would be better off if nobody would resort to personal connections, although it may occur in a positive manner of familiarity between regular patients and healthcare personnel. A major concern for patients was affording to pay for healthcare, with negative experiences including being over-charged, denied access, or having to forego treatment.

There is clear recognition of the variable dedication that healthcare personnel have for patients. Patients have expectations to be treated with dignity and respect, including empathy and compassion. They consider this to be intrinsic to the healthcare profession, and are disappointed when they are let down. They tolerate some of this for the sake of ‘following their doctors’ and getting the right treatment. But they are also interested in knowing which are the better performing hospitals. Cleanliness and regular contact with the doctor were highly prioritized by patients.

Our findings aligned with previous research on what matters to patients. They want humane, informative and available health professionals who are not financially-driven money-driven¹⁹⁹⁻²⁰³. While the purpose of being hospitalized is to be diagnosed or treated, patients place high value on humane personnel and on cleanliness. This concurs with reports that suggest that compassion may be prioritized above medical outcome²⁰⁴.

Personnel responsiveness may play a large role in shaping patient perceptions, and this has also been linked to the incidence of hospital-acquired infections, suggesting wider hospital problems²⁰⁵. Individual experiences have a strong role in shaping patient perceptions to both the hospital and health system, which concurs with other research^{206 207}. We also concur with previous research that the patient experience generally reflects the quality of care perceived by the patient²⁰⁸. This is a more objective patient perspective than satisfaction, despite their relation.

Most theories attempted to explain patient satisfaction in terms of expectancy theories (see 2.4.2)¹¹⁴. However, despite its conceptual and clarity limitations, it remains important for patients. Satisfaction is an emotion reflecting patients’ feelings, and is therefore not entirely explainable through objective reasoning. Patients are not ‘cool-headed consumers’¹¹⁷. Satisfaction should not be used as a global objective measure, as often seen in surveys. Instead, it should be acknowledged as just one of the various patient perspectives.

Pay-for-performance can be made more responsive through a broader consideration of patient perspectives.

Patient surveys are typically developed by experts, without being informed by patients themselves. This limits survey construct validity. The same had applied for the MoPH tool (based on HCAHPS). Our findings identify several issues that are not sufficiently captured by the MoPH tool, such as discharge and information clarity, risk of theft and time spent with personnel. In 2018, based on our findings, the MoPH improved the validity of its survey by updating several items.

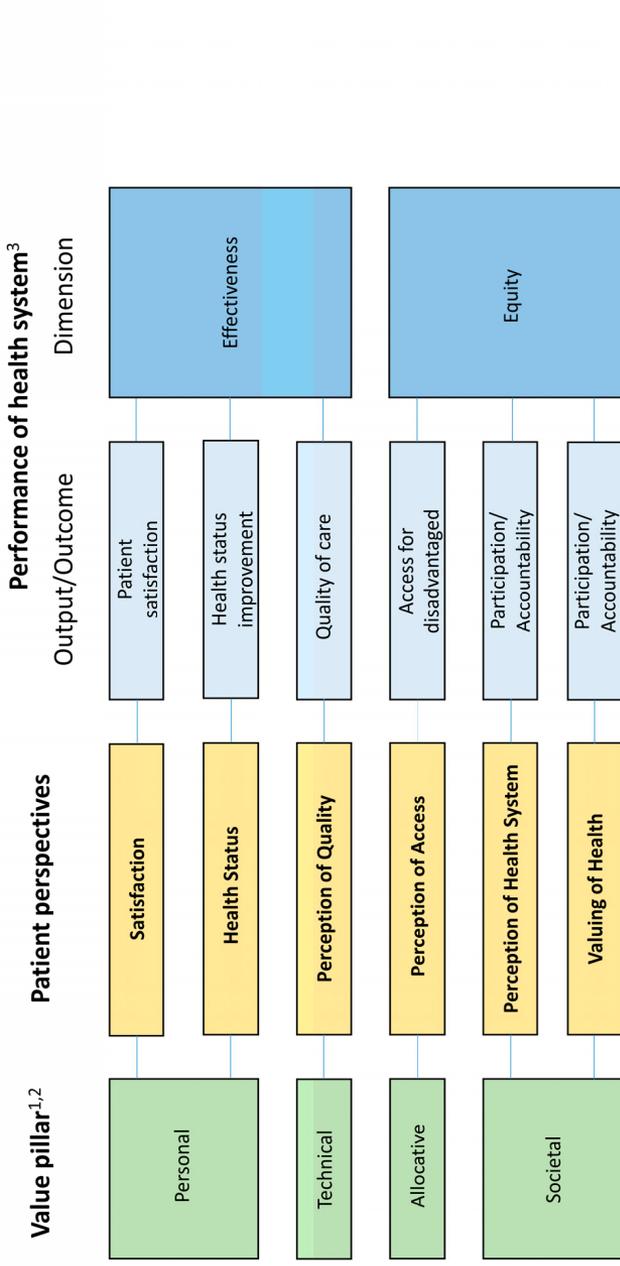
Health systems have an opportunity for wider engagement of patients for their perspectives. This benefits patient-centered care, health system effectiveness and equity.

Patients are usually engaged regarding their satisfaction, treatment (health status), and experience (care quality). However, engagement rarely occurs on the three other patient perspectives we identify: perception of access, perception of health system, and valuing of health. These represent a wider form of engagement, which may be aimed at health system development or reform. Such engagement recognizes patients not solely as healthcare recipients, but also as essential contributors to shaping the values and functions of a health system.

We developed a framework to relate patient perspectives to health systems performance and value-based care (see figure 16). The purpose of this framework is to illustrate the conceptual linkages between these three domains, thus facilitating the organization of wider patient engagement. We use the six patient perspectives we identified, alongside the Kruk and Freedman framework for health systems performance, and the value pillars recently proposed by the WHO EU Health Observatory and the European Commission^{4 209 210}. The value pillars include personal, technical, allocative and societal values.

Linking patient perspectives to the value pillars allows value-based programs such as pay-for-performance to consider a wider aim, including allocative and societal values, which are generally neglected. Wider patient engagement may involve assessing healthcare access and the overall health system, thus contributing towards decreasing population inequity. Capturing how patients value health relates to both accountability and participation within a health system. More broadly, health systems which engage widely may not only benefit patient-centeredness, but also form a bridge between patient-centeredness and people-centeredness.

In our research, Lebanese patients placed a high value on health, and strongly supported accountability and the improvement of public hospitals. These are likely not reflected in the priorities and spending of the Lebanese government. In the Lebanese context, examples of wider engagement may include public participation in determining local healthcare needs, and in the development of national strategies.



¹ Smith PC, et al (2020). Building on value-based health care: Towards a health system perspective. European Observatory Policy Briefs. World Health Organization.

² European Commission (2019). Defining Value in 'Value-Based Healthcare'. Report of the Expert Panel on effective ways of investing in Health.

³ Kruk M., Freedman L. (2008). Assessing Health System Performance in Developing Countries: A Review of the Literature, p. 263-276.

Figure 16: Relating patient perspectives to value-based care and health systems performance.

8.2 Methodological considerations

“The seeker after the truth is, therefore, not he who studies the writings of the ancients and, following his natural disposition, puts his trust in them, but rather the one who suspects his faith in them and questions what he gathers from them, the one who submits to argument and demonstration, and not to the sayings of a human being whose nature is fraught with all kinds of imperfection and deficiency. It is thus the duty of the man who studies the writings of scientists, if learning the truth is his goal, to make himself an enemy of all that he reads, and, applying his mind to the core and margins of its content, attack it from every side. He should also suspect himself as he performs his critical examination of it, so that he may avoid falling into either prejudice or leniency.”

– *The Optics Of Ibn Al Haytham, Books I-III, On Direct Vision* ²¹¹

This thesis used a mixed methods approach, which we consider to be advantageous in investigating complex interventions such as P4P (see 6.1). We were able to address research questions relating to different aspects of P4P, using quantitative and qualitative analyses. It also included investigations of two different P4P interventions, within the same overall context. The four papers differ considerably in their focus and analytical approach. Investigation of the 2014 and 2018 P4P interventions used a quasi-experimental ITS design, which may be considered the most appropriate, given the nature of the intervention and context. We first reflect on the four validities most relevant for ITS design for Papers 2 and 3, and other general aspects, followed by reflections on Papers 1 and 4.

8.2.1 Casemix and readmissions

1. Internal validity

We do not consider a threat to internal validity from history to be plausible. No other events occurred around the time of the 2014 or 2018 P4P interventions that may have offered an alternative explanation for our findings. Instrumentation changes did occur following the 2014 P4P (chemotherapy codes), and was detected in code-level analysis. This did not threaten the internal validity of our findings, as the contribution of instrumentation was measurable, and other changes were detected that were not attributable to instrumentation. Regarding the 2018 P4P, it is possible that some hospitals may have intentionally miss-coded readmissions to avoid algorithm detection. However, we consider this to be unlikely or at least of negligible impact, due to the wide case definitions used. Our use of an ITS design allowed us to control for any underlying process (maturation), which may have

otherwise explained detected changes. Also, since both our 2014 (casemix) and 2018 (readmissions) time series were stable and of medium length, we do not expect regression to the mean to have been plausible. No selection bias is expected, since we used historical self-controls for all of our time series.

2. External validity

There is a wide variability in P4P designs and contexts. From a design perspective, this may include different incentives, components types and weights. The locations and health systems of countries undertaking hospital P4P should also be considered. These factors limit the generalizability of our findings to other designs and contexts. Nevertheless, our findings apply for the Lebanese P4P and contribute to the limited body of evidence regarding hospital P4P impact. Temporal drift with regards to readmissions was addressed using age-adjusted readmission rates. Although we had no age adjustment for the 2014 P4P casemix investigation, we do not consider this to have had a considerable impact on the changes in casemix found, particularly as our study design controlled for secular trends.

3. Statistical conclusion validity

Our 2014 and 2018 P4P time series were stable in the pre-intervention period, and had a sufficient number of data points. This renders a threat from low statistical power or miss-interpretation less plausible. For the readmissions time series, we had missing data for December readmissions, due to the datasets available from the MoPH. This may be considered a threat to statistical conclusion validity. However, we consider this to be less plausible, since our intervention did not coincide with the missing data period, and the time series was stable. Our analytical approach ensured that test assumptions were not violated, particularly normality and independence.

4. Construct validity

For both 2014 and 2018 P4P time series, we used calendar monthly data points. This allows for sufficient data points, compared to using quarterly data points. Also, we did not use weekly or daily data points which would have been less likely to be normal (requiring data transformation). The use of 28-day months may have increased the construct validity of our investigations, as we can expect some differences in hospitalizations between weekends and weekdays. However, we opted for calendar months, to improve the interpretation of our findings. It is also relevant to note that the first few months of 2011 were the earliest period in the operation of the hospitalization database, with some hospitals lagging behind in the reporting of data to the MoPH. This represents a threat to construct validity. However, given its short duration and distance from the interventions, we do not expect this to have been a major threat. For readmissions in particular, the inability to include the entire spectrum of hospital visits may be considered a threat to construct validity, which should be addressed in future developments.

Overall, we consider that all four validities have been met in our 2014 and 2018 P4P time series investigations, while noting the above limitations. We cannot identify alternative explanations for the major changes we have attributed to the interventions, and express high confidence in the causal inference made.

Investigation of both P4P interventions used ITS design, but involved differences in our analytical approach. The 2018 P4P investigation used ARIMA, which may be considered superior over the Newey-OLS regression approach used for the 2014 P4P investigation¹⁶⁷. This is mainly due to the greater capability of ARIMA to account for seasonality, and the ability to difference the time series if it is found to be non-stationary. On the other hand, the 2014 P4P investigation may be considered more rigorous, in that it involved not only testing for change, but also attributing this to specific diagnoses and procedures. The analogue of this for the 2018 P4P would be an analysis of which readmission diagnoses or procedures changed, and should be a subject for future research. Given the complexity of outcome-based P4P, it is particularly useful to explain the underlying changes, rather than only estimate impact^{212 213}. This may also involve investigation of the causal pathways, particularly regarding complex interventions²¹⁴. In this thesis, we did not investigate the hospital response mechanisms or ‘black box’ (see chapter 4). This remains an important aspect of P4P evaluation, as it would provide insight into the motivations, decisions and actions of managers and health professionals.

8.2.2 Context of P4P and patient perspectives

The first paper provided a descriptive analysis addressing why and how P4P was developed in Lebanon. The analytical approach involved using documents developed by the project team, which included this author, in a role coordinating across the committees and being the link between them and the MoPH Director-General. As such, it may be impossible to avoid some bias due to the author's own perspective. Nevertheless, we entirely relied on the documentation produced by the project, and the discussions held with select project participants. This was supplemented by subsequent circulation of the paper's drafts across co-authors. Six of the eight co-authors were project participants, with two others involved in an annual evaluative capacity.

A weakness in the paper may be noted in the table on hospitalization changes, which included summary statistics across June to May for 2011-2014, and December 2014 to June 2015. Aside from being a simple pre/post design, it fails to account for seasonality. This renders the figures of the post-intervention period unreliable, though we note this had been labelled a preliminary analysis requiring subsequent investigation. An additional note is with regard to the quote from the then-minister of public health; this represented his view, but not that of the first author. While utilization review and improved auditing can considerably improve efficiency, we cannot expect to produce savings within a context of rising healthcare costs and increasing coverage for older citizens.

Trustworthiness, credibility, dependability and transferability were elaborated on in Paper 4. Our qualitative investigation did not seek to generalize regarding patient perspectives in Lebanon. However, we document a range of perspectives, many of which are also relevant in contexts other than Lebanon. We also sought a wide variation of patient experiences, with women and men participants coming from different geographic regions. Nevertheless, unforeseen selection bias may have occurred, threatening external validity. Another threat to external validity is due to the influence and interpretation of the researcher, which is impossible to avoid in such research design. However, we expect this was well-balanced, with no meaning imputed that was not present in the participants' discussion, and further improved through discussions with the two research assistants and co-authors.

Conclusion

This chapter provided a discussion of the main findings of this thesis. These included the use of participatory governance principles in developing P4P; the integration of P4P improving effectiveness by reducing unnecessary hospitalizations; how routine data and casemix can be used for hospital performance; the mixed findings of including a readmissions component to P4P; the six patient perspectives identified, and how these can be better considered to make P4P and broader health system development more responsive to the population. We also discussed the methodological considerations of our investigations. We find that the four most relevant validities have been met in our ITS investigations, but also note some limitations. We also note the limitations regarding the first paper, particularly the risk for some bias, and the measures taken for the fourth paper regarding trustworthiness, credibility, dependability and transferability.

In the next chapter, we provide the conclusions based on the research undertaken in this thesis, followed in the final chapter by a list of recommendations and suggestions regarding the P4P model of the MoPH, health system performance and P4P in other countries, and for future research investigation.

9 Implications for policy, practice and research

9.1 For the Lebanese health system and the pay-for-performance model

1. *We recommend* that the MoPH engage the public to capture their perspectives on health system redesign and crisis recovery, to improve access, participation, accountability and quality of care. This should include actively informing patients of their rights.
2. *We recommend* that the MoPH retain the casemix index component in its P4P model, until a time at which casemix is incorporated in setting global budgets of hospitals. Such a step should be made in the near-term, as unnecessary hospitalizations are better addressed through hospital budgets rather than reimbursement tiers. This would also free the P4P model to include new components that are more modifiable across hospitals. Capturing case complexity remains relevant for providing a fair comparison across hospitals, both in terms of casemix itself, as well as an adjustor for other indicators.
3. *We recommend* that the MoPH retain stroke and cholecystectomy within the P4P readmissions component, and remove general and pneumonia readmissions. *We suggest* including other readmissions, but avoiding higher risk conditions/procedures or subgroups (e.g. elderly, children). Maintaining the readmissions component should be conditional on meeting the next recommendation (#4). Alternatively, the readmissions component may be changed to replace readmissions with specific process indicators directly related to them, such as hospital-to-community transitions and discharge instructions, provided that appropriate measurement and monitoring are established. The future of readmissions in P4P likely lies in this last alternative, namely, in process-based components coupled with outcomes-based monitoring.

4. *We recommend* that the MoPH develop its system for hospital data collection, to include the entire spectrum of hospital visits. This would include observation and emergency room visits.
5. *We recommend* that the MoPH establish a linkage between hospital mortality data and its P4P model. Mortality should not be used as a component within P4P, but should be continuously monitored across a broad range of conditions/procedures, including those targeted by the P4P readmissions component.
6. *We suggest* that the MoPH establish linkages with other public and private payers regarding hospitalization data. This path may begin with data standardization and centralization, and leading up to eventual unification.
7. *We suggest* that the MoPH remove the accreditation component within P4P, and instead require all hospitals to be accredited prior to contracting with the MoPH. Hospitals should continue to be incentivized to gain accreditation.
8. *We recommend* that the Lebanese State improve public hospitals and dedicate greater resources for the health system, to more closely reflect the high value on health placed by Lebanese patients. A multi-sectoral initiative to decrease the influence of personal connections (*wasta*) and money is an important component of this. This should ultimately lead up to the establishment of a single Lebanese National Health Service, providing equitable coverage for the entire population.

9.2 For health systems and pay-for-performance in other countries

1. *We recommend* that P4P initiatives be integrated within health systems, rather than become stand-alone programs. P4P models should be developed using principles of participatory governance, and engage patients on their perspectives.
2. *We suggest* that health systems consider including casemix as a P4P component, provided that they do not already use casemix for budgeting and are challenged with unnecessary hospitalizations.
3. *We recommend* that health systems target readmissions using P4P on the condition that monitoring mechanisms are established for capturing the entire spectrum of hospital visits and mortality.
4. *We recommend* that P4P models include locally-relevant components, and exclude high-risk conditions/procedures and subgroups.
5. *We recommend* health systems to more widely engage people for their perspectives, including perception of access, perception of health system, and valuing of health, and not only satisfaction, experience and treatment. This may involve engaging the public in determining local healthcare needs and national strategies. Moving to a genuinely person- and people-centered health system cannot be achieved without wider engagement.

9.3 For future research

1. *We recommend* the use of appropriate interrupted time series analysis for the evaluation of P4P intervention impact in different contexts.
2. *We recommend* further research to investigate the impact of Lebanese P4P interventions on other components not yet investigated. Specifically, this relates to the 2018 impact on casemix, ICU capacity and utilization, and hospitalization of elderly persons, as well as the ‘black box’ representing hospital mechanisms used to change outputs or outcomes.
3. *We suggest* further research to investigate the impact of the economic crisis, COVID-19 and the Beirut port explosion on hospital performance, vis-à-vis the various P4P components.
4. *We recommend* further research to investigate the changes in causes of readmissions following cholecystectomy and stroke, to possibly attribute the observed readmission decreases to specific diagnoses or procedures.
5. *We suggest* further research to investigate patient perspectives in non-Lebanese contexts, to increase knowledge on the external validity of the framework relating patient perspectives to health systems performance and value-based care.

10 Conclusion

This thesis has described the development and investigated the impact of hospital pay-for-performance in Lebanon. Besides the contribution to the *episteme* on hospital P4P (see Preface), this thesis also documents the *techne* involved in developing and evaluating P4P in Lebanon, including the use of appropriate interrupted time series analysis.

The MoPH used principles of participatory governance to tackle the goals set by the 2009-2014 reforms. A major goal was the development of hospital P4P, which aimed to increase the fairness and transparency of the relation between hospitals and the Ministry. This was also to counter political or sectarian favoritism and clientelism.

The 2014 integration of P4P into the payer-provider relation between the MoPH and hospitals improved system effectiveness, by decreasing unnecessary hospitalizations and improving hospital coding quality. This was reflected in an increased hospital casemix, which was a major component of the P4P model, and confirmed by code-level analysis. Despite the limited resources available, the MoPH was able to use routine data and casemix towards improving hospital performance. This thesis demonstrated how casemix can be applied as a performance measure, provided that the context is appropriate.

The 2018 inclusion of readmissions into the P4P model led to a decrease in stroke and cholecystectomy readmissions, but not general and pneumonia readmission. There was no change in readmissions in the latter two readmission types, regardless of hospital size. Different factors may have contributed to these mixed findings, such as case complexity, dilution effect, and perceived or genuine ‘low hanging fruit’. The MoPH does not currently capture the entire spectrum of hospital visits for a readmission component, but this should be addressed in the future.

Patients in Lebanon highly valued health, and were strongly supportive of improving public hospitals, and of accountability to counter the influence of personal connections (*wasta*) and money. Affording to pay is a major concern for patients. Patients had a mix of positive and negative experiences under MoPH coverage in different hospitals. They expect to be treated with dignity and respect by humane health professionals, and highly prioritized hospital cleanliness and regular doctor contact.

We were able to increase the construct validity of the MoPH patient survey tool, using the findings from this thesis. We identified six patient perspectives, including the rarely-addressed patient perception of access, perception of health system, and valuing of health. We also developed a framework to relate patient perspectives to health system performance and value-based care. This may be helpful in improving health systems, particularly regarding public participation, accountability and access. It is important to more widely engage people on health system issues, recognizing them not solely as healthcare recipients, but also as essential contributors to shaping the values and functions of a health system.

Acknowledgements

I am very grateful for all the persons who have accompanied me in this journey. This thesis is the product of many people contributing directly or indirectly, to the work, the thinking and the spirit. This includes my supervisors, mentor, co-authors, project members, colleagues, friends and family. It also includes the many persons who have developed the materials I have cited, and the conceptual and analytical tools used. I am particularly grateful to the patients and health professionals I have engaged with, including during my medical studies and up to the focus group discussions that formed part of this thesis.

Thanks to Björn Ekman, for his supervision of this thesis, encouragement, critical thinking, support and guidance throughout the past several years. Although I lack a counterfactual, it is highly likely I would not have taken this path without your initiative. It has been a great experience climbing yet another mountain with you.

Thanks to Maria Emmelin, for her co-supervision of this thesis, encouragement, creativeness, and in-depth explorations into patient values and perspectives. Your guidance has been instrumental and your presence inspirational.

I have learned much from Björn and Maria over the past several years, and they have kindly shared their knowledge and expertise throughout this journey, in a most interesting experience of multi-disciplinary collaboration.

Thanks to Walid Ammar, for getting me involved in health systems, and being the personification of Aristotle's *phronesis*. His wisdom and insight has not only guided the development of the Lebanese health system over three decades, but also influenced people of varied backgrounds to develop their expertise, and strive for better and fairer systems and patient outcomes.

Thanks to Anette Agardh, for kindly hosting me at the Social Medicine and Global Health research group at Lund University over several years, and providing a very friendly and supportive environment for undertaking research.

Thanks to Josephine Borghi, for kindly accepting to be the opponent for my PhD defense, for travelling from the United Kingdom to Sweden for this event, and for sharing her expertise, time and effort.

Thanks to Johan Jarl, Lina Magnusson and Lars Kayser, for kindly accepting to be members of the examining committee for my PhD defense, and for sharing their

expertise, time and effort. Also thanks to Anders Halling and Johan Harl for their previous roles and feedback during the halfway review of this thesis.

Thanks to my friends and colleagues in Lebanon, with whom I have shared challenges throughout many years as we sought to improve the health system and patient outcomes, while maneuvering around political interference. I am thinking of you Rita Frieha, Hilda Harb, Jihad Makouk, Fadi El Jardali, Jenny Romanos, Abeer Al Halabi, Elise Barakat, Sizar Akoum, Nabil Kronfol, Ghassan Hamadeh, Georges Saade, Nadwa Rafeh, Rasha Hamra, Alissar Rady, Michelle Kosremelli-Asmar, Michel Matta, Ziyad Mahfoud, Kassem Kassak, Rima Afifi, Rita Karam, Randa Hamadeh, Carin Alsokhn, Nader Nassif, Hovig Jarkhedian and many other health professionals who continue the good struggle across hospitals, universities, institutions and elsewhere.

Thanks to Anne Mills, for sharing her wisdom, experience and encouragement as senior advisor on the HSRI project for the Lebanese P4P during 2016-2020.

Thanks to the SMGH administrative staff, Ditte Mårtensson, Elena Aguilar, Katarina Ambohm, Maria Agardh, as well as LU's Anette Saltin, for guidance, support and positive spirit throughout the administrative processes.

Thanks to my SMGH colleagues, past and present, for the great company and interesting discussions over several years, particularly before the pandemic upturned the world. I am thinking of you Natalia Vincens, Benedict Oppong Asamoah, Mats Sundbeck, Markus Larsson, Tanya Andersson Nystedt, Becky Nelson, Jens Wilkens, Tobias Holder, Rebecca Hayes Mejia, Jack Palmieri, Mia Kolak, Pia Svensson, Ochieng Malachi Arunda, Hannah Kaiser, Frida Pilgaard, Joseph Kazibwe, Gilbert Tumwine, Minilik Demissie Amogne, Mahnaz Moghaddassi, Martin Stafström, Dell Saulnier, Jesper Sundewall, Per-Olof Östergren, Devika Mehra, Vikas Choudhry, Susanne Sundell-Lecerof, Cecilia Fernbrant, Fiffi Boman and Anders Emmelin.

Thanks to the many patients I have encountered throughout my life, for teaching me what matters most, and how national priorities should reflect this. Everyone has the right to compassion, dignity and good healthcare. The memory of many of you will always accompany me.

Thanks to my friends, who now inhabit many lands, and are always close to heart and mind.

Thanks to Ramzi Semrani, for his creativity, passion for nature, longstanding friendship and the thesis cover photograph.

Thanks to my family and relatives, including my grandmother Sylvie, my cousins, aunts and uncles, for the support and lifelong experiences. The presence of those departed will also always accompany us: Tony Helou, Joe Khaled, Mahfouz Fawaz,

Ferdinand Gubbels, Marcello Moscatello, Josephine Hajj, Najla and Tanous Khalife, and Ishak Sejaan.

My heartfelt appreciation goes to my parents, my siblings and their families, my wife and my children.

To Liliane and Paul, for their enduring support, wisdom, integrity and dignity. Their influence has been immeasurable.

To Imane, Imad and Noor, for their incredible companionship on my life's path, and to Marco, Patrick, Paul and Jesper, for enriching my journey even further.

To Caroline, for her encouragement and regular visits supporting the family.

To Sophie, for the unwavering dedication and support, and the many stimulating discussions on health, health systems and life.

To Oliver and Nadine, whose companionship means more to me than anything in this world, and for all the moments of play, learning and adventures we have shared, with many more to come.

Køge (Denmark) and Jaj (Lebanon)

August 2023

Jade Khalife

Epilogue

Does hospital pay-for-performance work? It depends! We have a wide diversity in context, designs, incentives, measures and other factors that determine whether or not P4P ‘works’. From my perspective, based on the evidence developed in this thesis, hospital pay-for-performance in Lebanon resulted in several positive impacts, improving the relation between hospitals and the Ministry of Public Health, and providing a tool for continuous development of the health system. An important effect has also been curtailing the influence of favoritism and clientelism, through a fairer, more appropriate and transparent evaluation of hospital performance.

Pay-for-performance should also be designed to contribute to health systems, beyond the goals of their components. This includes developing approaches to make better use of routinely collected data. Considering unintended consequences, P4P should be able to ‘afford to fail’, without harming patients or health professionals in the process. This underlines the importance of monitoring and response within P4P.

Overall, I think hospital pay-for-performance may have an important role to play within health systems, particularly in an integrated form. However, this role is as one among several tools, and should not be the focus of the system. Rather, health systems should be centered on the people, and patient engagement is central to this. P4P initiatives should be developed to reflect this.

In Lebanon specifically, I think a redesign of the health system is needed, taking the best from the past system and developing a Lebanese National Health Service. Such an institution would provide equitable coverage for the entire population, and should be protected from political interference.

All countries will face increasing threats in this new age of pandemics, climate catastrophe and political-economic upheavals. This also represents an opportunity to re-center our health systems on the people, reflecting their high priority on health. A brighter future is one of humane and compassionate health systems.

References

1. Parry R. Episteme and Techne: The Stanford Encyclopedia of Philosophy; 2021 [cited 2023 28 July]. Winter 2021 Edition:[Available from: <https://plato.stanford.edu/archives/win2021/entries/episteme-techne/>].
2. Latham J. The research canvas: Framework for designing and aligning the "DNA" of your research study. 4th edition ed: Organization Design Studio®Ltd., 2022.
3. Nigel Nicholson NS. The Rhetoric of Medicine: Lessons on Professionalism from Ancient Greece. New York: Oxford University Press 2019.
4. Kruk M, Freedman L. Assessing Health System Performance in Developing Countries: A Review of the Literature 2008.
5. Donabedian A. Evaluating the quality of medical care. *Milbank Mem Fund Q* 1966;44(3):Suppl:166-206. [published Online First: 1966/07/01]
6. Gergen J, Josephson E, Coe M, et al. Quality of Care in Performance-Based Financing: How It Is Incorporated in 32 Programs Across 28 Countries. *Glob Health Sci Pract* 2017;5(1):90-107. doi: 10.9745/ghsp-d-16-00239 [published Online First: 2017/03/17]
7. Chee TT, Ryan AM, Wasfy JH, et al. Current State of Value-Based Purchasing Programs. *Circulation* 2016;133(22):2197-205. doi: 10.1161/circulationaha.115.010268 [published Online First: 2016/06/02]
8. Smith R, Hanson K. Health systems in low- and middle-income countries : an economic and policy perspective. Oxford: Oxford University Press 2012.
9. Grace FC, Meurk CS, Head BW, et al. An analysis of policy levers used to implement mental health reform in Australia 1992-2012. *BMC Health Serv Res* 2015;15:479. doi: 10.1186/s12913-015-1142-3 [published Online First: 2015/10/27]
10. Arrow KJ. American economic review. *Uncertainty and the Welfare Economics of Medical Care* 1963;53(5):941-73.
11. Savedoff WD. Kenneth Arrow and the birth of health economics. *Bulletin of the World Health Organization* 2004;82(2):139-40.
12. Josephson E, Gergen J, Coe M, et al. How do performance-based financing programmes measure quality of care? A descriptive analysis of 68 quality checklists from 28 low- and middle-income countries. *Health Policy and Planning* 2017;32(8):1120-26. doi: 10.1093/heapol/czx053
13. Shroff ZC, Bigdeli M, Meessen B. From Scheme to System (Part 2): Findings from Ten Countries on the Policy Evolution of Results-Based Financing in Health Systems. *Health Systems & Reform* 2017;3(2):137-47. doi: 10.1080/23288604.2017.1304190

14. Singh NS, Kovacs RJ, Cassidy R, et al. A realist review to assess for whom, under what conditions and how pay for performance programmes work in low- and middle-income countries. *Social Science & Medicine* 2021;270:113624. doi: <https://doi.org/10.1016/j.socscimed.2020.113624>
15. Eijkenaar F, Emmert M, Scheppach M, et al. Effects of pay for performance in health care: a systematic review of systematic reviews. *Health Policy* 2013;110(2-3):115-30. doi: 10.1016/j.healthpol.2013.01.008 [published Online First: 2013/02/06]
16. Emmert M, Eijkenaar F, Kemter H, et al. Economic evaluation of pay-for-performance in health care: a systematic review. *Eur J Health Econ* 2012;13(6):755-67. doi: 10.1007/s10198-011-0329-8 [published Online First: 2011/06/11]
17. Vlaanderen FP, Tanke MA, Bloem BR, et al. Design and effects of outcome-based payment models in healthcare: a systematic review. *Eur J Health Econ* 2019;20(2):217-32. doi: 10.1007/s10198-018-0989-8 [published Online First: 2018/07/06]
18. Borghi J, Singh NS, Brown G, et al. Understanding for whom, why and in what circumstances payment for performance works in low and middle income countries: protocol for a realist review. *BMJ Glob Health* 2018;3(3):e000695. doi: 10.1136/bmjgh-2017-000695 [published Online First: 2018/07/11]
19. Cashin C, Chi Y-LS, P. C., Borowitz M, et al. Paying for Performance in Health Care: implications for health system performance and accountability. England: World Health Organization 2014.
20. Sutton M, Nikolova S, Boaden R, et al. Reduced mortality with hospital pay for performance in England. *N Engl J Med* 2012;367(19):1821-8. doi: 10.1056/NEJMsa1114951 [published Online First: 2012/11/09]
21. Kristensen SR, Meacock R, Turner AJ, et al. Long-term effect of hospital pay for performance on mortality in England. *N Engl J Med* 2014;371(6):540-8. doi: 10.1056/NEJMoa1400962 [published Online First: 2014/08/08]
22. Andrew MR. Effects of the Premier Hospital Quality Incentive Demonstration on Medicare patient mortality and cost. *Health Services Research* 2009 doi: 10.1111/j.1475-6773.2009.00956.x
23. Ashish KJ, Karen EJ, Orav EJ, et al. The Long-Term Effect of Premier Pay for Performance on Patient Outcomes. *The New England Journal of Medicine* 2012 doi: 10.1056/nejmsa1112351
24. Terry S, Lauren Hersch N, Jyothi RT, et al. Does pay-for-performance improve surgical outcomes? An evaluation of phase 2 of the Premier Hospital Quality Incentive Demonstration. *Annals of Surgery* 2014 doi: 10.1097/sla.0000000000000425
25. Ryan A, Blustein J, Casalino L. Improvement Among Low-Performing Hospitals Medicare 's Flagship Test Of Pay-For-Performance Did Not Spur More Rapid Quality. *Health Affairs* 2012
26. Andrew MR, Matt S, Tim D. Does Winning a Pay-for-Performance Bonus Improve Subsequent Quality Performance? Evidence from the Hospital Quality Incentive Demonstration. *Health Services Research* 2014 doi: 10.1111/1475-6773.12097

27. Ryan AM, Burgess JF, Jr., Pesko MF, et al. The early effects of Medicare's mandatory hospital pay-for-performance program. *Health Serv Res* 2015;50(1):81-97. doi: 10.1111/1475-6773.12206 [published Online First: 2014/07/22]
28. Figueroa JF, Tsugawa Y, Zheng J, et al. Association between the Value-Based Purchasing pay for performance program and patient mortality in US hospitals: observational study. *BMJ* 2016;353:i2214. doi: 10.1136/bmj.i2214 [published Online First: 2016/05/11]
29. Souvik B, Danny M, Michael KP-O, et al. Association between degree of exposure to the Hospital Value Based Purchasing Program and 30-day mortality: experience from the first four years of Medicare's pay-for-performance program. *BMC Health Services Research* 2019 doi: 10.1186/s12913-019-4562-7
30. Gilman M, Adams EK, Hockenberry JM, et al. Safety-net hospitals more likely than other hospitals to fare poorly under Medicare's value-based purchasing. *Health Aff (Millwood)* 2015;34(3):398-405. doi: 10.1377/hlthaff.2014.1059 [published Online First: 2015/03/04]
31. Hsueh-Fen C, Saleema AK, Fei W, et al. Financial Performance of Hospitals in the Mississippi Delta Region Under the Hospital Readmissions Reduction Program and Hospital Value-based Purchasing Program. *Medical Care* 2017 doi: 10.1097/mlr.0000000000000808
32. Aaron SK, Young-Rock H, Oliver N, et al. Early Performance of Hospital Value-based Purchasing Program in Medicare: A Systematic Review. *Medical Care* 2020 doi: 10.1097/mlr.0000000000001354
33. Nathan WC, Jan PC. Hospital Performance in the First 6 Years of Medicare's Value-Based Purchasing Program. *Medical Care Research and Review* 2020 doi: 10.1177/1077558720927586
34. Aaron S, Hanadi H, Jing X, et al. Race to the Top of the Hospital Value-Based Purchasing Program. *Journal of healthcare management / American College of Healthcare Executives* 2021 doi: 10.1097/jhm-d-20-00087
35. Adriana GR, Katherine MM, Katherine MM, et al. How Total Performance Scores of Medicare Value-Based Purchasing Program Hospitals Change Over Time. *Journal for Healthcare Quality* 2021 doi: 10.1097/jhq.0000000000000321
36. Lee R, Barbara L-O, John TL, et al. Evaluating the robustness of the CMS Hospital Value-Based Purchasing measurement system. *Health Services Research* 2021 doi: 10.1111/1475-6773.13608
37. Igna B, Jose FF, Jie Z, et al. Impact of Financial Incentives on Early and Late Adopters among US Hospitals: observational study. *BMJ* 2018 doi: 10.1136/bmj.j5622
38. Desai NR, Ross JS, Kwon JY, et al. Association Between Hospital Penalty Status Under the Hospital Readmission Reduction Program and Readmission Rates for Target and Nontarget Conditions. *JAMA* 2016;316(24):2647-56. doi: 10.1001/jama.2016.18533
39. Zuckerman RB, Sheingold SH, Orav EJ, et al. Readmissions, Observation, and the Hospital Readmissions Reduction Program. *New England Journal of Medicine* 2016;374(16):1543-51. doi: 10.1056/NEJMsa1513024

40. Wadhera RK, Joynt Maddox KE, Kazi DS, et al. Hospital revisits within 30 days after discharge for medical conditions targeted by the Hospital Readmissions Reduction Program in the United States: national retrospective analysis. *BMJ* 2019;366:14563. doi: 10.1136/bmj.l4563
41. Psotka MA, Fonarow GC, Allen LA, et al. The Hospital Readmissions Reduction Program: Nationwide Perspectives and Recommendations: A JACC: Heart Failure Position Paper. *JACC Heart Fail* 2020;8(1):1-11. doi: 10.1016/j.jchf.2019.07.012 [published Online First: 2019/10/14]
42. Chatterjee P, Joynt Maddox KE. US National Trends in Mortality From Acute Myocardial Infarction and Heart Failure: Policy Success or Failure? *JAMA Cardiology* 2018;3(4):336-40. doi: 10.1001/jamacardio.2018.0218
43. Dharmarajan K, Wang Y, Lin Z, et al. Association of Changing Hospital Readmission Rates With Mortality Rates After Hospital Discharge. *Jama* 2017;318(3):270-78. doi: 10.1001/jama.2017.8444 [published Online First: 2017/07/19]
44. Fonarow GC, Konstam MA, Yancy CW. The Hospital Readmission Reduction Program Is Associated With Fewer Readmissions, More Deaths: Time to Reconsider. *J Am Coll Cardiol* 2017;70(15):1931-34. doi: 10.1016/j.jacc.2017.08.046 [published Online First: 2017/10/07]
45. Gupta A, Allen LA, Bhatt DL, et al. Association of the Hospital Readmissions Reduction Program Implementation With Readmission and Mortality Outcomes in Heart Failure. *JAMA Cardiology* 2018;3(1):44-53. doi: 10.1001/jamacardio.2017.4265
46. Khera R, Dharmarajan K, Wang Y, et al. Association of the Hospital Readmissions Reduction Program With Mortality During and After Hospitalization for Acute Myocardial Infarction, Heart Failure, and Pneumonia. *JAMA Network Open* 2018;1(5):e182777-e77. doi: 10.1001/jamanetworkopen.2018.2777
47. Wadhera RK, Joynt Maddox KE, Wasfy JH, et al. Association of the Hospital Readmissions Reduction Program With Mortality Among Medicare Beneficiaries Hospitalized for Heart Failure, Acute Myocardial Infarction, and Pneumonia. *Jama* 2018;320(24):2542-52. doi: 10.1001/jama.2018.19232 [published Online First: 2018/12/24]
48. Khera R, Pandey A, Ayers CR, et al. Contemporary Epidemiology of Heart Failure in Fee-For-Service Medicare Beneficiaries Across Healthcare Settings. *Circ Heart Fail* 2017;10(11) doi: 10.1161/circheartfailure.117.004402 [published Online First: 2017/11/14]
49. Jha AK. Death, Readmissions, and Getting Policy Right. *JAMA Network Open* 2018;1(5):e182776-e76. doi: 10.1001/jamanetworkopen.2018.2776
50. Gupta A, Fonarow GC. The Hospital Readmissions Reduction Program-learning from failure of a healthcare policy. *Eur J Heart Fail* 2018;20(8):1169-74. doi: 10.1002/ejhf.1212 [published Online First: 2018/05/24]
51. Charles NK, Thomas A, Lisa P, et al. Assessing Medicare's Hospital Pay-For-Performance Programs And Whether They Are Achieving Their Goals. *Health Affairs* 2015 doi: 10.1377/hlthaff.2015.0158

52. Quality AfHRA. Quality Indicator User Guide: Patient Safety Indicators (PSI) Composite Measures, 2021.
53. Rajaram R, Barnard C, Bilimoria KY. Concerns About Using the Patient Safety Indicator-90 Composite in Pay-for-Performance Programs. *JAMA* 2015;313(9):897-98. doi: 10.1001/jama.2015.52
54. Kyle HS, Justin BD, Michael JE, et al. Hospital-Acquired Condition Reduction Program Is Not Associated With Additional Patient Safety Improvement. *Health Affairs* 2019 doi: 10.1377/hlthaff.2018.05504
55. Roshun S, Devraj S, Ushapoorna N, et al. Changes in hospital safety following penalties in the US Hospital Acquired Condition Reduction Program: retrospective cohort study. *BMJ* 2019 doi: 10.1136/bmj.l4109
56. Emily A, Justin BD, Ushapoorna N, et al. Changes in Hospital-acquired Conditions and Mortality Associated With the Hospital-acquired Condition Reduction Program. *Annals of Surgery* 2019 doi: 10.1097/sla.0000000000003641
57. Christina AM, Allison RD, Cynthia B, et al. Association Between Hospital Characteristics and Performance on the New Hospital-Acquired Condition Reduction Program's Surgical Site Infection Measures. *JAMA Surgery* 2016 doi: 10.1001/jamasurg.2016.0408
58. Jose FF, David EW, Ashish KJ. Characteristics of hospitals receiving the largest penalties by US pay-for-performance programmes. *BMJ Quality & Safety* 2016 doi: 10.1136/bmjqs-2015-005040
59. Cheryl KZ, Jyothi RT, Andrew MR, et al. Medicare's Hospital Acquired Condition Reduction Program Disproportionately Affects Minority-serving Hospitals: Variation by Race, Socioeconomic Status, and Disproportionate Share Hospital Payment Receipt. *Annals of Surgery* 2019 doi: 10.1097/sla.0000000000003564
60. Sukruth AS, Waken RJ, Alina AL, et al. Association of Stratification by Proportion of Patients Dually Enrolled in Medicare and Medicaid With Financial Penalties in the Hospital-Acquired Condition Reduction Program. *JAMA Internal Medicine* 2020 doi: 10.1001/jamainternmed.2020.7386
61. Boamah SA, Hamadi HY, Spaulding AC. Hospital-Acquired Conditions Reduction Program, Racial and Ethnic Diversity, and Magnet Designation in the United States. *J Patient Saf* 2022;18(7):e1090-e95. doi: 10.1097/pts.0000000000001014 [published Online First: 2022/05/10]
62. Hanadi H, Shalmali RB, LaRee M, et al. Hospital-Acquired Conditions Reduction Program, Patient Safety, and Magnet Designation in the United States. *Journal of Patient Safety* 2020 doi: 10.1097/pts.0000000000000628
63. Matthew RB, Kelly IS, Helen K, et al. Association of Magnet Nursing Status With Hospital Performance on Nationwide Quality Metrics. *Journal for Healthcare Quality* 2019 doi: 10.1097/jhq.0000000000000202
64. Ravi R, Ravi R, Jeanette WC, et al. Hospital Characteristics Associated With Penalties in the Centers for Medicare & Medicaid Services Hospital-Acquired Condition Reduction Program. *JAMA* 2015 doi: 10.1001/jama.2015.8609

65. Emily JL, Kyle HS, Andrew MR. Improving the Hospital-Acquired Condition Reduction Program Through Rulemaking. *JAMA Health Forum* 2020 doi: 10.1001/jamahealthforum.2020.0416
66. Lane K, Samuel AS, Berna D, et al. Complication Rates, Hospital Size, and Bias in the CMS Hospital-Acquired Condition Reduction Program. *American Journal of Medical Quality* 2017 doi: 10.1177/1062860616681840
67. Richard LF, Norbert G, Norbert G, et al. Is the CMS Hospital-Acquired Condition Reduction Program a Valid Measure of Hospital Performance? *American Journal of Medical Quality* 2017 doi: 10.1177/1062860616640883
68. Santé HAd. Dotation financière pour l'amélioration de la qualité 2023 [Available from: https://www.has-sante.fr/jcms/c_493937/fr/dotation-financiere-pour-l-amelioration-de-la-qualite accessed 25 July.
69. Santé HAd. Healthcare Quality and Safety Indicators 2019 [Available from: https://www.has-sante.fr/jcms/c_2044563/en/healthcare-quality-and-safety-indicators.
70. Benoît L, Shu J, Anne G, et al. Evaluation of the effects of the French pay-for-performance program-IFAQ pilot study. *International Journal for Quality in Health Care* 2017 doi: 10.1093/intqhc/mzx111
71. Richardson D, Tarnow-Mordi WO, Lee SK. Risk adjustment for quality improvement. *Pediatrics* 1999;103(1 Suppl E):255-65. [published Online First: 1999/01/26]
72. Hornbrook MC. Hospital case mix: its definition, measurement and use. Part II: Review of alternative measures. *Med Care Rev* 1982;39(2):73-123. [published Online First: 1983/02/05]
73. Hornbrook MC. Hospital case mix: its definition, measurement and use: Part I. The conceptual framework. *Med Care Rev* 1982;39(1):1-43. [published Online First: 1983/01/06]
74. Kramer AA, Higgins TL, Zimmerman JE. Intensive care unit readmissions in U.S. hospitals: patient characteristics, risk factors, and outcomes. *Crit Care Med* 2012;40(1):3-10. doi: 10.1097/CCM.0b013e31822d751e [published Online First: 2011/09/20]
75. Frost SA, Alexandrou E, Bogdanovski T, et al. Severity of illness and risk of readmission to intensive care: A meta-analysis. *Resuscitation* 2009;80(5):505-10. doi: <https://doi.org/10.1016/j.resuscitation.2009.02.015>
76. Bradley TB, Kominski GF. Contributions of case mix and intensity change to hospital cost increases. *Health Care Financ Rev* 1992;14(2):151-63. [published Online First: 1992/12/04]
77. Bjorvatn A. Private or public hospital ownership: Does it really matter? *Social Science & Medicine* 2018;196:166-74. doi: <https://doi.org/10.1016/j.socscimed.2017.11.038>
78. Street A, Sivey P, Mason A, et al. Are English treatment centres treating less complex patients? *Health Policy* 2010;94(2):150-57. doi: <https://doi.org/10.1016/j.healthpol.2009.09.013>
79. Moscone F, Siciliani L, Tosetti E, et al. Do public and private hospitals differ in quality? Evidence from Italy. *Regional Science and Urban Economics* 2020;83:103523. doi: <https://doi.org/10.1016/j.regsciurbeco.2020.103523>

80. Fetter RB, Shin Y, Freeman JL, et al. Case mix definition by diagnosis-related groups. *Med Care* 1980;18(2 Suppl):iii, 1-53. [published Online First: 1980/02/01]
81. Scheller-Kreinsen D. GA, Busse R. The ABC of DRGs. *EuroObserver*, 2009.
82. Dixon J. US health care. II: The cost problem. *Bmj* 1992;305(6858):878-80. doi: 10.1136/bmj.305.6858.878 [published Online First: 1992/10/10]
83. Or Z. Implementation of DRG Payment in France: Issues and recent developments. *Health Policy* 2014;117(2):146-50. doi: <https://doi.org/10.1016/j.healthpol.2014.05.006>
84. Yang CM, Reinke W. Feasibility and validity of International Classification of Diseases based case mix indices. *BMC Health Serv Res* 2006;6:125. doi: 10.1186/1472-6963-6-125 [published Online First: 2006/10/07]
85. Lichtig L. Hospital Information System for Case Mix Management.: New York: John Wiley and Sons Press 1986.
86. Services CfMM. Details for title: FY 2011 Final Rule Data Files 2011 [Available from: <https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/FY-2011-IPPS-Final-Rule-Home-Page-Items/CMS1237932.html>] accessed 15 June 2023.
87. Jencks SF, Williams MV, Coleman EA. Rehospitalizations among patients in the Medicare fee-for-service program. *N Engl J Med* 2009;360(14):1418-28. doi: 10.1056/NEJMsa0803563 [published Online First: 2009/04/03]
88. Kristensen SR, Bech M, Quentin W. A roadmap for comparing readmission policies with application to Denmark, England, Germany and the United States. *Health Policy* 2015;119(3):264-73. doi: 10.1016/j.healthpol.2014.12.009 [published Online First: 2014/12/31]
89. Donabedian A. The quality of care. How can it be assessed? *JAMA* 1988;260(12):1743-8. [published Online First: 1988/09/23]
90. Benbassat J, Taragin M. Hospital readmissions as a measure of quality of health care: advantages and limitations. *Arch Intern Med* 2000;160(8):1074-81. doi: 10.1001/archinte.160.8.1074 [published Online First: 2000/05/02]
91. Kim Y, Ejaz A, Xu L, et al. Understanding recurrent readmission after major surgery among patients with employer-provided health insurance. *Am J Surg* 2016;212(2):305-14.e2. doi: 10.1016/j.amjsurg.2016.01.028 [published Online First: 2016/05/10]
92. Gohil SK, Datta R, Cao C, et al. Impact of Hospital Population Case-Mix, Including Poverty, on Hospital All-Cause and Infection-Related 30-Day Readmission Rates. *Clin Infect Dis* 2015;61(8):1235-43. doi: 10.1093/cid/civ539 [published Online First: 2015/07/02]
93. Barnett ML, Hsu J, McWilliams JM. Patient Characteristics and Differences in Hospital Readmission Rates. *JAMA Intern Med* 2015;175(11):1803-12. doi: 10.1001/jamainternmed.2015.4660 [published Online First: 2015/09/15]
94. Calvillo-King L, Arnold D, Eubank KJ, et al. Impact of social factors on risk of readmission or mortality in pneumonia and heart failure: systematic review. *J Gen Intern Med* 2013;28(2):269-82. doi: 10.1007/s11606-012-2235-x [published Online First: 2012/10/12]

95. Friebel R, Hauck K, Aylin P, et al. National trends in emergency readmission rates: a longitudinal analysis of administrative data for England between 2006 and 2016. *BMJ Open* 2018;8(3):e020325-e25. doi: 10.1136/bmjopen-2017-020325
96. Roshanghalb A, Mazzali C, Lettieri E, et al. Stability over time of the "hospital effect" on 30-day unplanned readmissions: Evidence from administrative data. *Health Policy* 2021;125(10):1393-97. doi: 10.1016/j.healthpol.2021.07.009 [published Online First: 2021/08/08]
97. Tsai TC, Orav EJ, Joynt KE. Disparities in surgical 30-day readmission rates for Medicare beneficiaries by race and site of care. *Ann Surg* 2014;259(6):1086-90. doi: 10.1097/sla.0000000000000326 [published Online First: 2014/01/21]
98. Horwitz LI, Partovian C, Lin Z, et al. Development and use of an administrative claims measure for profiling hospital-wide performance on 30-day unplanned readmission. *Annals of internal medicine* 2014;161(10 Suppl):S66-S75. doi: 10.7326/M13-3000
99. Fekri O, Manukyan E, Klazinga N. Associations between hospital deaths (HSMR), readmission and length of stay (LOS): a longitudinal assessment of performance results and facility characteristics of teaching and large-sized hospitals in Canada between 2013-2014 and 2017-2018. *BMJ Open* 2021;11(2):e041648. doi: 10.1136/bmjopen-2020-041648 [published Online First: 2021/02/08]
100. Lindenauer PK, Bernheim SM, Grady JN, et al. The performance of US hospitals as reflected in risk-standardized 30-day mortality and readmission rates for medicare beneficiaries with pneumonia. *J Hosp Med* 2010;5(6):E12-8. doi: 10.1002/jhm.822 [published Online First: 2010/07/29]
101. Kansagara D, Englander H, Salanitro A, et al. Risk prediction models for hospital readmission: a systematic review. *Jama* 2011;306(15):1688-98. doi: 10.1001/jama.2011.1515 [published Online First: 2011/10/20]
102. Metcalfe D, Zogg CK, Judge A, et al. Pay for performance and hip fracture outcomes: an interrupted time series and difference-in-differences analysis in England and Scotland. *Bone Joint J* 2019;101-b(8):1015-23. doi: 10.1302/0301-620x.101b8.Bjj-2019-0173.R1 [published Online First: 2019/08/01]
103. Oakley B, Nightingale J, Moran C, et al. Does achieving the best practice tariff improve outcomes in hip fracture patients? An observational cohort study. *BMJ Open* 2017;7(2):e014190. doi: 10.1136/bmjopen-2016-014190
104. Balint E. The possibilities of patient-centered medicine. *J R Coll Gen Pract* 1969;17(82):269-76. [published Online First: 1969/05/01]
105. Balint M. The doctor, his patient, and the illness. *The Lancet* 1955;265(6866):683-88. doi: [https://doi.org/10.1016/S0140-6736\(55\)91061-8](https://doi.org/10.1016/S0140-6736(55)91061-8)
106. Engel GL. A unified concept of health and disease. *Perspect Biol Med* 1960;3:459-85. doi: 10.1353/pbm.1960.0020 [published Online First: 1960/01/01]
107. Griffiths R, Bett M, Blyth J, et al. NHS Management Inquiry. London, 1983.
108. Institute of Medicine Committee on Quality of Health Care in A. Crossing the Quality Chasm: A New Health System for the 21st Century. Washington (DC): National Academies Press (US) 2001.
109. World Health Organization. Framework on integrated, people-centred health services: report by the Secretariat. Geneva: World Health Assembly, 2016.

110. National Academies of Sciences E, Medicine, Health, et al. Crossing the Global Quality Chasm: Improving Health Care Worldwide. Washington (DC): National Academies Press (US) 2018.
111. World Health Organization, Organisation for Economic Co-operation and Development, The World Bank. Delivering quality health services: a global imperative for universal health coverage. Geneva, 2018.
112. Speakers and moderators at the Policy Forum on the Future of Health: statement by Donald M. Berwick.; 2017; Paris. Organisation for Economic Cooperation and Development.
113. Ware JE, Jr., Davies-Avery A, Stewart AL. The measurement and meaning of patient satisfaction. *Health Med Care Serv Rev* 1978;1(1):1, 3-15. [published Online First: 1977/12/13]
114. Batbaatar E, Dorjdagva J, Luvsannyam A, et al. Conceptualisation of patient satisfaction: a systematic narrative literature review. *Perspectives in Public Health* 2015;135(5):243-50. doi: 10.1177/1757913915594196 [published Online First: 2015/07/19]
115. Staniszevska S, Ahmed L. The concepts of expectation and satisfaction: do they capture the way patients evaluate their care? *J Adv Nurs* 1999;29(2):364-72. doi: 10.1046/j.1365-2648.1999.00897.x [published Online First: 1999/04/10]
116. Bowling A, Rowe G, Lambert N, et al. The measurement of patients' expectations for health care: a review and psychometric testing of a measure of patients' expectations. *Health Technol Assess* 2012;16(30):i-xii, 1-509. doi: 10.3310/hta16300 [published Online First: 2012/07/04]
117. Junewicz A, Youngner SJ. Patient-satisfaction surveys on a scale of 0 to 10: improving health care, or leading it astray? *Hastings Cent Rep* 2015;45(3):43-51. doi: 10.1002/hast.453 [published Online First: 2015/03/11]
118. Ng JHY, Luk BHK. Patient satisfaction: Concept analysis in the healthcare context. *Patient Educ Couns* 2019;102(4):790-96. doi: 10.1016/j.pec.2018.11.013 [published Online First: 2018/11/28]
119. Beattie M, Murphy DJ, Atherton I, et al. Instruments to measure patient experience of healthcare quality in hospitals: a systematic review. *Syst Rev* 2015;4:97. doi: 10.1186/s13643-015-0089-0 [published Online First: 2015/07/24]
120. HCAHPS: Patients' Perspectives of Care Survey: Centers for Medicare & Medicaid Services; 2023 [Available from: <https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/HospitalHCAHPS> accessed 15 June 2023.
121. Nicholas C, Rahul A, Yang S, et al. Association of the Medicare Value-Based Purchasing Program With Changes in Patient Care Experience at Safety-net vs Non-Safety-net Hospitals. *JAMA health forum* 2022 doi: 10.1001/jamahealthforum.2022.1956
122. Khic-Houy P, Rachel C, Marie MB, et al. Associations between patient experiences and clinical outcomes: a cross-sectional data linkage study of the Australian private healthcare sector. *BMJ Open Quality* 2019 doi: 10.1136/bmjog-2019-000637

123. Carter J, Ward C, Wexler D, et al. The association between patient experience factors and likelihood of 30-day readmission: a prospective cohort study. *BMJ Qual Saf* 2018;27(9):683-90. doi: 10.1136/bmjqs-2017-007184 [published Online First: 2017/11/18]
124. Wang DE, Tsugawa Y, Figueroa JF, et al. Association Between the Centers for Medicare and Medicaid Services Hospital Star Rating and Patient Outcomes. *JAMA Intern Med* 2016;176(6):848-50. doi: 10.1001/jamainternmed.2016.0784 [published Online First: 2016/04/12]
125. Kemp KA, Santana MJ, Southern DA, et al. Association of inpatient hospital experience with patient safety indicators: a cross-sectional, Canadian study. *BMJ Open* 2016;6(7):e011242. doi: 10.1136/bmjopen-2016-011242 [published Online First: 2016/07/03]
126. Kennedy GD, Tevis SE, Kent KC. Is there a relationship between patient satisfaction and favorable outcomes? *Ann Surg* 2014;260(4):592-8; discussion 98-600. doi: 10.1097/sla.0000000000000932 [published Online First: 2014/09/10]
127. Hachem F CJ, Fullam F, Gallan AS, Hohmann S, Johnson C. The relationships between HCAHPS communication and discharge satisfaction items and hospital readmissions. *Patient Experience Journal* 2014;1(2):71-77. doi: 10.35680/2372-0247.1022
128. Doyle C, Lennox L, Bell D. A systematic review of evidence on the links between patient experience and clinical safety and effectiveness. *BMJ Open* 2013;3(1) doi: 10.1136/bmjopen-2012-001570 [published Online First: 2013/01/08]
129. Irwin P, Irwin P, Francis F, et al. Patient satisfaction in pay for performance programs. *Quality management in health care* 2011 doi: 10.1097/qmh.0b013e318213aed0
130. Patient Satisfaction Surveys. *Catalyst Carryover* 2018;4(1) doi: doi:10.1056/CAT.18.0288
131. Bastemeijer CM, Boosman H, van Ewijk H, et al. Patient experiences: a systematic review of quality improvement interventions in a hospital setting. *Patient Relat Outcome Meas* 2019;10:157-69. doi: 10.2147/prom.S201737 [published Online First: 2019/06/14]
132. Banda S, Nkungula N, Chiumia IK, et al. Tools for measuring client experiences and satisfaction with healthcare in low- and middle-income countries: a systematic review of measurement properties. *BMC Health Services Research* 2023;23(1):133. doi: 10.1186/s12913-023-09129-9
133. Bate P, Robert G. Experience-based design: from redesigning the system around the patient to co-designing services with the patient. *Qual Saf Health Care* 2006;15(5):307-10. doi: 10.1136/qshc.2005.016527 [published Online First: 2006/11/01]
134. Yt M. Myth: High patient satisfaction means high-quality care. *Journal of Health Services Research and Policy* 2013 doi: 10.1177/1355819613501426

135. Papanicolas I, Figueroa JF, Orav EJ, et al. Patient Hospital Experience Improved Modestly, But No Evidence Medicare Incentives Promoted Meaningful Gains. *Health Aff (Millwood)* 2017;36(1):133-40. doi: 10.1377/hlthaff.2016.0808 [published Online First: 2017/01/11]
136. Ryan AM, Krinsky S, Maurer KA, et al. Changes in Hospital Quality Associated with Hospital Value-Based Purchasing. *N Engl J Med* 2017;376(24):2358-66. doi: 10.1056/NEJMsa1613412 [published Online First: 2017/06/15]
137. Joynt KE, Jha AK. Characteristics of hospitals receiving penalties under the Hospital Readmissions Reduction Program. *Jama* 2013;309(4):342-3. doi: 10.1001/jama.2012.94856 [published Online First: 2013/01/24]
138. Weeks JC, Catalano PJ, Cronin A, et al. Patients' expectations about effects of chemotherapy for advanced cancer. *N Engl J Med* 2012;367(17):1616-25. doi: 10.1056/NEJMoa1204410 [published Online First: 2012/10/26]
139. Fenton JJ, Jerant AF, Bertakis KD, et al. The cost of satisfaction: a national study of patient satisfaction, health care utilization, expenditures, and mortality. *Arch Intern Med* 2012;172(5):405-11. doi: 10.1001/archinternmed.2011.1662 [published Online First: 2012/02/15]
140. Davidson KW, Shaffer J, Ye S, et al. Interventions to improve hospital patient satisfaction with healthcare providers and systems: a systematic review. *BMJ Qual Saf* 2017;26(7):596-606. doi: 10.1136/bmjqs-2015-004758 [published Online First: 2016/08/05]
141. Malpani R, Adrados M, Mercier MR, et al. Characteristics and Predictors of HCAHPS Nonresponse After Spine Surgery. *Spine (Phila Pa 1976)* 2020;45(8):E448-e56. doi: 10.1097/brs.0000000000003287 [published Online First: 2019/10/15]
142. Mercier MR, Pathak N, Adrados M, et al. Patient Factors and Perioperative Outcomes Affect Hospital Consumer Assessment of Healthcare Providers and Systems Survey Response Rates After Primary Total Hip Replacement. *J Am Acad Orthop Surg Glob Res Rev* 2021;5(4) doi: 10.5435/JAAOSGlobal-D-21-00052 [published Online First: 2021/04/03]
143. Dunsch F, Evans DK, Macis M, et al. Bias in patient satisfaction surveys: a threat to measuring healthcare quality. *BMJ Global Health* 2018;3(2):e000694. doi: 10.1136/bmjgh-2017-000694
144. Nations-ESCWA U. Multidimensional poverty in Lebanon (2019-2021): Painful reality and uncertain prospects, 2021.
145. Ammar W. Health beyond politics. Health beyond politics. Lebanon: World Health Organization, Eastern Mediterranean Regional Office 2009.
146. Ammar W, Wakim IR, Hajj I. Accreditation of hospitals in Lebanon: a challenging experience. *East Mediterr Health J* 2007;13(1):138-49. [published Online First: 2007/06/06]
147. Ammar W, Khalife J, El-Jardali F, et al. Hospital accreditation, reimbursement and case mix: links and insights for contractual systems. *BMC Health Serv Res* 2013;13:505. doi: 10.1186/1472-6963-13-505 [published Online First: 2013/12/07]

148. Joynt Maddox KE, Sen AP, Samson LW, et al. Elements of Program Design in Medicare's Value-based and Alternative Payment Models: a Narrative Review. *J Gen Intern Med* 2017;32(11):1249-54. doi: 10.1007/s11606-017-4125-8 [published Online First: 2017/07/19]
149. McCleary R, McDowall D, Bartos B. Design and analysis of time series experiments. New York, NY: Oxford University Press 2017.
150. Bhopal R. Concepts of Epidemiology. Oxford: Oxford University Press 2016.
151. Respirator Standards, The National Personal Protective Technology Laboratory: National Institute for Occupational Safety and Health; [Available from: <https://www.cdc.gov/niosh/npptl/RespStds.html>] accessed 1 July 2023.
152. GB 2626-2019 Respiratory protection: Non-powered air-purifying particle respirator: State Administration for Market Regulation, Standardization Administration of China; [Available from: <https://openstd.samr.gov.cn/bzgk/gb/newGbInfo?hcno=16D8935B45AD7AE40228801B7FADFC6C>] accessed 1 July 2023.
153. Jimenez JL, Marr LC, Randall K, et al. What were the historical reasons for the resistance to recognizing airborne transmission during the COVID-19 pandemic? *Indoor Air* 2022;32(8):e13070. doi: <https://doi.org/10.1111/ina.13070>
154. Leo BF, Lin CY, Markandan K, et al. An overview of SARS-CoV-2 transmission and engineering strategies to mitigate risk. *Journal of Building Engineering* 2023;106737. doi: <https://doi.org/10.1016/j.jobe.2023.106737>
155. Wang CC, Prather KA, Sznitman J, et al. Airborne transmission of respiratory viruses. *Science* 2021;373(6558):eabd9149. doi: 10.1126/science.abd9149
156. Greenhalgh T, Jimenez JL, Prather KA, et al. Ten scientific reasons in support of airborne transmission of SARS-CoV-2. *The Lancet* 2021;397(10285):1603-05. doi: 10.1016/S0140-6736(21)00869-2
157. Cheng Y, Ma N, Witt C, et al. Face masks effectively limit the probability of SARS-CoV-2 transmission. *Science* 2021;372(6549):1439-43. doi: doi:10.1126/science.abg6296
158. Hill AB. The environment and disease: association or causation? *Proc R Soc Med* 1965;58(5):295-300.
159. Moyses Szklo FJN. Epidemiology: Beyond the basics. Ontario: Jones and Bartlett 2007.
160. Porta M. A Dictionary of Epidemiology. 6th ed: Oxford University Press 2014.
161. Katz M. Evaluating clinical and public health evaluations. Cambridge 2010.
162. Gordis L. Epidemiology. 4th edition ed. Philadelphia: Saunders, Elsevier 2009.
163. David Savitz GW. Interpreting Epidemiologic Evidence. 2nd edition ed. Oxford: Oxford University Press 2016.
164. Noel S, Weiss TDK. Epidemiologic Methods: Studying the Occurrence of Illness. 2nd ed. Oxford: Oxford University Press 2014.
165. Campbell DT. Assessing the impact of planned social change. *Evaluation and Program Planning* 1979;2(1):67-90. doi: [https://doi.org/10.1016/0149-7189\(79\)90048-X](https://doi.org/10.1016/0149-7189(79)90048-X)

166. Bernal JL, Cummins S, Gasparrini A. Interrupted time series regression for the evaluation of public health interventions: a tutorial. *International Journal of Epidemiology* 2016;46(1):348-55. doi: 10.1093/ije/dyw098
167. Turner SL, Karahalios A, Forbes AB, et al. Comparison of six statistical methods for interrupted time series studies: empirical evaluation of 190 published series. *BMC Medical Research Methodology* 2021;21(1):134. doi: 10.1186/s12874-021-01306-w
168. Graneheim UH, Lundman B. Qualitative content analysis in nursing research: concepts, procedures and measures to achieve trustworthiness. *Nurse Educ Today* 2004;24(2):105-12. doi: 10.1016/j.nedt.2003.10.001 [published Online First: 2004/02/11]
169. Krippendorff K. *Content Analysis: An Introduction to Its Methodology*. Fourth Edition ed. Thousand Oaks, California: SAGE Publications, Inc., 2019.
170. Nancy LK, Nancy SW, Daniel RA, et al. Content analysis: review of methods and their applications in nutrition education. *Journal of Nutrition Education and Behavior* 2002 doi: 10.1016/s1499-4046(06)60097-3
171. Michael Quinn P. *How to use qualitative methods in evaluation*. 1987
172. Paley J. *Phenomenology as qualitative research. a critical analysis of meaning attribution*: Routledge, Taylor & Francis Group 2017.
173. Urquhart C. *Grounded theory for qualitative research. a practical guide*: SAGE 2013.
174. Krueger R. *Focus Groups: A Practical Guide for Applied Research*. 2nd ed. Thousand Oaks, California: SAGE 1994.
175. Hennink M. *International Focus Group Research*. Cambridge: Cambridge University Press 2007.
176. Leavy P. *Research design. [Elektronisk resurs] quantitative, qualitative, mixed methods, arts-based, and community-based participatory research approaches*: Guilford Press 2017.
177. Johnson RB, Onwuegbuzie AJ. Mixed Methods Research: A Research Paradigm Whose Time Has Come. *Educational Researcher* 2004;33(7):14-26.
178. Outcome and Payment Measures: Centers for Medicare & Medicaid Services; 2023 [Available from: <https://www.cms.gov/medicare/quality-initiatives-patient-assessment-instruments/hospitalqualityinits/outcomeandpaymentmeasures> accessed 5 July 2023.
179. Krumholz HM, Brindis RG, Brush JE, et al. Standards for statistical models used for public reporting of health outcomes: an American Heart Association Scientific Statement from the Quality of Care and Outcomes Research Interdisciplinary Writing Group: cosponsored by the Council on Epidemiology and Prevention and the Stroke Council. Endorsed by the American College of Cardiology Foundation. *Circulation* 2006;113(3):456-62. doi: 10.1161/circulationaha.105.170769 [published Online First: 2005/12/21]
180. Burton ML, Romney AK. A multidimensional representation of role terms. *American Ethnologist* 1975;2(3):397-407. doi: <https://doi.org/10.1525/ae.1975.2.3.02a00020>
181. Shipstone EI. Some variables affecting pattern conception. *Psychological Monographs: General and Applied* 1960;74, No. 17(Whole No. 504)

182. Yeh HW, Gajewski BJ, Perdue DG, et al. Sorting it Out: Pile Sorting as a Mixed Methodology for Exploring Barriers to Cancer Screening. *Qual Quant* 2014;48(5):2569-87. doi: 10.1007/s11135-013-9908-3 [published Online First: 2014/08/22]
183. Quintiliani LM, Campbell MK, Haines PS, et al. The Use of the Pile Sort Method in Identifying Groups of Healthful Lifestyle Behaviors among Female Community College Students. *Journal of the American Dietetic Association* 2008;108(9):1503-07. doi: <https://doi.org/10.1016/j.jada.2008.06.428>
184. Fitzpatrick R. Surveys of patients satisfaction: I--Important general considerations. *Bmj* 1991;302(6781):887-9. doi: 10.1136/bmj.302.6781.887 [published Online First: 1991/04/13]
185. Giordano LA, Elliott MN, Goldstein E, et al. Development, implementation, and public reporting of the HCAHPS survey. *Med Care Res Rev* 2010;67(1):27-37. doi: 10.1177/1077558709341065 [published Online First: 2009/07/30]
186. Collaborative Institutional Training Initiative (CITI Program) [Available from: <http://www.citiprogram.org/> accessed 7 August 2023.
187. Ederer F. Randomized controlled clinical trial. National Eye Institute workshop for ophthalmologists. Why do we need controls? Why do we need to randomize? *Am J Ophthalmol* 1975;79(5):758-62. doi: 10.1016/0002-9394(75)90733-3 [published Online First: 1975/05/01]
188. To T, Feldman W, Young W, et al. Hospitalization rates of children with gastroenteritis in Ontario. *Can J Public Health* 1996;87(1):62-5. [published Online First: 1996/01/01]
189. Kronfol N, Khalife J, Romanos J, et al. Observations on the Ministry of Public Health program of support to the hospitalization of patients in Lebanon. *J Med Liban* 2014;62(1):33-9. [published Online First: 2014/04/02]
190. Ginsburg PB, Carter GM. Medicare case-mix index increase. *Health Care Financ Rev* 1986;7(4):51-65. [published Online First: 1987/02/04]
191. Goldfarb MG, Coffey RM. Change in the Medicare case-mix index in the 1980s and the effect of the prospective payment system. *Health Serv Res* 1992;27(3):385-415. [published Online First: 1992/08/01]
192. Borghi J, Little R, Binyaruka P, et al. In Tanzania, The Many Costs Of Pay-For-Performance Leave Open To Debate Whether The Strategy Is Cost-Effective. *Health Affairs* 2015;34(3):406-14. doi: 10.1377/hlthaff.2014.0608
193. Soucat A, Dale E, Mathauer I, et al. Pay-for-Performance Debate: Not Seeing the Forest for the Trees. *Health Systems & Reform* 2017;3(2):74-79. doi: 10.1080/23288604.2017.1302902
194. Meessen B, Shroff ZC, Ir P, et al. From Scheme to System (Part 1): Notes on Conceptual and Methodological Innovations in the Multicountry Research Program on Scaling Up Results-Based Financing in Health Systems. *Health Systems & Reform* 2017;3(2):129-36. doi: 10.1080/23288604.2017.1303561

195. Fisher ES, Wennberg JE, Stukel TA, et al. Associations among hospital capacity, utilization, and mortality of US Medicare beneficiaries, controlling for sociodemographic factors. *Health Serv Res* 2000;34(6):1351-62. [published Online First: 2000/02/02]
196. Horwitz LI, Bernheim SM, Ross JS, et al. Hospital Characteristics Associated With Risk-standardized Readmission Rates. *Med Care* 2017;55(5):528-34. doi: 10.1097/mlr.0000000000000713 [published Online First: 2017/03/21]
197. Marang-van de Mheen PJ, Putter H, Bastiaannet E, et al. Competing risks in quality and safety research: a framework to guide choice of analysis and improve reporting. *BMJ Quality & Safety* 2021;30(12):1031-37. doi: 10.1136/bmjqs-2021-012988 [published Online First: 2021/09/19]
198. Batbaatar E, Dorjdagva J, Luvsannyam A, et al. Determinants of patient satisfaction: a systematic review. *Perspectives in Public Health* 2016;137(2):89-101. doi: 10.1177/17579139166634136
199. Deledda G, Moretti F, Rimondini M, et al. How patients want their doctor to communicate. A literature review on primary care patients' perspective. *Patient Education and Counseling* 2013;90(3):297-306. doi: <https://doi.org/10.1016/j.pec.2012.05.005>
200. Luthy C, Cedraschi C, Perrin E, et al. How do patients define "good" and "bad" doctors? *Swiss Med Wkly* 2005;135(5-6):82-6. [published Online First: 2005/02/25]
201. Borracci RA, Álvarez Gallesio JM, Ciambrone G, et al. What patients consider to be a good doctor, and what doctors consider to be a good patient. *Revista médica de Chile* 2020;148:930-38.
202. Quirk M, Mazor K, Haley H-L, et al. How patients perceive a doctor's caring attitude. *Patient Education and Counseling* 2008;72(3):359-66. doi: <https://doi.org/10.1016/j.pec.2008.05.022>
203. Wensing M, Jung HP, Mainz J, et al. A systematic review of the literature on patient priorities for general practice care. Part 1: Description of the research domain. *Soc Sci Med* 1998;47(10):1573-88. doi: 10.1016/s0277-9536(98)00222-6 [published Online First: 1998/11/21]
204. HealthWatch-England. What do people want from the NHS and social care in the future, 2018.
205. Saman DM, Kavanagh KT, Johnson B, et al. Can Inpatient Hospital Experiences Predict Central Line-Associated Bloodstream Infections? *PLOS ONE* 2013;8(4):e61097. doi: 10.1371/journal.pone.0061097
206. Rapport F, Hibbert P, Baysari M, et al. What do patients really want? An in-depth examination of patient experience in four Australian hospitals. *BMC Health Serv Res* 2019;19(1):38. doi: 10.1186/s12913-019-3881-z [published Online First: 2019/01/17]
207. Schoenfelder T, Klewer J, Kugler J. Determinants of patient satisfaction: a study among 39 hospitals in an in-patient setting in Germany. *Int J Qual Health Care* 2011;23(5):503-9. doi: 10.1093/intqhc/mzr038 [published Online First: 2011/07/01]

208. Berkowitz B. The Patient Experience and Patient Satisfaction: Measurement of a Complex Dynamic. *Online J Issues Nurs* 2016;21(1):1. doi: 10.3912/OJIN.Vol21No01Man01 [published Online First: 2016/11/18]
209. Smith PC, Sagan A, Siciliani L, et al. European Observatory Policy Briefs. Building on value-based health care: Towards a health system perspective. Copenhagen (Denmark): World Health Organization 2020.
210. European Commission. Defining Value in 'Value-Based Healthcare'. Report of the Expert Panel on effective ways of investing in Health., 2019.
211. Sabra AI. The Optics Of Ibn Al Haytham, Books I-III, On Direct Vision. Translated With Introduction And Commentary. London: The Warburg Institute, University of London 1989.
212. Renmans D, Holvoet N, Orach CG, et al. Opening the ‘black box’ of performance-based financing in low- and lower middle-income countries: a review of the literature. *Health Policy and Planning* 2016;31(9):1297-309. doi: 10.1093/heapol/czw045
213. de Walque D, Robyn PJ, Saidou H, et al. Looking into the performance-based financing black box: evidence from an impact evaluation in the health sector in Cameroon. *Health Policy and Planning* 2021;36(6):835-47. doi: 10.1093/heapol/czab002
214. Anselmi L, Binyaruka P, Borghi J. Understanding causal pathways within health systems policy evaluation through mediation analysis: an application to payment for performance (P4P) in Tanzania. *Implementation Science* 2017;12(1):10. doi: 10.1186/s13012-016-0540-1

Paper 1





Hospital Contracting Reforms: The Lebanese Ministry of Public Health Experience

Jade Khalife, Nadwa Rafeh, Jihad Makouk, Fadi El-Jardali, Bjorn Ekman, Nabil Kronfol, Ghassan Hamadeh & Walid Ammar

To cite this article: Jade Khalife, Nadwa Rafeh, Jihad Makouk, Fadi El-Jardali, Bjorn Ekman, Nabil Kronfol, Ghassan Hamadeh & Walid Ammar (2017) Hospital Contracting Reforms: The Lebanese Ministry of Public Health Experience, *Health Systems & Reform*, 3:1, 34-41, DOI: [10.1080/23288604.2016.1272979](https://doi.org/10.1080/23288604.2016.1272979)

To link to this article: <https://doi.org/10.1080/23288604.2016.1272979>

 Published online: 31 Jan 2017.

 Submit your article to this journal [↗](#)

 Article views: 4292

 View related articles [↗](#)

 View Crossmark data [↗](#)

 Citing articles: 10 View citing articles [↗](#)

Research Article

Hospital Contracting Reforms: The Lebanese Ministry of Public Health Experience

Jade Khalife^{1,2,3}, Nadwa Rafeh⁴, Jihad Makouk¹, Fadi El-Jardali³, Bjorn Ekman ², Nabil Kronfol³, Ghassan Hamadeh⁵ and Walid Ammar^{1,3,*}

¹Ministry of Public Health, Beirut, Lebanon

²Department of Clinical Sciences, Lund University, Malmö, Sweden

³Department of Health Management and Policy, Faculty of Health Sciences, American University of Beirut, Beirut, Lebanon

⁴World Bank Group, Washington, DC, USA

⁵Department of Family Medicine, Faculty of Medicine, American University of Beirut, Beirut, Lebanon

CONTENTS

Introduction and Health Sector Overview

Hospital Sector Overview

MoPH Hospital Contracting Reforms

Strengthening the Utilization Review Function

Developing Admission Criteria

Developing Mixed-Model Hospital Contracting

Moving Forward

References

Abstract—Since 2009, the Ministry of Public Health (MoPH) in Lebanon has been going through a major reform initiative to improve its contracting system with private and public hospitals. The private sector is the main provider of hospital care in the country and the main contractor to the MoPH for the provision of curative care. As an “insurer of last resort,” the MoPH plays an important role in providing hospital coverage to 53% of the population who lack coverage by private or public insurance schemes, through contractual arrangements with the private sector. Historically, the MoPH used hospital accreditation as the basis for contracting and for determining the reimbursement rate. However, recent studies by the MoPH showed that reimbursing hospitals solely on accreditation results was not appropriate and led to an unfair and inefficient reimbursement system. The reform program included the development of several components, in particular, an automated billing system, a utilization review function, standardized admission criteria, and a hospital case mix index that accounts for case complexity. In 2014, the MoPH started implementing a new mixed-model contracting system with private and public hospitals. Preliminary evaluation of the new model suggests that the system incentivized hospitals to admit fewer inappropriate cases and more cases that are more complex/serious. This article shares one experience of how to introduce a merit-based system to face the common practice of political clientelism and confessional/religious-based favoritism in Lebanon. It highlights the importance of stakeholder engagement in a framework of networking and participatory governance that proved to be a key element behind the resilience of a diversified health system.

Keywords: casemix, contracting, governance, health reform

Received 22 August 2016; revised 3 December 2016; accepted 3 December 2016.

*Correspondence to: Walid Ammar; Email: mphealth@cyberia.net.lb

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/khsr.

INTRODUCTION AND HEALTH SECTOR OVERVIEW

Lebanon is a middle-income country with a population of about 4.5 million citizens and 1.5 million refugees, the

majority of the latter group being due to the conflict in neighboring Syria.¹ Almost 88% of the Lebanese population resides in urban areas, two thirds of which is between the ages of 15 and 64 years.¹ Since the mid-1970s, Lebanon has been grappling with a myriad of political and socioeconomic hardships due to prolonged periods of conflict, including 15 years of civil unrest, war in 2006, and lately fallout from the conflict in Syria. The fragile and conflict-prone environment has led to political instability, shrinking economy, limited governance capacity, and inadequate public services across sectors.

The long history of conflict contributed to the weakening of the public sector and to the development of private institutions and nongovernmental organizations (NGOs). Accordingly, the private sector became the main provider of health care services in the country and the main contractor to the Ministry of Public Health (MoPH) for the provision of curative care. The unregulated growth led to the oversupply of services and created supply-induced demand with implications on the quality of care.^{2,3} Consequently, more dedicated efforts were required on behalf of the ministry to strengthen its regulatory role and build strong partnerships with the private sector.

Despite the chronic instability and constraints, Lebanon's health outcomes compare favorably with other countries in the Middle East North Africa (MENA) region as well as middle-income countries in other regions. Life expectancy at birth is 80 years, compared to 68 years for the region and 74 years for middle-income countries.⁴ Lebanon was one of the few countries to have achieved both Millennium Development Goals 4 and 5, namely, reducing maternal and child mortality. Between 1990 and 2012, the infant mortality ratio decreased from 32 to 9 per 1,000 live births, and the maternal mortality ratio dropped from 64 to 16 per 100,000 live births.⁴ These improvements were achieved while decreasing total health expenditures from 12.4% in 1998 to 6.4% in 2012.⁵ Lebanon was recognized by the World Health Organization as having been able to remarkably improve the health status of its population while significantly decreasing the gross domestic product share of health expenditures, mainly by decreasing individual out-of-pocket spending on health care.⁵ Lebanon is also well rated in terms of cost-effectiveness. According to a 2014 global comparative study by the *Economist*, Lebanon was ranked 32 among 166 countries in terms of outcomes with a remarkably lower cost per outcome than the countries in the same category.⁶

As such, the health system in Lebanon proved to be resilient to the massive influx of refugees and to the rapid increase of its population by 30%. This was largely attributed

to the governance arrangements led by the MoPH based on collaborative networking with the private sector and civil society organizations. Such governance was instrumental in overcoming the disadvantages of fragmentation of the health system while benefiting from the diversity of stakeholders involved.⁷

This article discusses the experience of the MoPH in the reform initiatives to improve governance and regulation of Lebanon's hospital care in particular, including hospital contracting reforms between the MoPH and public and private hospitals. The focus of this experience is of major relevance for the health sector in Lebanon, given the prominent role of the MoPH as payer for hospital care, as well as that of private (and public) hospitals as providers. It also describes the approach of the MoPH to address the limitations due to asymmetric information in this payer-provider relationship and to improve the effectiveness and efficiency of health services while countering favoritism and clientelism with a merit-based system.

HOSPITAL SECTOR OVERVIEW

The Lebanese health system is highly diverse, including a mix of public and private payers and providers. Health financing is mobilized from a range of resources, including general government revenues, social security contributions, and the private sector. Total expenditures on health constitute 6.4% of national gross domestic product, 40% of which is accounted for by hospitals alone.^{5,8} The private sector also accounts for 71% of health care financing, of which 37% is out-of-pocket payments made by households.⁸ Though the public sector is the main payer of hospital care, the private sector dominates in terms of service provision. There are 165 hospitals in Lebanon, 82% of which are privately owned and managed by physicians or by charitable organizations, often religious.⁸ Both public and private hospitals have similar average bed capacities. Public hospitals operate under a semi-autonomous model with hospital boards composed of various stakeholders involved, thus having a certain degree of autonomy.

Around 47% of Lebanese citizens have health insurance coverage. About 23% of those are covered by the National Social Security Fund, 9% by military schemes, 7% by private insurance, 4% by the Civil Servants Cooperative, and 4% by other schemes.² The remaining 53% lack any formal coverage and are covered by the MoPH, which serves as an "insurer of last resort."⁹ This has meant a strong role for the ministry not only in preventive care, public health leadership, and regulation but also in curative care.

To provide hospital coverage to about 250,000 cases per year, the MoPH contracts 26 public and 105 private hospitals. Individual patient copayment to the hospital constitutes 5% (public hospital) or 15% (private hospital) of the hospitalization costs, with the MoPH directly reimbursing the hospital for the 85%–95% difference. A ministerial waiver is also applied in specific cases where the patient is unable to afford copayment. As such, the MoPH is the main financier of private hospitals, allocating about 64% of its total annual budget (of about 367 million USD) for hospitalization coverage in 2012.¹⁰

MoPH HOSPITAL CONTRACTING REFORMS

Over the years, the MoPH has undertaken several reforms of its hospital contracting system, in line with its broad approach emphasizing participatory governance and involving stakeholders in policy making (Figure 1). The first reforms began with the development of a hospital rating system in 1983, leading to linking the ratings to hospital reimbursement in the 1990s and finally the adoption of hospital accreditation in 2001. Since then, the accreditation program has completed three cycles and has expanded to include numerous medical specialties, patient care, performance appraisals, staff competency testing, and appropriateness and implementation of policies and procedures. It remains a key part of the hospital contracting system.

Between 2001 and 2014, the reimbursement rate was determined solely by the hospital’s results in the most recent accreditation cycle. A three-tiered reimbursement rate was used, with about 10%–30% difference depending on whether the case is a medical or surgical (procedure) admission. This created a strong incentive for hospitals to have a higher reimbursement tier. The MoPH reimburses hospitals for surgical procedures using a predefined flat rate, whereas for medical cases a fee-for-service structure is used, where the latter is more likely to lead to overutilization and inappropriate admissions.

It became apparent to both MoPH and hospitals that using hospital accreditation as the sole determinant of the reimbursement rates is limiting. This was prompted by the heterogeneity among hospitals and the difficulty in developing a composite score for the increasing number of accreditation standards. The interest of all involved parties to go beyond accreditation, while still maintaining it as an integral component for contracting, formed a favorable environment within which to undertake new contracting reforms.

With grant funding from the World Bank, in 2009 the MoPH embarked on a new project to improve its hospital contracting system. The desire for a new reform was accompanied by a broad vision to improve the hospital reimbursement system, increase MoPH hospitalization coverage despite limited resources, and improve the quality of hospital care. More specifically, the project focused on three main goals: improve the appropriateness and fairness of contracting between MoPH and hospitals, improve the efficiency of MoPH spending, and promote good provider practices and discourage misuse or abuse of services.

To achieve the three interrelated goals, the MoPH first formed three committees: Utilization Review Committee, Admission Criteria Committee, and Performance Contracting Committee, each focusing on one of the goals yet working synergistically together. Each committee was composed of a mix of medical, public health, and information technology (IT) professionals whose affiliations included the MoPH, universities, and hospitals. The multidisciplinary composition of the committees provided strong credibility with which to address the goals of the program. The primary task of each committee was to discuss and identify objectives and activities to be undertaken in the new reform. It is relevant to note that the broad setting of the goals was an opportunity for the committees to assess the situation, identify gaps, and explore a wide variety of improvement interventions. Throughout the first year, committee meetings were held on

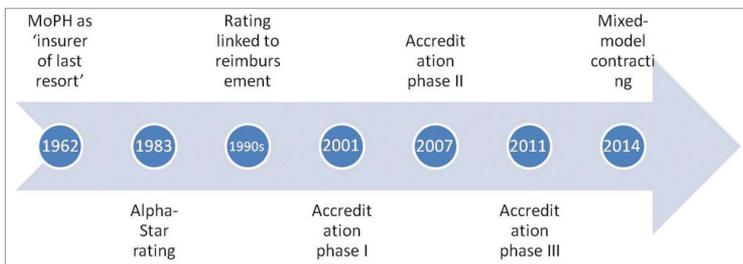


FIGURE 1. Timeline of MoPH Hospital Contracting Reforms

a monthly basis and somewhat less frequently afterwards. The committees were kept abreast of discussions among them by a few common members and by intercommittee events held semi-annually.

The initial phase also included a considerable amount of data extraction and analysis of MoPH hospitalization database. This provided more in-depth understanding of the frequency of admissions and costs across medical and surgical conditions, as well as variations among hospitals and regions. This work was done in close collaboration with the MoPH IT department, which played a central role in the reform process.

An important component of the project included the development of an automated billing system (ABS) aimed at simplifying the hospital billing procedures and expediting the reimbursement process. The development of the ABS included the purchase of data servers (jointly by the MoPH and private hospitals), software installation, user interface program development, and trainings for hospitals and MoPH staff on bill entry and submission to the MoPH. The impact of the ABS implementation was beneficial to both hospitals and the MoPH because it reduced the number of days for processing a hospital bill from 35 days in 2008 to only 5 days in 2014.

STRENGTHENING THE UTILIZATION REVIEW FUNCTION

The Utilization Review Committee was primarily tasked with developing the utilization review function to help improve MoPH spending efficiency as well reduce misuse or abuse of hospitalization. The committee started by analyzing the hospital database, which included a wealth of information on medical diagnoses, surgical procedures, cost, and length of stay for every case treated at public or private hospitals. Using this database, the MoPH became more aware of the wide variation among cases and hospitals costs, specifically for medical admissions. This was to some extent expected, given the fee-for-service structure of such cases. The work of the committee focused on the following:

1. Extraction, exploration, and analysis of hospitalization data, including comparison of admission to discharge diagnoses and cost indicators.
2. Development of a hospital coding nomenclature to support analysis.
3. Review of several medical diagnoses and surgical procedures, primarily those that were among the most frequent, most expensive, and/or most prone for abuse/misuse (e.g., appendectomy, headache, tonsillitis).
4. Identification of cost/frequency–outlier hospitals/cases.
5. Development of a rules engine to define which surgical procedures are allowed to be coupled together.
6. Development of an algorithm for identification of outlier bills for referral to MoPH auditing.
7. Interviews and collection of feedback and recommendations from MoPH medical controllers regarding improvement of hospitalization system.

Prior to this activity, the MoPH auditing body made limited use of the database and had a practice of randomly selecting around 10% of all cases for detailed auditing. The work under this function provided auditors with a nonrandom method for selecting hospital bills using the algorithm developed to identify outlier bills. The identified bills were those that markedly differed in cost and length of stay, totaling up to about 3% of total annual cases. This complemented the existing practice of random bill selection by the MoPH auditing body but helped focus the efforts on a small proportion of bills (3%), which accounted for over 20% of total hospitalization costs. This approach also led to the identification of select hospitals that had a relatively greater proportion of such bills. Subsequent notifications and deductions were made by the MoPH for some of these hospitals. In 2014, this was formally adopted as a regular process conducted in coordination between the IT department and the auditing body at the MoPH.

The analysis of hospitalization data also included a study whose key findings suggested that “[...] a significant proportion of the hospital admissions under [the MoPH] are for conditions that could generally have been described as relatively minor and hence that may not have needed hospitalization, at least not to that level.”¹¹ The results also highlighted the differences in both cost and average length of stay of medical cases of private versus public hospitals by region and by reimbursement category. The study was crucial to inform the work of the Admission Criteria Committee in the selection of conditions for criteria development and of the Performance Contracting Committee in the selection of indicators.

DEVELOPING ADMISSION CRITERIA

During the 1990s, the health sector in Lebanon witnessed a large increase in the number of physicians in the country, reaching around 8.3% increase per year.¹² This was in large part due to physicians returning to Lebanon after the end of the 1975–1990 civil unrest and those returning from fellowships and grants to study medicine outside Lebanon. This led to a high physician-to-population ratio of 3.2 per 1,000 inhabitants (2007–2013), well above the regional and global

averages of 1.3 and 1.4 per 1,000, respectively.¹³ This also meant a multiplicity of graduating countries and wide variation in educational backgrounds, making it particularly difficult to standardize clinical practices and protocols among providers.¹²

To address this situation, the Admission Criteria Committee was tasked with addressing the promotion of good provider practices by developing standardized admission criteria. The main approach of this committee was to explore interventions to support physicians in evidence-based decision making for hospital admission. It also provided support for MoPH medical controllers who are often required to decide on admission of cases outside their specialty.

Building on the information gained through the utilization review function, the committee evaluated medical and surgical conditions and developed admission criteria for 40 high-cost, high-volume, and/or misuse- and abuse-prone conditions (Table 1). The criteria developed benefited from the different medical backgrounds and specialties of committee members and the in-depth knowledge of the challenges and barriers faced by medical controllers. An extensive review of existing national and sub-national medical guidelines was also conducted through sources such as PubMed, US National Guideline

Clearinghouse, UK National Institute for Clinical Excellence, and the French Haute Autorité de Santé. Guidelines were reviewed and adapted to the Lebanese context in order to increase the likelihood of physician compliance.

Several tools were also developed to support the medical controllers in their case review process. Brief checklists to support decision making for 34 conditions were prepared. Similarly, an indication level notification was also added to hospital admission forms of all cases under MoPH coverage, whereby medical controllers could inform the MoPH of their opinion on whether an admission had a low, medium, or high indication for admission. Low-indication cases were particularly monitored by the MoPH. Though such cases may not be inappropriate to admit, a high proportion prompted closer review by the MoPH.

DEVELOPING MIXED-MODEL HOSPITAL CONTRACTING

The Performance Contracting Committee had a considerably lengthy period of discussion and planning, both to build on information gained through utilization review and to review the various approaches of evaluating hospital case complexity and indicators of performance.

Analysis of hospital case complexity was conducted using a locally adapted classification of *International Classification of Diseases*, 10th Revision (ICD-10)¹⁴ code grouping and surgical procedures, with greater weight given to diagnoses expected to be more complex/serious or more costly. However, this approach was subsequently replaced with the development of hospital case mix index (CMI) calculation using ICD-10 code and surgical procedure code, with the standard formula below:

$$CMI_h = \frac{\sum_g [W_g * N_{gh}] / \sum_g N_{gh}}{\sum_g [W_g * N_{gn}] / \sum_g N_{gn}}$$

where *h* is the hospital; *W_g* is the weight calculated for each ICD/common procedural terminology code; *N_{gh}* is the number of cases within each code in hospital *h*; and *N_{gn}* is the number of cases within each code in the total population.

Despite the absence of diagnosis related groups (DRGs) in Lebanon, which is commonly used for case mix index calculation, and following review of the literature and local adaptation, it was possible to calculate hospital case mix indices based directly on ICD-10 code and surgical procedure code. To our knowledge this was the first instance where such an approach was used at the national level where DRGs were not available. Despite the more limited precision expected in case mix index calculation in using this approach, it allowed

Aortic valve replacement due to atrial stenosis/regurgitation	Gastric bypass	Plasmapheresis
Asthma	Gastroenteritis and dehydration in children < 5 years	Prostatic stent
Bronchiolitis	Headache	Renal colic
Carpel tunnel syndrome	Hemorrhoidectomy	Septoplasty
Cataract	Hepatitis A	Sinusotomy
Chest pain	Hypertension	Spondylolisthesis
Cholecystectomy	Hysterectomy	Tonsillectomy
Chronic obstructive pulmonary disease	Idiopathic thrombocytopenic purpura	Tonsillitis
Cochlear implant	Laminectomy and discectomy	Total hip replacement
Colporrhaphy	Mastectomy	Total knee replacement
Community-acquired pneumonia	Orchiopexy	Tympanostomy and myringotomy
Coronary artery bypass graft	Otitis media	Varicocele
Febrile neutropenia	Ovarian cyst	Vitrectomy
Fever in children < 5 years		

TABLE 1. List of Conditions for Admission Criteria Development

the committee to overcome the limitations posed by the lack of DRGs and measure the complexity of cases admitted at hospitals throughout the country.

Importantly, a study was also undertaken examining hospitals across reimbursement rate tiers and found that CMI varied considerably both among and within reimbursement tiers.¹⁵ This argued against the previously prevailing assumption that hospitals with higher accreditation (and reimbursement) necessarily admitted more complex cases. The conclusion drawn from this was that the linkage of reimbursement rate solely to accreditation was not appropriate and led to unfair and inefficient reimbursement system.¹⁵ This provided further support for including additional factors besides accreditation to the contracting system.

Based on the results, the MoPH decided to move forward with the new model, linking private hospitals' performance to reimbursement. This represented a great challenge considering the climate of political instability and the strong affiliation of most hospitals to different political and confessional/religious factions. MoPH policy makers supported the decision with scientific evidence, particularly in using the results of the assessment study to highlight the gaps in the system.¹⁵ This enhanced the credibility and objectivity of the work and highlighted the technical approach of the new contracting system. MoPH conducted a broad meeting with all hospitals to disseminate the evidence and project approach and gather feedback from hospitals. This was a decisive step to bring the discussions to a scientific ground and cut short political and confessional/religious interference.

By November 2014, the MoPH changed its hospital contracting design to a mixed model that included hospital CMI, patient satisfaction, and three select policy indicators intended to incentivize increased intensive care unit capacity and decrease unnecessary admissions and inappropriate billing. This resulted in changes in reimbursement rate tiers in about half of contracted hospitals. More specifically, 65 hospitals

Tier	Before		After	
	Count	Percentage	Count	Percentage
High	44	34%	38	29%
Medium	58	45%	51	40%
Low	28	21%	40	31%
Total	130		129	

TABLE 2. Group-Level Changes in Hospital Reimbursement Rate Tiers

had a change in their reimbursement tier and 66 stayed within the same tier. The distribution of hospitals before and after the implementation of the mixed model is detailed in Table 2.

An evaluation of the hospital case mix model a year following implementation of the new system revealed that the average case mix index among private hospitals significantly increased from 1.14 to 1.18. This suggested that the new system incentivized hospitals to admit fewer inappropriate cases and more cases that are more complex/serious. The small number of public hospitals limited the statistical power needed to detect changes in such in the short term.

Preliminary analysis also suggested that there are some improvements, especially among medical cases, that are more informative due to their fee-for-service structure. In the seven months following implementation (December 2014–June 2015), the average number of cases per month and the average cost per case decreased by 22.7% and 5.1%, respectively, compared to the period prior to implementation (Table 3). However, the average cost per inpatient day increased slightly (0.6%). This figure is of limited accuracy due to the short analysis period, which would not capture the entire length of stay of some long-term cases. Further analysis covering longer periods is needed in the future to provide more reliable results.

These findings complement the results of the case mix index, which suggests that fewer unnecessary and more complex cases are admitted following the implementation of the new system. This is not surprising given that a strong incentive

	June 2011–May 2012	June 2012–May 2013	June 2013–May 2014	December 2014–June 2015
Total cost (1,000 LBP)	94,160,183	112,632,757	123,684,592	52,977,762
Number of cases, total	93,340	96,842	105,000	47,372
Number of cases, average per month	7,778	8,070	8,750	6,767
% Change	—	3.8	8.4	-22.7
Cost per case, average (1,000 LBP)	1,009	1,163	1,178	1,118
% Change	—	15.3	1.3	-5.1
Inpatient days, total	379,412	397,748	430,099	183,187
Cost per inpatient day, average (1,000 LBP)	248	283	287	289
% Change	—	14.1	1.6	0.6
Average length of stay	4.06	4.11	4.10	3.87

TABLE 3. Changes in Cost and Length of Stay of Medical Hospitalization Cases (Excluding Surgical Cases) Since June 2011. *Note.* LBP = Lebanese Pounds

was created by the new system for hospitals to avoid unnecessary admissions. It is also likely that changes in hospital practices may have been influenced by the various activities introduced by this reform. Numerous hospitals sought to improve the quality of coding of hospital cases, which in the new system was linked to their reimbursement rate. Several trainings on ICD-10 coding for medical diagnoses and common procedural terminology for surgical procedures have been conducted for hospital coders. Nevertheless, these preliminary findings will need to be further examined as more hospitalization data from the postimplementation period accumulate, allowing for trend analysis for longer-term impact assessment.

MOVING FORWARD

The new mixed-model hospital contracting model implemented in 2014 was an important development in the relationship between the MoPH and hospitals. The new system is also more easily adaptable to subsequent improvements in the indicators used, which is an ongoing process. Though the formal activities of the three committees have been completed, the work on performance contracting, utilization review, and admission criteria has been incorporated into the MoPH's programs. This experience to date has laid the ground for improved hospital contracting, better collaboration among stakeholders, and informed new MoPH policy. In July 2016, the minister of public health announced a new policy that provides full hospital coverage to all uninsured citizens above the age of 64 years and publicly stated that "[this] will be funded through savings [expected to be] achieved from utilization review and auditing reforms"¹⁶ implemented in 2014.

Meanwhile, the MoPH is continuing the work to strengthen further the contracting system. This includes the development of more precise weights to increase the accuracy of the hospital case mix index, increase the use of hospitalization data for utilization review in medical auditing, and the development of performance indicators that reflect actual patient outcomes. As part of an ongoing process, this work will benefit from additional and more in-depth evaluation in the near future. As more postimplementation data are gathered, an evaluation study is planned to provide greater insight into the actual impact of the reform.

Several lessons may be drawn from the hospital contracting reform initiative of 2009–2014. The involvement from the conception of the initiative of multidisciplinary professionals with affiliations to key stakeholders and academia was instrumental in developing the activities of the three committees. The active involvement of MoPH policy makers was similarly crucial in providing the guidance, resources, and institutional commitment toward achieving the goals of improving

contracting appropriateness/fairness, MoPH spending efficiency, and provider practices. The use of financing as a leverage to introduce the new initiative was also a major advantage to incentivize stakeholders. In addition, the multipronged approach to interrelated goals was highly beneficial and allowed the committees to build on each other's efforts and undertake a wide range of interventions.

The trainings provided to MoPH's medical controllers (stationed at all hospitals for approval of admissions) were important in building their capacity for admission criteria implementation, as well as using evidence-based material to empower their role as gatekeepers for hospital admission. The extensive use of hospitalization data throughout this process has been critical in increasing the knowledge of the MoPH regarding the coverage it provides to uninsured persons in Lebanon, thus providing the basis for the interventions undertaken. Many hospitals have responded favorably to the recent changes, although a more formal assessment has yet to be undertaken. The close partnership between the Syndicate of Private Hospitals and MoPH was key to supporting the new hospital contracting system.

The Lebanese health care system has exhibited considerable resilience and sustained its achievements in a context of political instability.⁷ Collaborating with the private sector based on transparent and participatory governance enhanced commitment of all parties to shared goals and provided a framework to manage diversity and optimize the use for national resources. The MoPH's role as facilitator and steward of the partnership was enhanced by the introduction of a fairer and more appropriate contracting model based on local evidence that also helped counter prevailing political and confessional/religious favoritism. The new hospital contracting system is one example of a merit system that the MoPH succeeded in introducing and gaining the acceptability by a major player in the health sector: the private hospitals.

DISCLOSURE OF POTENTIAL CONFLICTS OF INTEREST

The authors declare they have no conflicts of interest. Authors JK, JM and WA have affiliations to the Ministry of Public Health, and author NR to the World Bank.

ACKNOWLEDGMENTS

The authors acknowledge the WHO country office (Lebanon) for their support and all individuals who have been involved

in the Performance Contracting, Utilization Review, and Admission Criteria committees.

FUNDING

The second Emergency Social Protection Implementation Support Project (ESPISP-2) was financed by a grant from the World Bank to the Lebanese Ministry of Public Health during 2009–2014.

ORCID

Bjorn Ekman  <http://orcid.org/0000-0001-7630-7904>

REFERENCES

- [1] United Nations Department of Economic and Social Affairs. 2015 Revision of world population prospects. July 2015. Available at <https://esa.un.org/unpd/wpp/Download/Standard/Population/> (accessed 22 August 2016).
- [2] Ammar W. Health beyond politics. Beirut: World Health Organization; 2009.
- [3] Ammar W, Wakim R, Hajj I. Accreditation of hospitals in Lebanon: a challenging experience. *East Mediterr Health J* 2007; 13(1): 138-149.
- [4] World Health Organization. Global Health Observatory data. January 2015. Available at http://who.int/gho/mortality_burden_disease/en/ (accessed 22 August 2016).
- [5] World Health Organization. The world health report: health systems financing: the path to universal coverage. Geneva: World Health Organization; 2010.
- [6] Economist Intelligence Unit. Health outcomes and cost: a 166-country comparison. London: Economist Intelligence Unit; 2014.
- [7] Ammar W, Kdouh O, Hammoud R, Hamadeh R, Harb H, Ammar Z, Atun R, Christiani D, Zalloua PA. Health system resilience: Lebanon and the Syrian refugee crisis. *J Glob Health* 2016; 6(2): 299-307.
- [8] Ministry of Public Health, Lebanon. National health accounts 2012. Available at <http://www.moph.gov.lb/en/Pages/8/141/national-health-accounts> (accessed 22 August 2016)
- [9] Ammar W. Health system and reform in Lebanon. Beirut: World Health Organization; 2003.
- [10] Ministry of Public Health, Lebanon. Statistical bulletin 2012. October 2014. Available at <http://www.moph.gov.lb/en/Pages/8/327/statistical-bulletins> (accessed 22 August 2016).
- [11] Kronfol N, Khalife J, Romanos J, Makouk J, Noun P, Ammar W. Observations on the Ministry of Public Health program of support to the hospitalization of patients in Lebanon. *J Med Liban* 2014; 62(1): 33-39.
- [12] Daher M, Husseini H, Kasparian R, Kasparian C. La démographie médicale au Liban. Pléthore, féminisation et rajeunissement [Medical demography in Lebanon. Plethora, feminization, youthfulness.] *J Med Liban* 1998; 46(1): 43-46.
- [13] World Health Organization. World health statistics 2015. Geneva: World Health Organization; 2015.
- [14] World Health Organization. International statistical classification of diseases and related health problems, 10th revision. Geneva: World Health Organization; 2003.
- [15] Ammar W, Khalife J, El-Jardali F, Romanos J, Harb H, Hamadeh G, Dimassi H. Hospital accreditation, reimbursement and case mix: links and insights for contractual systems. *BMC Health Serv Res* 2013; 13: 505.
- [16] Baalbaki S, An-Nahar News. Details of the complete health coverage for those older than 64 years. 16 July 2016. Available at: <http://newspaper.annahar.com/article/426628-%D9%88%D8%B2%D8%A7%D8%B1%D8%A9-%D8%A7%D9%84%D8%B5%D8%AD%D8%A9-%D8%AA%D8%B7%D9%84%D9%82-%D8%A7%D9%84%D8%AB%D9%84%D8%AB%D8%A7%D8%A1-%D9%85%D8%B4%D8%B1%D9%88%D8%B9-%D8%AA%D8%A3%D9%85%D9%8A%D9%86-%D8%A7%D9%84%D8%AA%D8%A7%D9%84%D8%AA%D8%BA%D8%B7%D9%8A%D8%A9-%D8%A7%D9%84%D8%B4%D8%A7%D9%85%D9%84%D8%A9-%D9%84%D9%85%D9%86-%D8%AA%D8%AC%D8%A7%D9%88%D8%B2-%D8%A7%D9%84-64-%D8%B9%D8%A7%D9%85%D8%A7-%D8%B9%D8%AF%D8%AF> (accessed 22 August 2016).

Paper 2



RESEARCH ARTICLE

REVISED

Hospital performance and payment: impact of integrating pay-for-performance on healthcare effectiveness in Lebanon [version 2; peer review: 2 approved]

Jade Khalife ^{1,2}, Walid Ammar^{2,3}, Maria Emmelin¹, Fadi El-Jardali ³, Bjorn Ekman ¹

¹Faculty of Medicine at Lund University, Lund, Sweden

²Ministry of Public Health, Beirut, Lebanon

³Faculty of Health Sciences, American University of Beirut, Beirut, Lebanon

V2 First published: 14 May 2020, 5:95
<https://doi.org/10.12688/wellcomeopenres.15810.1>
 Latest published: 10 Dec 2020, 5:95
<https://doi.org/10.12688/wellcomeopenres.15810.2>

Abstract

Background: In 2014 the Lebanese Ministry of Public Health integrated pay-for-performance into setting hospital reimbursement tiers, to provide hospitalization service coverage for the majority of the Lebanese population. This policy was intended to improve effectiveness by decreasing unnecessary hospitalizations, and improve fairness by including risk-adjustment in setting hospital performance scores.

Methods: We applied a systematic approach to assess the impact of the new policy on hospital performance. The main impact measure was a national casemix index, calculated across 2011-2016 using medical discharge and surgical procedure codes. A single-group interrupted time series analysis model with Newey ordinary least squares regression was estimated, including adjustment for seasonality, and stratified by case type. Code-level analysis was used to attribute and explain changes in casemix index due to specific diagnoses and procedures.

Results: Our final model included 1,353,025 cases across 146 hospitals with a post-intervention lag-time of two months and seasonality adjustment. Among medical cases the intervention resulted in a positive casemix index trend of 0.11% per month (coefficient 0.002, CI 0.001-0.003), and a level increase of 2.25% (coefficient 0.022, CI 0.005-0.039). Trend changes were attributed to decreased cases of diarrhea and gastroenteritis, abdominal and pelvic pain, essential hypertension and fever of unknown origin. A shift from medium to short-stay cases for specific diagnoses was also detected. Level changes were attributed to improved coding practices, particularly for breast cancer, leukemia and chemotherapy. No impact on surgical casemix index was found.

Conclusions: The 2014 policy resulted in increased healthcare

Open Peer Review

Approval Status  

	1	2
version 2 (revision) 10 Dec 2020		 view
version 1 14 May 2020	 view	 view

1. **Kristiina Kahur** , Nordic Casemix Centre, Helsinki, Finland
2. **Linghan Shan**, Harbin Medical University, Harbin, China

Any reports and responses or comments on the article can be found at the end of the article.

effectiveness, by increasing the casemix index of hospitals contracted by the Ministry. This increase was mainly attributed to decreased unnecessary hospitalizations and was accompanied by improved medical discharge coding practices. Integration of pay-for-performance within a healthcare system may contribute to improving effectiveness. Effective hospital regulation can be achieved through systematic collection and analysis of routine data.

Keywords

performance, health systems, reform, casemix index, low and middle income countries, interrupted time series analysis, unnecessary hospitalization, coding

Corresponding author: Jade Khalife (jade.khalife@med.lu.se)

Author roles: **Khalife J:** Conceptualization, Data Curation, Formal Analysis, Funding Acquisition, Investigation, Methodology, Project Administration, Resources, Software, Visualization, Writing – Original Draft Preparation; **Ammar W:** Conceptualization, Funding Acquisition, Resources, Writing – Review & Editing; **Emmelin M:** Funding Acquisition, Supervision, Writing – Review & Editing; **El-Jardali F:** Funding Acquisition, Project Administration, Writing – Review & Editing; **Ekman B:** Conceptualization, Funding Acquisition, Project Administration, Supervision, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: This work was supported by the Wellcome Trust [204020] and a project grant [MR/N015916/1] from the Joint Health Systems Research Initiative (DFID/MRC/Wellcome Trust/ESRC).

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2020 Khalife J *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Khalife J, Ammar W, Emmelin M *et al.* **Hospital performance and payment: impact of integrating pay-for-performance on healthcare effectiveness in Lebanon [version 2; peer review: 2 approved]** Wellcome Open Research 2020, 5:95 <https://doi.org/10.12688/wellcomeopenres.15810.2>

First published: 14 May 2020, 5:95 <https://doi.org/10.12688/wellcomeopenres.15810.1>

REVISED Amendments from Version 1

Based on the feedback received from the two reviewers, we have updated the manuscript with some additional information. We have clarified that the policy intervention was directed at the categorization of hospitals into the three reimbursement tiers, but not the reimbursement structure itself. We have clarified the weight-setting process for diagnosis and procedure codes. We have defined the systematic approach used in this investigation. We have acknowledged that inclusion of age and comorbidity variables would allow more accurate calculation of hospitalization complexity.

Any further responses from the reviewers can be found at the end of the article

Introduction

The linkage of performance and payment has been increasingly used in healthcare during the past two decades. In recent years this has extended towards hospital performance designs, despite mixed evidence regarding its effects.

In 2014 the Lebanese Ministry of Public Health (MoPH) changed the basis by which it determined the payment reimbursement tiers of about 140 public and private hospitals, which it contracts for providing hospitalization service coverage for the majority of the Lebanese population. A pay-for-performance (P4P) framework was integrated within this system, the most prominent component of which was the hospital casemix index (CMI), which reflects the average risk or illness severity of all patients within a hospital (Khalife *et al.*, 2017). This intervention was aimed at improving effectiveness by incentivizing hospitals to decrease unnecessary hospitalizations, as well as to improve fairness in determining hospital reimbursement (by including risk-adjustment), within an integrated evaluation framework.

Hospitalizations that are not compliant with any medical or social criteria may be considered as unnecessary hospitalizations. These arguably differ from potentially preventable hospitalizations; in that the latter may have an indication for admission but would have otherwise been avoided with appropriate outpatient or primary care. Unnecessary hospitalizations are common in various countries and challenge the functioning of healthcare systems (Caminiti *et al.*, 2013; Macinko *et al.*, 2010; Stranges & Friedman, 2006; To *et al.*, 1996). Such cases may be primarily regarded as a question of appropriateness of care. Using the Kruk and Freedman performance framework, unnecessary hospitalizations may be categorized as a quality of care output measure, under the effectiveness dimension (Kruk & Freedman, 2008). Downstream association with outcomes on patient health status (effectiveness) and on maximizing value of resources (efficiency) would also be expected.

The current evidence on the impact of P4P in healthcare is weak, particularly in low- and middle-income countries (LMICs) (Witter *et al.*, 2012). In this study we assess the impact

of the integration of the P4P policy by the MoPH, and contribute to the evidence base on P4P effectiveness, using routine data and a specially developed CMI.

Pay for performance

A health reform that links measures to payment creates a financial incentive for service providers to improve their performance vis-à-vis these measures. Seen through the lens of principal-agent theory, such P4P seeks to address the recognized information problems within healthcare, particularly information asymmetry (Smith & Hanson, 2012). This alignment of interests provides the principal (i.e. the payer) a tool to incentivize the provider to improve healthcare outputs and outcomes (Grace *et al.*, 2015). Such tools may function at a system level within a complex environment and should be adjustable by the principal (Roberts *et al.*, 2004). Aspects of a P4P intervention that have a large role in determining impact include the actual measures used, context and incentive size.

Most performance frameworks have either been developed for use in high-income countries (HICs) or are heavily influenced by such contexts, and likely require adjustment for use in LMICs (Tashobya *et al.*, 2014). Based on a review of commonly used performance indicators, Kruk and Freedman provide a framework for LMICs (Kruk & Freedman, 2008). Adapting Donabedian's system evaluation of structure, process and outcome measures, this framework categorized indicators as outputs/processes and outcomes/impact along with the dimensions of effectiveness, equity and efficiency (Donabedian, 1966; Donabedian, 1988).

In LMICs, healthcare P4P initiatives commonly include structural measures of quality. More widely, however, a transition is ongoing towards outcome measures, which are the ultimate target for performance improvement (Chee *et al.*, 2016; Gergen *et al.*, 2017). Establishing broad outcome measurements may enhance P4P impact and integration into routine systems may make such initiatives more cost-effective (Borghi *et al.*, 2015; Chee *et al.*, 2016).

The effects of healthcare P4P in HIC contexts have been mixed, largely confirming the scarcity of the evidence base (Eijkenaar *et al.*, 2013; Emmert *et al.*, 2012; Gillam *et al.*, 2012; Van Herck *et al.*, 2010). A recent review on P4P schemes in the United Kingdom found some positive findings, but cautioned that overall effects on care quality were unclear, underscoring the need for long-term monitoring and evaluation (Mandavia *et al.*, 2017).

The potential for P4P impact may be larger in LMICs, considering the relatively lower resources of providers and more dynamic health reform context (Witter *et al.*, 2012). However, the scarcity of the evidence from LMICs is particularly pronounced; the evidence base is too weak to make general conclusions (Witter *et al.*, 2012). A recent review of P4P in maternal and child care in LMICs found positive impact on process quality, but weak evidence on health outcomes and out-of-pocket expenses (Das *et al.*, 2016).

Hospital-based P4P initiatives in England and the United States targeting 30-day readmissions and mortality resulted in improved short-term performance that was not sustained in the long-term, further highlighting challenges due to contextual changes, spill-over, measurement limitations and overall design (Jha *et al.*, 2012; Kristensen *et al.*, 2014; Lindenauer *et al.*, 2007; Sutton *et al.*, 2012; Werner *et al.*, 2011). Long-term investigations of previously favorable initiatives have found improvement in readmissions, but no improvement or worsening of mortality, as well as undesirable practices having misled some earlier findings (Gupta & Fonarow, 2018; Wadhwa *et al.*, 2018; Wasfy *et al.*, 2017). Anticipatory or short-term behaviors in response to P4P scheme engagement and implementation may differ from long-term behavior, likely contributing to the findings of short but not long-term effectiveness of P4P (Ryan *et al.*, 2015).

It has been proposed that the debate on P4P should move from distinct projects towards integration within the health system, with broad system objectives (Soucat *et al.*, 2017). Such integration may be seen as an extension of strategic purchasing of health services, working towards achieving universal health coverage. An approach that considers the overall health system rather than more narrow objectives would avoid ‘not seeing the forest for the trees’ (Soucat *et al.*, 2017).

Aim and objectives

We apply a systematic approach to assess whether the new pay-for-performance policy had an impact on the healthcare system’s effectiveness. We analyze whether this policy affected the complexity of the average hospitalization case (i.e. CMI), stratifying by hospitalization case types, length of stay, and hospital ownership. We further quantify any changes, with plausible explanations, at the level of diagnoses and procedures.

The specific objectives are:

1. To determine the impact on CMI level and trend, across public and private hospitals
 - a. by case types: medical, surgical and mixed
 - b. by length of stay: medical short, medium and long-stays
2. To detect any changes in diagnoses and procedures
 - a. by contribution to CMI changes
 - b. by hospitalized cases

The Lebanese health care context and reforms

Lebanon is a small Mediterranean country of 4.5 million citizens and 2 million refugees (primarily Syrian) (United Nations, 2017). The diversified healthcare system is dominated by public payers and private providers (Ammar, 2003). The MoPH is the largest public payer, covering hospitalization for about 52% of Lebanese, who otherwise lack any insurance coverage for hospitalization (Ammar, 2009). This role is a legacy of the 1975–1990 war that disrupted developmental reforms at various stages.

The MoPH has engaged in different healthcare supply-side reforms. Contracting private hospitals to provide hospitalization services took place in the 1960s due to public coverage limitations and political reasons, preceding neoliberal influences that supported New Public Management agendas in other systems (Smith & Hanson, 2012). In the late 1990s public hospitals were granted semi-autonomous status, similar to contemporary experiences of public hospitals in HICs and LMICs.

Hospital accreditation was linked to payment in 2001 and incentivized hospitals towards quality improvement, with accreditation results used to categorize hospitals across three reimbursement tiers (Ammar *et al.*, 2007). Hospital budgets set on an annual basis were not targeted by this change, as they continued to be historically and politically determined. Therefore, a moral hazard existed for hospitals interested in admitting more predictable low-risk patients to efficiently reach their pre-allocated budgets. Unnecessary hospitalizations became increasingly recognized as a major problem, particularly for medical (non-surgical) cases where lower barriers to treatment existed (Kronfol *et al.*, 2014). These were often one or two-day stays with variable costs but generally less complex diagnoses or symptoms, such as nausea, vomiting, diarrhea and gastroenteritis (Kronfol *et al.*, 2014).

Policy reform

To address unnecessary hospitalizations and increase fairness in hospital performance assessment, the MoPH created a new policy intervention in late 2014, linking reimbursement tier to a composite hospital total performance score (TPS) (Ammar *et al.*, 2013). The first public announcement of the intervention was through an engagement event held in August 2014 for hospital executives and managers. The event highlighted the new model and its components, as well as the importance of accurate coding for appropriate casemix assessment. The TPS included components on CMI, patient satisfaction and other minor policy indicators, as well as accreditation (Khalife *et al.*, 2017). Specifically, this policy had a regulatory aspect in the form of accreditation, and more distinct payment aspect for other components. The CMI reflects the average risk or illness severity of all patients within a hospital, and was effectively the greatest determinant of a hospital’s TPS (Hornbrook, 1982). Therefore, a decrease in unnecessary hospitalizations would be expected to increase a hospital’s CMI and consequently its TPS. Additional information regarding this process has been described elsewhere (Khalife *et al.*, 2017).

The incentive size for tier classification remained unchanged from pre- to post-intervention periods. Broadly, the difference in reimbursement tier is 10% additional change per tier among surgical procedures, and about 15% per tier among medical cases; e.g. a bottom-tier ‘T3’ hospital charges 1,000,000 LBP (USD 670) for a surgical procedure, while a top-tier ‘T1’ hospital charges 1,200,000 (USD 800) for the same procedure.

Consequently, the policy intervention was directed at the mechanism whereby hospitals were categorized into the different

reimbursement tiers. However, the intervention did not change the reimbursement structure of hospitalization cases. Reimbursement of medical cases remained fee-for-service, while surgical cases used a pre-defined flat-fee.

Methods

Ethical statement

Research protocol approval was granted by the Institutional Review Board (IRB) at the American University of Beirut (ID: FHS.FE.21). The requirement for patient consent was waived by the IRB.

Data sources

Hospitalization data including all cases under MoPH coverage from January 2011 to December 2016 was extracted by the MoPH Information Technologies Department and shared with the research team in a format with patient identifiers anonymized. The fields included record number (unique per admission), case identifier, hospital code, admission date, discharge date, length of stay, total charge, medical code on each admission and discharge, and surgical procedure code. STATA software package version 11 was used for all calculations and analyses.

CMI calculation

The MoPH overcame limitations in developing a hospital CMI for its contracted 146 public and private hospitals, due to the lack of local Diagnosis Related Groups (DRGs) on which most casemix systems rely on. The MoPH CMI calculation approach used average costs for weight-setting among medical cases based on International Classification of Diseases, 10th Revision (ICD-10) discharge code, and Common Procedural Terminology (CPT) procedure code among surgical cases (Ammar *et al.*, 2013; World Health Organization, 2006; Yang & Reinke, 2006). Among medical cases the weight-setting process was separated across short-stay (<2 days), medium-stay (2–15 days) and long-stay cases (>15 days). This approach has been detailed elsewhere (Ammar *et al.*, 2013; Khalife *et al.*, 2017).

For greater comparability across case types, we standardized the weight-setting previously used by the MoPH; we used five-year cost averages rather than the MoPH two-year averages; and assigned the average weight among medical cases for low-volume medical conditions (less than 20 cases). Surgical CMI did not require average cost figures as surgical services have fixed flat-rate charges. However, 11 procedures with weights ten times above the standard reference (1 million LBP) were capped at a weight of 10.00 to limit excessive impact of outliers.

The weight for each ICD and procedure code was the same throughout the period investigated. We note that the MoPH updated procedure costs in March 2013, which remained in use until the 2018 update. This increased the base-rate of all procedures, to account for inflation. We used weights based on the March 2013 update, also retrospectively up to 2011. The MoPH undertook a hospital-bed update (base-rate), which is only one component of the bill charged by hospitals to the MoPH, but nevertheless represents an internal inflation adjustment.

For medical cases, we used the five-year code average without further adjustment.

We developed algorithms to calculate monthly CMI (rather than yearly) and developed ‘mixed’ cases algorithms (cases concurrently including medical and surgical care). We incorporated secondary procedures into surgical CMI, but this was not done for medical CMI as additional diagnoses or comorbidities are not currently utilized.

CMI was calculated using the formula below, which excludes the denominator correction component used in some versions, as this was calculated at system rather than hospital level (Lichtig, 1986). This is also the standard generic CMI formula used by the US Centers for Medicare and Medicaid Services (Services, 2011).

$$\text{CMI} = \frac{\sum[W_g * N_{gn}]}{\sum g N_{gn}}$$

where W_g is the weight calculated for each ICD or CPT, and N_{gn} is the number of cases within each ICD in the total population.

CMI was calculated for medical, surgical and mixed cases separately, and repeated using cases only at public hospitals and only at private hospitals separately. Among the medical cases, short-stay, medium-stay and long-stay casemix indices were also calculated separately at each stage and combined using a case-weighted approach to also obtain an all-stay medical CMI. Such a combination was not conducted for all cases combined as it would likely conceal meaningful findings.

We excluded all chemotherapy coded cases (ICD Z51.1 and Z51.2). Chemotherapy is generally a low-cost short-stay hospitalization, and the MoPH has since 2014 communicated to hospitals the need for correct coding specifically for cancer patients. As a result, the MoPH has documented a decrease in chemotherapy miscoding under general cancer codes (e.g. C50 code being corrected as Z51.1). Retaining chemotherapy codes would have had the effect of artificially decreasing CMI. Typically, chemotherapy cases form around 7.9% of medical cases or 4.4% of total cases.

CMI algorithms were first run on an annual level for developing descriptive statistics. Similar algorithms were used to develop a monthly CMI with discharge date used to categorize records into calendar months.

Study design and methods

Impact on CMI level and trend, across public and private hospitals (Objective 1). To detect a change in CMI attributable to our intervention, it is necessary to make use of a control, whether outside of the group (using randomization) or within it (using historical control), while accounting for potential confounders. Randomization was not possible in this situation due to legal regulations that necessitate the MoPH to use the

same performance assessment and reimbursement framework for all acute-care hospitals. Considering the availability of multi-year data, we chose to use a single-group interrupted time series (ITS), with Newey ordinary least squares regression. ITS analysis is considered the most appropriate quasi-experimental design, given the research context and aims (Penfold & Zhang, 2013). With such an approach we seek to identify an ‘interruption’ of a continuous sequence of observations in a population (a time series) by a specific intervention. ITS analysis is particularly useful when randomization is not possible, and may have greater external validity than randomized designs when occurring in a real-world setting (Bernal *et al.*, 2016). This approach uses a historical control group, and more broadly the Bradford Hill criteria for causality to offer plausible causal explanations (Habicht *et al.*, 1999; World Health Organization, 2009; World Health Organization, 2012). Such an approach has also been noted as the most suitable for the evaluation of P4P interventions (World Health Organization, 2009).

ITS analysis has been increasingly used in population-level evaluations of health interventions, with advantages including the control for secular trends, ease of population-level analysis and ability to evaluate intended and unintended outcomes (Bernal *et al.*, 2016; Lopez Bernal *et al.*, 2018). It also allows us to evaluate both the short- and long-term impact of an intervention, by analysis of level and trend changes, respectively, and when appropriately used allows us to limit the effects of history bias in what is a complex real-world setting (Naci & Soumerai, 2016).

We applied ITS analysis on a dataset including all cases between January 2011 to December 2016, with the exception of the mixed cases dataset that began January 2012 (see Table 1). Statistical significance was set at $p < 0.05$. It is relevant to note that discharge codes were recorded in the hospitalization database as of January 2011 for most hospitals, with a few hospitals reaching full compliance within five months. Mixed cases data was available as of January 2012.

The intervention was given a two-month lag for expected effect (i.e. as of October 2014), based on the MoPH experience of the hospital response time (personal communication; Dr. Jihad Makouk, MoPH). We adjusted for seasonality using calendar months. Using 72 monthly data points, and CMI as the dependent variable, we ran ITS separately for each of medical all-stay, short-stay, medium-stay and long-stay cases; surgical cases and mixed cases. We subsequently stratified into public and private hospital cases.

As part of the sensitivity analysis we also varied lag period between intervention and impact by zero to four months, with the overall results being unchanged except in magnitude. To ensure the adequacy of our ITS model, we assessed auto-correlation using the Cumby-Huizinga test; in most cases auto-correlation was not present, and where detected was otherwise accounted for by the lag period.

Changes in diagnoses and procedures (Objective 2). We used a before and after approach to quantify the change in individual

diagnoses and procedures in terms of CMI change (contribution to changes detected in Objective 1), and in terms of absolute and relative change to all hospitalizations. We used algorithms on a dataset including all medical and surgical cases, with a pre-intervention period combining 2013 and 2014 cases and a post-intervention period with 2015 and 2016 cases. This objective used annual cut-offs, comparing 2013 and 2014 cases with 2015 and 2016 cases, in contrast to the first objective (monthly). This was to allow for the development of policy-relevant information and allow CMI algorithm looping and comparison across around 10,000 ICD and CPT codes.

For each case type and stay, we selected the top twenty codes with the greatest CMI change from pre- to post-intervention periods, for code-level analysis. We therefore restricted our analysis to examine the codes with the greatest impact on CMI in any direction (i.e. positive or negative).

In our approach, the codes which had the greatest *impact* on overall CMI are identified as a function of code weight in relation to the overall CMI ‘average’, the code weight itself, and frequency. While the codes with the greatest change in terms of their *share* of CMI are identified as a function of code weight and frequency only.

CMI change formulae:

$$WN_0 = (W_g N_{gn})^{2013} + (W_g N_{gn})^{2014}$$

$$WN_1 = (W_g N_{gn})^{2015} + (W_g N_{gn})^{2016}$$

$$\text{Code Count Effect (CCE)} = \left[(WN_1 - WN_0) * (W_g - CMI_{ref}) \right]^2$$

$$\text{Code attributable change} = \left(\frac{CCE_g}{\Sigma CCE} \right) * 100\%$$

Where W_g is the weight calculated for each ICD or CPT, and N_{gn} is the number of cases within each ICD in the total population.

CMI *share* change formula:

$$\text{Code share change} = \frac{(WN_1 - WN_0)}{WN_0}$$

Results

Descriptive statistics

The total study population across 2011–2016 included 1,353,025 inpatient hospitalizations, or about 230,000 per year, which were composed of 55% medical cases, 43% surgical cases and 2% mixed cases (see Table 1). Medium-stay medical cases were about four times more frequent than short-stay cases (43% to 11%), with long-stay cases being a small minority (1%). Among surgical cases the ratio of case to procedure was 1 case to 1.1 procedure, and among mixed cases 1 case to 1.5 procedure.

Table 1. Hospitalization cases under the Ministry of Public Health coverage at public and private hospitals in Lebanon, 2011–2016.

Case type	2011		2012		2013		2014		2015		2016		All years	
	n	%	n	%	n	%	n	%	n	%	n	%	n	%
All hospitals														
Medical	19,574	10.4	26,044	11.5	27,119	11.5	26,720	11.3	24,578	10.4	24,220	10.5	148,310	11.0
Short-stay	79,264	42.3	94,523	41.8	101,132	43.0	101,524	42.9	107,379	45.3	101,772	44.2	585,809	43.3
Medium-stay	2,156	1.1	2,851	1.3	2,793	1.2	2,770	1.2	2,674	1.1	2,494	1.1	15,744	1.2
Long-stay	100,994	53.8	123,418	54.6	131,044	55.8	131,014	55.4	134,631	56.8	128,486	55.7	749,863	55.4
All medical cases	93,516	-	111,355	-	112,523	-	114,021	-	110,168	-	109,083	-	650,666	-
Surgical	83,025	44.3	98,396	43.6	99,767	42.5	100,934	42.7	97,037	40.9	95,602	41.5	574,975	42.5
Cases	5,247	-	6,358	-	6,164	-	6,668	-	8,241	-	8,932	-	41,610	-
Procedures	3,584	1.9	4,120	1.8	4,114	1.8	4,446	1.9	5,517	2.3	6,396	2.8	28,187	2.1
Total cases	187,603	100	225,934	100	234,925	100	236,394	100	237,185	100	230,484	100	1,353,025	100
Public														
Case type														
Medical	6,301	12.7	9,263	13.3	9,178	13.0	9,541	12.8	10,051	12.8	10,634	12.4	55,033	12.8
Short-stay	20,968	42.2	29,229	41.8	30,672	43.3	33,680	45.3	37,492	47.8	40,832	47.7	193,094	45.0
Medium-stay	501	1.0	747	1.1	633	0.9	630	0.8	669	0.9	704	0.8	3,889	0.9
Long-stay	27,770	56.0	39,239	56.2	40,483	57.2	43,851	59.0	48,212	61.5	52,170	60.9	252,015	58.7
All medical cases	24,049	-	33,463	-	33,031	-	33,055	-	32,871	-	36,471	-	192,940	-
Procedures	20,906	42.1	29,167	41.8	28,845	40.7	28,896	38.9	28,485	36.3	31,491	36.8	167,990	39.1
Cases	1,476	-	2,430	-	2,275	-	2,367	-	2,423	-	2,726	-	13,697	-
Procedures	957	1.9	1,453	2.1	1,492	2.1	1,593	2.1	1,678	2.1	1,997	2.3	9,180	2.1
Cases	49,633	100.0	69,859	100.0	70,820	100.0	74,340	100.0	78,375	100.0	85,658	100.0	429,185	100.0
Total cases														
Private														
Case type														
Medical	13,273	9.6	16,781	10.8	17,941	10.9	17,179	10.6	14,527	9.1	13,586	9.4	93,287	10.1
Short-stay	58,296	42.3	65,294	41.8	70,460	42.9	67,844	41.9	69,887	44.0	60,940	42.2	392,721	42.5
Medium-stay	1,655	1.2	2,104	1.3	2,160	1.3	2,140	1.3	2,005	1.3	1,790	1.2	11,854	1.3
Long-stay	73,224	53.1	84,179	53.9	90,561	55.2	87,163	53.8	86,419	54.4	76,316	52.8	497,862	53.9
All medical cases	69,467	-	77,892	-	79,492	-	80,966	-	77,297	-	72,612	-	457,726	-
Procedures	62,119	45.0	69,229	44.4	70,922	43.2	72,038	44.5	68,552	43.2	64,111	44.4	406,971	44.1
Cases	3,771	-	3,928	-	3,889	-	4,301	-	5,818	-	6,206	-	27,913	-
Procedures	2,627	1.9	2,667	1.7	2,622	1.6	2,853	1.8	3,839	2.4	4,058	2.8	18,666	2.0
Cases	137,970	100.0	156,075	100.0	164,105	100.0	162,054	100.0	158,810	100.0	144,485	100.0	923,499	100.0
Total cases														

There was limited variation across 2011–2016 in terms of total cases, the exception being 2011, during which not all hospitals were transmitting discharge diagnoses to the MoPH hospitalization database until June 2011. We observed a minor decrease of surgical cases (absolute and relative) and medical short-stay cases in 2015 and 2016.

Approximately one in every three hospitalizations occurs in public hospitals. The proportion of hospitalization at public hospitals increased from 31% to 37% between 2012 and 2016. Most of this change is due to increased medical cases (32% to 41%) and, to a lesser extent, increasing surgical cases in public hospitals (30% to 33%).

Impact on CMI level and trend, across public and private hospitals (Objective 1)

Pre-intervention

The pre-intervention CMI monthly coefficients were 0.975, 1.284 and 1.783 for each of medical, surgical and mixed cases, respectively. Among medical cases, CMI was lowest for short-stays (0.352) and highest for long-stay cases (3.326).

Post-intervention

Overall the intervention had variable impacts depending on case type and hospital ownership, resulting in either increased or unchanged CMI (see [Table 2](#), [Figure 1](#) and [Figure 2](#)).

a. Medical cases

A decreasing pre-intervention trend was reversed, resulting in an increasing trend across all hospitals. Large level changes in CMI were also found at two months post-intervention, among public and private hospitals. No level changes were found among medium-stay cases. Overall, the main impact was on short and medium-stay cases, specifically a level change in the former and a trend change in the latter.

b. Surgical cases

An increasing pre-intervention trend continued unchanged in the post-intervention period across all hospitals. However, among public hospitals the absence of a pre-intervention trend was replaced with an increasing CMI trend.

c. Mixed cases

The only change found among mixed cases was an increasing trend in the medical component of private hospital cases. This had minimal impact on the overall CMI, as mixed cases comprised a very modest proportion of all hospitalizations.

We observed seasonality in medical CMI, and in particular medium-stay cases, with a March minor peak, a July–August trough, and a November–December major peak. Surgical casemix seasonality was limited to private hospitals only, with an April–June major peak, and a November–December minor peak. Mixed casemix seasonality had March and October peaks.

Changes in diagnoses and procedures (Objective 2)

The number of medical case discharge ICD codes used across 2013–2016 was 3,164 for short-stay, 5,828 for medium-stay

and 1,566 for long-stay cases, and 3,911 CPT codes for surgical cases. The 20 selected codes in each case type accounted for a majority of the change in CMI (see [Table 3](#)).

CMI change

Codes with the greatest changes from pre- to post-intervention are displayed in [Table 4](#) (for detailed changes see [Table 5](#)). These included 2,970 fewer cases of abdominal and pelvic pain, 698 fewer cases of intestinal infectious diseases, 1,001 fewer cases of fever of unknown origin, and 783 fewer cases of essential hypertension; altogether this represents about 4.3% of all annual medical cases. The greatest change on medium-stay CMI was due to a decrease of 2,237 cases of diarrhea and gastroenteritis. Two forms of shifting were also noted among several codes: from medium to short-stay cases (e.g. A09, I10, I20–I25.9), and from three-digit to four-digit ICD10 codes (e.g. J18, J44, P22). The greatest change on short-stay CMI was due to the improved coding of chemotherapy cases under the relevant code (Z51.1) rather than under neoplasms (C00–D49).

Given the magnitude of the change in vaginal deliveries, we further examined deliveries throughout 2013–2016 (vaginal and cesarean section). Private hospitals had 36.9% less vaginal deliveries (4,022) in 2015–2016 than in 2013–2014, while public hospitals increased by less than 1%. The decreasing trend in private hospitals began in early 2014, which was prior to the P4P intervention in late 2014. Concurrently, cesarean deliveries decreased at private hospitals by 7.5% and increased at public hospitals by 9.2%. The latter change also had an impact in increasing surgical CMI at public hospitals. Overall, vaginal deliveries under MoPH coverage decreased by 21.0% (18,843 to 14,894), while cesarean deliveries decreased by 2.5% in the aforementioned two-year periods (23,607 to 23,011).

CMI share change

The range of change of CMI share per code from pre- to post-intervention is shown in [Table 3](#). The conditions with the greatest change in their share of CMI included: percutaneous transluminal coronary angioplasty (PTCA) and vaginal delivery (surgical); malignant neoplasm of breast and acute lymphoblastic leukemia (short-stay); bronchopneumonia, diarrhea and gastroenteritis (medium-stay) (see [Table 5](#)).

Discussion

This study investigated the impact of integrating a P4P policy on the complexity of hospitalizations at hospitals contracted by the Lebanese Ministry of Public Health. We found that the intervention resulted in increased case complexity, specifically among medical cases. We used a systematic approach, first examining all hospitalizations by case type, lengths of stay and hospital ownerships for changes, and subsequently identifying and quantifying which diagnoses and procedures explain this change.

Case complexity was greatest for mixed cases, followed by surgical and medical cases. This was unsurprising considering that mixed cases are typically more complex and include ICU cases. Furthermore, a large proportion of medical cases are for short-term evaluation, basic treatment, or chemotherapy.

Table 2. Results of the interrupted time-series analysis (ITSA) on casemix index at Lebanese hospitals, adjusted for seasonality, 2011–2016, with the intervention point of August 2014 (two-month effect lag).

Case type		Hospitals	Before intervention			After intervention		
			Monthly CMI coefficient	TREND		TREND	LEVEL	
				% (CI)	Explained by		% (CI)	Explained by
Medical	All	0.975	↘ 0.10% (0.06 - 0.13%)	-	↗ 0.11% (0.02 - 0.21%)	Medium-stay cases	↗ 2.25% (0.51 - 3.98%)	Short-stay cases
	Public	0.941	↘ 0.17% (0.11 - 0.23%)	-	↗ 0.15% (0.06 - 0.22%)	Medium-stay cases	-	-
	Private	0.989	↘ 0.06% (0.01 - 0.11%)	-	↗ 0.19% (0.06 - 0.32%)	Short-stay cases	↗ 2.70% (0.15 - 5.24%)	Short-stay cases
Surgical	All	1.284	↗ 0.05% (0.01 - 0.10%)	-	↗ 0.14% (0.06 - 0.21%)	-	-	-
	Public	1.179	-	No trend	↗ 0.13% (0.02 - 0.24%)	-	-	-
	Private	1.326	↘ 0.12% (0.03 - 0.21%)	-	↗ 0.24% (0.13 - 0.35%) ²	-	-	-
Mixed	All	1.783	-	-	-	No trend	-	-
	Public	1.964	-	No trend	-	No trend	-	-
	Private	1.689	-	-	↗ 0.35% (0.10 - 0.60%) ²	-	-	-

¹p=0.06, ²p=0.11, ³p=0.33; no significant change between pre and post-intervention

Hospitals		MEDICAL			SURGICAL			MIXED	
		All stays	Short-stay	Medium-stay	Long-stay	All components	Medical component	Surgical component	
									ITSA aspect
Starting level	0.975	0.352	1.078	3.326	1.284	1.783	1.530	1.991	
Level at two months post-intervention	0.022	0.009	-	0.417	-	-	-	-	
Monthly trend, pre-intervention	-0.001	-	-0.001	-	0.001	-	-	-	
Monthly trend, post-intervention	0.001	0.001	0.001	-	0.002	-	-	-	
Monthly trend, change	0.002	0.001	0.002	-	-	-	-	-	

	Starting level	0.941	0.356	1.049	2.697	1.179	1.964	1.708	2.200	
Public	Level at two months post-intervention	-	0.007	-	-	-	-	-	-	-
	Monthly trend, pre-intervention	-0.002	-	-0.001	-	-	-	-	-	-
Monthly trend, post-intervention	0.001	-	0.001	-	-	0.002	-	-	-	
Monthly trend, change	0.003	-	0.003	-	-	0.003	-	-	-0.014	
Starting level	0.989	0.352	1.089	3.498	1.326	1.689	1.421	1.906		
Level at two months post-intervention	0.027	0.009	-	0.489	-	-	-	-	-	
Monthly trend, pre-intervention	-0.001	-	-	-	0.002	-	-	-	-	
Monthly trend, post-intervention	0.002	0.001	0.002	-	0.003	0.006	-	0.008		
Monthly trend, change	0.003	0.001	0.002	-	-	-	-	-	-	
(Numerator only results significant at p<0.05)										
c. Relative changes, using respective monthly casemix index starting level, percentages										
Hospitals	MEDICAL									
	ITSa aspect	All stays	Short-stay	Medium-stay	Long-stay	SURGICAL		MIXED		Surgical component
Starting level	0.975	0.352	1.078	3.326	1.284	1.783	1.530	1.991		
Level at two months post-intervention	2.25%	2.61%	-	0.125	-	-	-	-	-	
Monthly trend, pre-intervention	-0.10%	-	-0.06%	-	0.05%	-	-	-	-	
Monthly trend, post-intervention	0.11%	0.14%	0.09%	-	0.14%	-	-	-	-	
Monthly trend, change	0.22%	0.17%	0.16%	-	-	-	-	-	-	
Starting level	0.941	0.356	1.049	2.697	1.179	1.964	1.708	2.200		
Level at two months post-intervention	-	2.05%	-	-	-	-	-	-	-	
Monthly trend, pre-intervention	-0.17%	-	-0.12%	-	-	-	-	-	-	
Monthly trend, post-intervention	0.15%	-	0.11%	-	0.13%	-	-0.81%	-	-	
Monthly trend, change	0.31%	-	0.24%	-	0.22%	-	-	-	-	

	Starting level	0.989	0.352	1.089	3.498	1.326	1.689	1.421	1.906								
										MEDICAL						MIXED	
										ITSA aspect		All stays	Short-stay	Medium-stay	Long-stay	SURGICAL	
Private	Level at two months post-intervention	2.70%	2.56%	-	13.98%	-	-	-	-								
	Monthly trend, pre-intervention	-0.06%	-	-	-	0.12%	-	-	-								
	Monthly trend, post-intervention	0.19%	0.23%	0.15%	-	0.24%	0.35%	-	0.41%								
	Monthly trend, change	0.25%	0.28%	0.18%	-	-	-	-	-								
d. Confidence intervals for table 2b.																	
Hospitals																	
	Starting level	0.9599 - 0.9896	0.3416 - 0.3630	1.0498 - 1.1068	2.5017 - 4.1500	1.2666 - 1.3013	1.6649 - 1.9012	1.3860 - 1.6741	1.8782 - 2.1043								
	Level at two months post-intervention	0.0050 - 0.0388	0.0031 - 0.0153	-	0.1414 - 0.6932	-	-	-	-								
All	Monthly trend, pre-intervention	(0.0013 - 0.0006)	-	(0.0012 - 0.0003)	-	0.0001 - 0.0013	-	-	-								
	Monthly trend, post-intervention	0.0002 - 0.0020	0.0002 - 0.0007	0.0002 - 0.0017	-	0.0008 - 0.0027	-	-	-								
	Monthly trend, change	0.0012 - 0.0030	0.0002 - 0.0010	0.0008 - 0.0026	-	-	-	-	-								
Public	Starting level	0.9095 - 0.9723	0.3394 - 0.3728	1.0377 - 1.0593	1.3111 - 4.0834	1.0898 - 1.2691	1.7092 - 2.2188	1.5207 - 1.8947	1.8338 - 2.5664								
	Level at two months post-intervention	-	0.0012 - 0.0135	-	-	-	-	-	-								
	Monthly trend, pre-intervention	(0.0022 - 0.0010)	-	(0.0017 - 0.0010)	-	-	-	-	-								
	Monthly trend, post-intervention	0.0006 - 0.0021	-	0.0009 - 0.0015	-	0.0002 - 0.0028	-	(0.0254 - 0.0024)	-								
	Monthly trend, change	0.0019 - 0.0039	-	0.0020 - 0.0030	-	0.0007 - 0.0046	-	-	-								
	Starting level	0.9657 - 1.0117	0.3409 - 0.3628	1.0523 - 1.1258	2.7808 - 4.2154	1.2820 - 1.3702	1.6016 - 1.7767	1.2363 - 1.6055	1.8291 - 1.9833								
Private	Level at two months post-intervention	0.0015 - 0.0518	0.0012 - 0.0169	-	0.1433 - 0.8347	-	-	-	-								
	Monthly trend, pre-intervention	(0.0011 - 0.0001)	-	-	-	0.0004 - 0.0028	-	-	-								
	Monthly trend, post-intervention	0.0006 - 0.0032	0.0004 - 0.0013	0.0002 - 0.0030	-	0.0017 - 0.0047	0.0017 - 0.0101	-	0.0018 - 0.0139								
	Monthly trend, change	0.0012 - 0.0038	0.0004 - 0.0016	0.0004 - 0.0035	-	-	-	-	-								

(Numerator only results significant at p<0.05)



Figure 1. Medical short and medium-stay hospitalizations monthly casemix index (CMI), at public and private hospitals, unadjusted, 2011-2016.

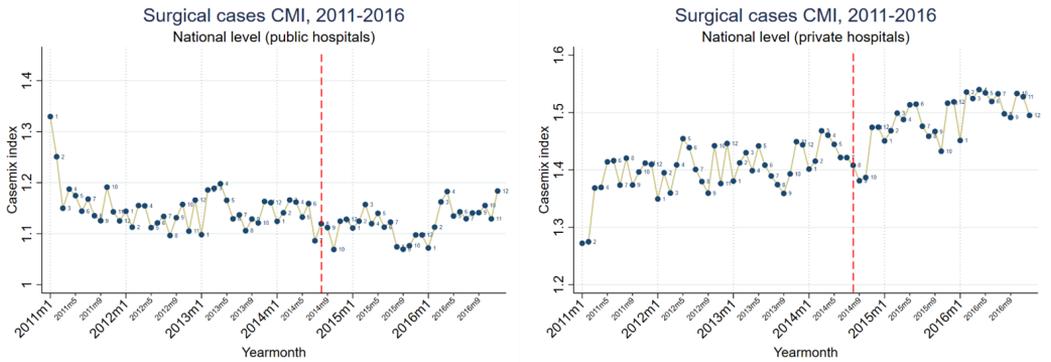


Figure 2. Surgical hospitalizations monthly casemix index (CMI), at public and private hospitals, 2011-2016.

Table 3. Summary results of casemix index (CMI) changes due diagnoses and procedures within different case types.

Case type		Number of codes (n)	Top 20 codes proportion of total cases (%)	Range of change in CMI share per code (%)
Medical	Short-stay	3,164	94.1%	-3.1% to 0.6%
	Medium-stay	5,828	80.6%	-0.8% to 1.4%
	Long-stay	1,566	55.8%	-0.7% to 1.4%
Surgical		3,911	96.2%	-0.6% to 1.3%

Similar reasoning explains medical casemix increasing with longer stays.

Medical casemix: decreased unnecessary hospitalizations and improved coding

During the pre-intervention period, medical casemix had a decreasing trend, possibly as a result of increasing unnecessary hospitalizations, for which an incentive exists. The intervention resulted in a reversal to an increasing casemix trend, as well as a short-term level change, at both public and private hospitals. The trend and level changes are attributable to medium and short-stay cases, respectively.

The decrease in diarrhea and gastroenteritis cases had a greater effect on medium-stay casemix than any other condition. Diarrhea and gastroenteritis cases are likely to be responsible for more unnecessary hospitalizations than any other diagnosis (Kronfol *et al.*, 2014; To *et al.*, 1996). Similar decreases were also found in abdominal and pelvic pain, essential hypertension cases and fever of unknown origin, all being major sources of unnecessary hospitalizations (Kronfol *et al.*, 2014). It is relevant to highlight that some of these conditions are more precisely symptoms (e.g. pain), which are not generally appropriate as discharge diagnoses and likely to be unnecessary hospitalizations. We find that the intervention decreased unnecessary hospitalizations by changing admission practices for these four conditions. The change in absolute cases for these conditions were at least one order of magnitude greater than any such changes since 2011 and are unlikely to be due to decreased disease burden.

We also found large shifts from medium to short-stay cases among each of diarrhea and gastroenteritis, abdominal and pelvic pain and essential hypertension cases. This suggests that hospitals not only decreased unnecessary hospitalizations, but also decreased unnecessary hospital stays. Although changes in length of stay are often difficult to associate with quality of care, among such conditions it is likely a result of improved hospital practices during hospitalization, in addition to the aforementioned improved pre-hospitalization practices. Such changes were seen among both public and private hospitals.

The large increase in chronic obstructive pulmonary disease (COPD) cases may be influenced by both practice changes and increasing burden. A steady increase in global COPD has been noted, with smoking prevalence and air pollution being the major risk factors (Lopez *et al.*, 2006). These two factors are also prevalent in Lebanon, and are likely to have contributed to the increased COPD burden (IHME, 2018). It is also possible that improved hospital admission practices influenced this increase. The increases in pneumonia and acute bronchitis cases are likely due to increased disease burden, with yearly and seasonal variations typical of infectious disease. We note the possibility of COPD acute exacerbation of COPD being misdiagnosed as acute bronchitis, as well as acute bronchitis also having smoking as a risk factor.

Given that this intervention had no direct impact on the separately pre-determined hospital budgets, it may be argued that the decreases in unnecessary hospitalizations provided hospitals with greater resources to hospitalize more necessary hospitalizations, such as COPD, pneumonia and acute bronchitis. Confirmation of such an impact would, however, require further investigation.

The decrease in ischemic heart disease cases, including a length of stay shift, may be associated with the concurrent increase in percutaneous transluminal coronary angioplasty (PTCA), which is indicated for certain ischemic heart disease cases. PTCA over-utilization has been documented in other healthcare systems and has had decreasing use in recent years (Mariotto *et al.*, 1999; Weiss & Elixhauser, 2014). It is not possible to evaluate medically indicated PTCA cases in this study, and further investigation would be required to ascertain the level of benefit or abuse.

The code-level findings indicate improved coding practices, which has been documented elsewhere to occur alongside real case changes following the introduction of CMI (Ginsburg & Carter, 1986; Goldfarb & Coffey, 1992). The improved coding for chemotherapy, particularly in breast cancer and leukemia, explained most of the short-stay casemix increase. Among medium and short-stays we find a shift to full coding digits and

Table 4. Diagnoses and procedures with the greatest change effect on casemix index (CMI), pre- versus post-intervention.

Description	ICD/CPT code	Major effects	Notes
Neoplasms	C00-D49	Increased ss-CMI (87%)	Greatest change on ss-CMI ↓10,179 cases, net Mainly due to malignant neoplasm of breast and acute lymphoblastic leukemia Concurrent with increase in chemotherapy Z51.1 ↑11,666 cases
Intestinal infectious diseases (category)	A00-A09		↓961 ms-cases; ↑263 ss-cases
Diarrhea and gastroenteritis of presumed infectious origin	A09	Increased ms-CMI (25%)	Greatest change on ms-CMI ↓2,237 ms-cases; ↑179 ss-cases
Unspecified non-infective gastroenteritis and colitis	K52.9	Decreased ms-CMI (2%)	↓745 ms-cases; ↓108 ss-cases
Abdominal and pelvic pain (category)	R10-R10.4		↑2,970 ms-cases, net
Abdominal and pelvic pain, other/ unspecified abdominal pain	R10, R10.4	Increased ms- and ss-CMI	↓1,975 ms-cases; ↑174 ss-cases
Influenza and pneumonia	J09-J18	Decreased ms-CMI (4%)	↑3,909 ms-cases; ↑298 ss-cases
Pneumonia, non-specific	J18		↓1,456 ms-cases
Pneumonia, specific	J18.0, J18.9		↑4,692 ms-cases
COPD	J44-J44.9	Increased ms-CMI (5%)	↑1,306 ms-cases, net
COPD with acute exacerbation	J44.1		↑625 ms-cases
COPD, non-specific	J44		↓234 ms-cases
Acute bronchitis	J20-J20.9	Decreased ms-CMI (3%)	↑1,145 ms-cases
	J20		↑747 ms-cases
Essential hypertension	I10		↓957 ms-cases; ↑174 ss-cases
Ischemic heart diseases	I20-I25.9	Decreased ms-CMI (7%)	↓1,100 ms-cases; ↑275 ss-cases Mainly due to angina pectoris and acute myocardial infarction
Fever of unknown origin	R50	Increased ms-CMI (3%)	↓989 ms-cases; ↓12 ss-cases
Stroke	I64	Decreased ms-CMI (2%)	↓383 ms-cases; ↑12 ss-cases
Respiratory distress of newborn, non-specific	P22		↓334 ms-cases
Respiratory distress of newborn, specific	P22.0		↑287 ms-cases
Vaginal delivery	F9410L1	Increased surgical CMI (43%)	↓3,939 cases Greatest change in absolute and in CMI share among all ICD/CPT codes
Percutaneous Transluminal Coronary Angioplasty (PTCA)	X2983/6	Increased surgical CMI (36%)	↑778 cases

ss, short-stay; ms, medium-stay; COPD, chronic obstructive pulmonary disease; ICD, International Classification of Diseases; CPT, Common Procedural Terminology.

Table 5. Identifying diagnoses/procedures with the greatest change in terms of casemix index, from pre-intervention to post-intervention.

a. Short-stay medical cases								
#	ICD code	Cases (n) 2013–2014	Cases (n) 2015–2016	Change in cases (n)	Change in cases (%)	Change in CMI share, relative	Change in CMI, relative	Description
1	C50	5,776	3,384	-2,392	-41.4%	-3.1%	43.3%	Malignant neoplasm of breast
2	C91.0	2,873	1,223	-1,650	-57.4%	-2.4%	19.9%	Acute lymphoblastic leukaemia
3	C76	297	89	-208	-70.0%	-0.6%	5.9%	Malignant neoplasm of other and ill-defined sites
4	C50.9	5,550	4,744	-806	-14.5%	-0.5%	5.2%	Breast, unspecified
5	C91	680	130	-550	-80.9%	-0.7%	5.1%	Lymphoid leukaemia
6	E88.0	92	2	-90	-97.8%	-0.4%	4.0%	Disorders of plasma-protein metabolism, not elsewhere classified
7	C90.0	1,135	474	-661	-58.2%	-0.9%	3.7%	Multiple myeloma
8	C25.9	524	130	-394	-75.2%	-0.6%	1.9%	Pancreas, unspecified
9	Z51	427	21	-406	-95.1%	-0.6%	1.5%	Other medical care
10	I10	697	871	174	25.0%	0.6%	0.7%	Essential (primary) hypertension
11	C18.9	1,046	691	-355	-33.9%	-0.5%	0.4%	Colon, unspecified
12	Z51.8	6	106	100	1666.7%	0.1%	0.3%	Other specified medical care
13	I46.9	82	103	21	25.6%	0.2%	0.3%	Cardiac arrest, unspecified
14	C56	1,414	947	-467	-33.0%	-0.6%	0.3%	Malignant neoplasm of ovary
15	N18	86	134	48	55.8%	0.2%	0.3%	Chronic renal failure
16	I25.1	13	38	25	192.3%	0.1%	0.3%	Atherosclerotic heart disease
17	I21.9	25	49	24	96.0%	0.1%	0.3%	Acute myocardial infarction, unspecified
18	I20.0	177	244	67	37.9%	0.3%	0.3%	Unstable angina
19	A09	947	1,126	179	18.9%	0.5%	0.3%	Diarrhoea and gastroenteritis of presumed infectious origin
20	I48	77	142	65	84.4%	0.2%	0.3%	Atrial fibrillation and flutter
b. Medium-stay medical cases								
#	ICD code	Cases (n) 2013–2014	Cases (n) 2015–2016	Change in cases (n)	Change in cases (%)	Change in CMI share, relative	Change in CMI, relative	Description
1	A09	18,440	16,203	-2,237	-12.1%	-0.8%	25.0%	Diarrhoea and gastroenteritis of presumed infectious origin
2	I21	745	350	-395	-53.0%	-0.5%	7.8%	Acute myocardial infarction
3	P22.0	420	707	287	68.3%	0.3%	6.1%	Respiratory distress syndrome of newborn
4	J18.0	6,487	10,134	3,647	56.2%	1.4%	5.6%	Bronchopneumonia, unspecified
5	J44.1	439	1,064	625	142.4%	0.5%	5.0%	Chronic obstructive pulmonary disease with acute exacerbation, unspecified
6	R10	11,406	10,034	-1,372	-12.0%	-0.6%	4.4%	Abdominal and pelvic pain
7	I46.9	127	262	135	106.3%	0.2%	3.3%	Cardiac arrest, unspecified
8	R50	3,348	2,359	-989	-29.5%	-0.4%	2.8%	Fever of unknown origin
9	K52.9	2,343	3,088	745	31.8%	0.2%	2.4%	Noninfective gastroenteritis and colitis, unspecified

10	Z38.0	75	262	187	249.3%	0.2%	2.4%	Singleton, born in hospital
11	P22	1,369	1,035	-334	-24.4%	-0.3%	2.4%	Respiratory distress of newborn
12	J20	2,291	3,038	747	32.6%	0.2%	2.4%	Acute bronchitis
13	R07.0	306	137	-169	-55.2%	-0.2%	2.0%	Pain in throat
14	A41.9	525	708	183	34.9%	0.2%	1.8%	Septicaemia, unspecified
15	I64	1,454	1,071	-383	-26.3%	-0.3%	1.7%	Stroke, not specified as haemorrhage or infarction
16	A08.4	206	804	598	290.3%	0.2%	1.3%	Viral intestinal infection, unspecified
17	J18.9	2,079	3,124	1,045	50.3%	0.6%	1.2%	Pneumonia, unspecified
18	J98	166	3	-163	-98.2%	-0.2%	1.1%	Other respiratory disorders
19	R10.4	3,167	2,564	-603	-19.0%	-0.2%	1.1%	Other and unspecified abdominal pain
20	G81.9	428	198	-230	-53.7%	-0.2%	1.0%	Hemiplegia, unspecified
c. Long-stay medical cases								
#	ICD code	Cases (n) 2013–2014	Cases (n) 2015–2016	Change in cases (n)	Change in cases (%)	Change in CMI share, relative	Change in CMI, relative	Description
1	P22	215	108	-107	-49.8%	-0.4%	17.9%	Respiratory distress of newborn
2	F20.9	1	31	30	3000.0%	1.4%	4.9%	Schizophrenia, unspecified
3	F25.0	2	26	24	1200.0%	1.1%	3.1%	Schizoaffective disorder, manic type
4	A41.9	74	121	47	63.5%	0.3%	2.7%	Septicaemia, unspecified
5	I20	74	32	-42	-56.8%	-0.2%	2.5%	Angina pectoris
6	F10	39	6	-33	-84.6%	-0.1%	2.5%	Mental and behavioural disorders due to use of alcohol
7	F11	45	13	-32	-71.1%	0.0%	2.4%	Mental and behavioural disorders due to use of opioids
8	G81	49	11	-38	-77.6%	-0.2%	2.4%	Hemiplegia
9	J98	42	0	-42	-100.0%	-0.2%	2.2%	Other respiratory disorders
10	I50	139	102	-37	-26.6%	-0.1%	2.1%	Heart failure
11	F20.0	88	58	-30	-34.1%	0.0%	2.0%	Paranoid schizophrenia
12	Z38.0	18	55	37	205.6%	0.2%	1.7%	Singleton, born in hospital
13	P22.0	114	149	35	30.7%	0.3%	1.5%	Respiratory distress syndrome of newborn
14	F11.2	0	16	16		0.7%	1.4%	Mental and behavioural disorders due to use of opioids, dependence syndrome
15	M96.9	16	0	-16	-100.0%	-0.7%	1.4%	Postprocedural musculoskeletal disorder, unspecified
16	F25	45	21	-24	-53.3%	0.0%	1.3%	Schizoaffective disorders
17	R10.0	28	6	-22	-78.6%	0.0%	1.0%	Acute abdomen
18	I46.9	38	73	35	92.1%	0.3%	1.0%	Cardiac arrest, unspecified
19	J18.9	72	99	27	37.5%	0.2%	0.9%	Pneumonia, unspecified
20	G45.9	13	37	24	184.6%	0.1%	0.9%	Transient cerebral ischaemic attack, unspecified

d. Surgical cases								
#	CPT code	Cases (n) 2013–2014	Cases (n) 2015–2016	Change in cases (n)	Change in cases (%)	Change in CMI share, relative	Change in CMI, relative	Description
1	F9410L1	18,833	14,894	-3,939	-20.9%	-0.6%	43.4%	Vaginal delivery only including postpartum care
2	X2986L1	1,153	1,687	534	46.3%	1.3%	33.7%	Percutaneous transluminal coronary angioplasty: with stent; each additional vessel
3	M7447G1	745	954	209	28.1%	0.5%	4.9%	Arthroplasty, knee condyle and plateau ("total knee replacement")
4	C3510G	1,323	1,179	-144	-10.9%	-0.3%	3.2%	Coronary artery bypass, vein only: any number W/CPB
5	X2983L1	3,154	3,398	244	7.7%	0.5%	2.2%	Percutaneous transluminal coronary angioplasty: with stent
6	C3000Gb	423	809	386	91.3%	0.4%	1.5%	Cardiopulmonary bypass, including cannulation: add to primary procedure (W/CPB)
7	X3548L	11,900	12,556	656	5.5%	0.2%	1.1%	Combined left heart catheterization, selective coronary angiography, one or more.
8	C3411G	189	115	-74	-39.2%	-0.2%	1.0%	Replacement, aortic valve: w/wout aortic annulus enlargement, mechanical.
9	M7125G1	294	447	153	52.0%	0.2%	0.7%	Partial hip replacement, prosthesis (eg, femoral stem prosthesis, bipolar arthroplasty)
10	R0140Gb	3,600	3,959	359	10.0%	0.0%	0.7%	Submucous resection turbinate, complete or partial
11	M2842G1	346	450	104	30.1%	0.2%	0.7%	Spinal instrumentation, posterior: segmental fixation.
12	M7130G1	622	715	93	15.0%	0.2%	0.6%	Arthroplasty, acetabular and proximal femoral prosthetic replacement (Age <60y)
13	F9812G	3,386	2,959	-427	-12.6%	-0.1%	0.5%	Treatment of incomplete or missed abortion, any trimester, completed surgically
14	M2843G1	108	182	74	68.5%	0.1%	0.4%	Spinal instrumentation, posterior: segmental fixation
15	X2987L	23	58	35	152.2%	0.1%	0.4%	Closure of PDA or ASD by Amplatzer
16	C3411Ga	44	9	-35	-79.5%	-0.1%	0.3%	Replacement, aortic valve: w/wout aortic annulus enlargement, biological.
17	C5301G	132	94	-38	-28.8%	-0.1%	0.2%	Thromboendarterectomy, with or without patch graft
18	U2332Gb	656	852	196	29.9%	0.0%	0.2%	Cystourethroscopy, with insertion of indwelling ureteral stent
19	D2820G	2,651	2,349	-302	-11.4%	0.0%	0.2%	Tonsillectomy with or without adenoidectomy
20	M7715G	111	25	-86	-77.5%	-0.1%	0.2%	Osteoplasty, tibia and fibula, lengthening

ICD, International Classification of Diseases; CPT, Common Procedural Terminology; CMI, casemix index; CPB, cardiopulmonary bypass; PDA, patent ductus arteriosus; ASD, atrial septal defect.

more specific codes, specifically among neoplasms, pneumonia, COPD, respiratory distress of newborn, and diarrhea and gastroenteritis.

Surgical casemix

Changes in surgical casemix were limited to public hospitals only, with an increasing trend emerging in the post-intervention period. This was explained by increased cesarean deliveries in public hospitals, which compensated a concurrent decrease in private hospitals, and was identified as a trend preceding the intervention by several months, and therefore not likely impacted by the intervention itself.

Private hospitals continued their increasing casemix trend without any change following the intervention. This trend was explained by two factors: increasing PTCA cases throughout 2011–2016 and decreasing vaginal deliveries since early 2014. The former may also be an example of private hospitals being faster adopters of new technologies than public hospitals or having greater potential for over-utilization. We had expected to find new (and costlier) procedures such as PTCA to explain a large proportion of an increasing surgical casemix at all hospitals. However, with the exception of PTCA, adoption of new procedures had a negligible effect on surgical casemix.

The large decrease in vaginal deliveries at private hospitals coincides with the sharp increase into Lebanon of refugees from Syria in early 2014. This change was not compensated by public hospitals, whose vaginal deliveries remained generally unchanged. Deliveries had been documented in 2013 as the leading cause for hospitalization among Syrian refugees in Lebanon (Huster *et al.*, 2014). This suggests that, under MoPH coverage, Lebanese women having vaginal deliveries had decreased access to private hospitals, or otherwise opted to deliver outside of this coverage (out-of-pocket payment). The hospitalization access of refugee and citizen communities requires further investigation, including the role of private hospitals.

Improvement potential, hospital ownership, and seasonality

Hospitals had a greater improvement space in medical cases than in surgical cases. Prior to the intervention, medical ICD10 coding had no link to reimbursement (fee-for-service), unlike surgical codes (flat-fee). The intervention incentivized improved medical coding (for accurate CMI assessment), but the same was not relevant for surgical coding.

Private hospitals had a greater overall casemix than public hospitals, suggesting that more complex cases were admitted to private hospitals, as had been observed in previous MoPH investigations (Ammar *et al.*, 2013). However, private and public hospitals had similar short-stay casemix, while public hospitals had a higher casemix for mixed cases. Two factors at private hospitals that are relevant to note are their possibly greater capacity for accurate coding and for healthcare technology. The former has been found elsewhere to explain some of

the casemix gap between public and private hospitals, while the latter is associated with increased casemix (Mendez *et al.*, 2014; Park *et al.*, 2017).

We also note the increasing proportion of hospitalizations at public hospitals throughout 2011–2016, which continues the trend observed by the MoPH for preceding years. Since the proportion of contracted public and private hospitals had remained relatively unchanged, this is attributable to the MoPH's broad policy of increasing the service delivery carried out at public rather than private hospitals.

The seasonality in the casemix within different case types is likely explained by a combination of disease burden variation throughout the year, and pre-planned hospitalizations that avoid vacation periods. The November–December peak for medical medium-stay may be related to influenza seasonality in Lebanon (World Health Organization, 2019). Further investigations would be required to associate casemix seasonality with specific diseases or conditions. It is noteworthy that surgical casemix seasonality was found only in private hospitals, peaking during the three months preceding the summer period, which suggests that pre-planned hospitalizations are more common in private than in public hospitals.

P4P design

The importance of detailing P4P designs has been highlighted, particularly considering the heterogeneity of such interventions (Chee *et al.*, 2016). Notable design features of the MoPH P4P is the linkage between performance and reimbursement tier, rather than a bonus/penalty. To our knowledge this has not been undertaken elsewhere. In Lebanon, such a feature was a result of the political and financial non-feasibility of having other financial mechanisms to reward hospitals for improved performance. However, this has the benefit of being more sustainable in avoiding bonus financing and maintaining a system aspect (tiers) familiar to hospitals. In effect, the design integrated P4P into the system of determining hospital reimbursement tier (Soucat *et al.*, 2017).

Another feature is the inclusion of CMI directly within the performance scoring, rather than as a risk adjustor for other outcomes (e.g. readmissions). This focus on casemix was intended to address unnecessary hospitalization, as well as the absence of risk-adjustment in hospital assessment. It also formed an objective that may be influenced by a wide range of hospitalizations, rather than a narrow set.

The type and magnitude of incentives are also important factors in determining P4P impact. Having casemix incentivized outside of a prospective payment system or similar approach likely limited the potential of providers to engage in up-coding or otherwise game the system, as has been observed in other casemix evaluations (Radu *et al.*, 2010; Sukul *et al.*, 2019). Such behavior is further limited as providers do not have access to the weights used in casemix evaluation.

Strengths and limitations

The use of ITS analysis with a large number of pre and post-intervention data points allowed us to account for background effects in what is a complex environment. Such effects may include other interventions or events that may have impacted hospitalization practice or burden of disease. No relevant system interventions were carried out by the MoPH throughout 2011–2016, besides the new intervention evaluated in this work. Other quality-focused activities undertaken by hospitals individually have not been assessed, though these are less likely to have system-level impact.

We relied on the data that is collected and input at hospital-level, and subsequently used to calculate CMI. As such, the validity of the casemix results relies on the quality of data input, specifically ICD10 and CPT codes. The intervention of August 2014 included advice to hospitals regarding improved coding accuracy, whose short-term results have been detected with CMI level changes. However, we do not account for coding quality initiatives that may have occurred within hospitals.

We recognize that the inclusion of age and comorbidities in the calculation of casemix index would allow a more accurate measure of hospitalization case complexity. Within the Lebanese healthcare setting there currently is insufficient information for selecting diagnoses and procedures for adjustment by age. The absence of comorbidity is due to the lack of routine recording of this variable across most hospitals, and policymaker concerns regarding potential miss-use. Age and comorbidities remain important areas for future development of casemix calculation and pay-for-performance, but their absence does not negate the findings of the current investigation.

Conclusions

This research suggests that the integration of a hospital performance-payment policy in 2014 increased the effectiveness of the healthcare system, primarily due to decreased unnecessary hospitalizations, as well as a decreased length of stay of such cases. An improvement in quality of care may subsequently benefit patient health and resource use (efficiency). The new policy also led to improved discharge coding quality. Although unrelated to effectiveness, the latter finding increases the validity of evidence and policymaking that makes use of such information, including but not limited to the P4P design. Changes in unnecessary hospitalizations took place at a gradual pace compared to the more immediate coding practice changes. We also confirm that CMI can be appropriate tool to detect changes in hospitalizations or performance improvement.

By using a systems perspective, we were able to investigate the impact across different hospitalization case types, length of stay and hospital ownership, and went further to quantify and attribute changes to specific diagnoses and procedures. We

also identified issues relevant for further investigation and policymaking (e.g. vaginal deliveries, COPD, ischemic heart disease treatment).

A similar analytical approach using interrupted time series may be used in the evaluation of other interventions on the hospitalization system. The algorithms developed for this research may also be adapted to investigate specific issues, as well as by a regulator or payer to actively monitor hospitalization trends across code and hospital attributes. This would support keeping providers accountable and increase the responsiveness capacity to address non-desirable or harmful hospitalization practices, such as unnecessary hospitalizations.

Despite the generally unfavorable evidence regarding the effectiveness of P4P in hospital settings, our findings suggest that certain P4P designs may be effective in specific contexts, such as that of the MoPH and Lebanese hospitals. This may be in part due to its integration within the system, rather than as a stand-alone external intervention, in addition to the existing space for improvement.

Future investigations would be required to evaluate longer-term impact, which would inform whether the impact was a one-time result of introducing a new policy, or if it retains sustainable benefits. This would necessarily require continued commitment and capacity-building by the MoPH towards this process.

Our findings suggest that effective hospital regulation can be achieved through the systematic collection and analysis of readily available routine data. Our analytical approach to such data reveals relevant patterns of change to performance measures. LMICs that lack casemix adjustment and incentives for improving hospital performance may choose to adopt similar approaches and monitoring systems to measure and improve hospital performance over time.

Data availability

Underlying data

The source data are owned by the Lebanese Ministry of Public Health and in line with the IRB approval granted for this study, the authors are not permitted to share the source patient-level data. In compliance with the MoPH's obligation on data privacy, the underlying data are accessible in a de-identified form upon request to the Ministry of Public Health (directorgeneral@moph.gov.lb), including a justification for the request.

Acknowledgements

We would like to thank Anne Mills (London School of Hygiene and Tropical Medicine, project senior advisor); Anders Halling and Johan Jarl (Lund University, for feedback on an earlier draft); and the contributions of the following individuals in hospital performance activities at the Lebanese MoPH: Jihad Makouk, Rita Freiha, Abeer El Halabi and Elise Barakat.

References

- Ammar W: **Health system and reform in Lebanon**. Entreprise universitaire d'études et de publications. 2003.
[Reference Source](#)
- Ammar W: **Health beyond politics**. 2009.
[Reference Source](#)
- Ammar W, Khalife J, El-Jardali F, et al.: **Hospital Accreditation, Reimbursement and Case Mix: Links and Insights for Contractual Systems**. *BMC Health Serv Res*. 2013; **13**: 505.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Ammar W, Wakim IR, Hajj I: **Accreditation of Hospitals in Lebanon: A Challenging Experience**. *East Mediterr Health J*. 2007; **13**(1): 138–149.
[PubMed Abstract](#)
- Bernal JL, Cummins S, Gasparrini A: **Interrupted Time Series Regression for the Evaluation of Public Health Interventions: A Tutorial**. *Int J Epidemiol*. 2016; **46**(1): 348–355.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Borghji J, Little R, Binyaruka P, et al.: **In Tanzania, the Many Costs of Pay-For-Performance Leave Open to Debate Whether the Strategy Is Cost-Effective**. *Health Aff (Millwood)*. 2015; **34**(3): 406–414.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Caminiti C, Meschi T, Braglia L, et al.: **Reducing Unnecessary Hospital Days to Improve Quality of Care Through Physician Accountability: A Cluster Randomised Trial**. *BMC Health Serv Res*. 2013; **13**: 14.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Chee TT, Ryan AM, Wasfy JH, et al.: **Current State of Value-Based Purchasing Programs**. *Circulation*. 2016; **133**(22): 2197–2205.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Das A, Gopalan SS, Chandramohan D: **Effect of Pay for Performance to Improve Quality of Maternal and Child Care in Low- And Middle-Income Countries: A Systematic Review**. *BMC Public Health*. 2016; **16**: 321.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Donabedian A: **Evaluating the quality of medical care**. *Milbank Mem Fund Q*. 1966; **44**(3): Suppl: 166–206.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Donabedian A: **The quality of care. How can it be assessed?** *JAMA*. 1988; **260**(12): 1743–8.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Eijkenaar F, Emmert M, Scheppach M, et al.: **Effects of pay for performance in health care: a systematic review of systematic reviews**. *Health Policy*. 2013; **110**(2–3): 115–30.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Emmert M, Eijkenaar F, Kemter H, et al.: **Economic evaluation of pay-for-performance in health care: a systematic review**. *Eur J Health Econ*. 2012; **13**(6): 755–67.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Gergen J, Josephson E, Coe M, et al.: **Quality of Care in Performance-Based Financing: How It Is Incorporated in 32 Programs Across 28 Countries**. *Glob Health Sci Pract*. 2017; **5**(1): 90–107.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Gillam SJ, Sriwardena AN, Steel N: **Pay-for-performance in the United Kingdom: impact of the quality and outcomes framework: a systematic review**. *Ann Fam Med*. 2012; **10**(5): 461–8.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Ginsburg PB, Carter GM: **Medicare case-mix index increase**. *Health Care Financ Res*. 1986; **7**(4): 51–65.
[PubMed Abstract](#) | [Free Full Text](#)
- Goldfarb MG, Coffey RM: **Change in the Medicare case-mix index in the 1980s and the effect of the prospective payment system**. *Health Serv Res*. 1992; **27**(3): 385–415.
[PubMed Abstract](#) | [Free Full Text](#)
- Grace FC, Meurck CS, Head BW, et al.: **An analysis of policy levers used to implement mental health reform in Australia 1992-2012**. *BMC Health Serv Res*. 2015; **15**: 479.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Gupta A, Fonarow GC: **The Hospital Readmissions Reduction Program-learning from failure of a healthcare policy**. *Eur J Heart Fail*. 2018; **20**(8): 1169–1174.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Habicht JP, Victora CG, Vaughan JP: **Evaluation designs for adequacy, plausibility and probability of public health programme performance and impact**. *Int J Epidemiol*. 1999; **28**(1): 10–8.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Hornbrook MC: **Hospital case mix: its definition, measurement and use. Part II: Review of alternative measures**. *Med Care Rev*. 1982; **39**(2): 73–123.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Huster KM, Patterson N, Schilperoord M, et al.: **Cesarean sections among Syrian refugees in Lebanon from december 2012/january 2013 to june 2013: probable causes and recommendations**. *Yale J Biol Med*. 2014; **87**(3): 269–288.
[PubMed Abstract](#) | [Free Full Text](#)
- IHME: **Lebanon profile**. 2018.
[Reference Source](#)
- Jha AK, Joynt KE, Orav EJ, et al.: **The long-term effect of premier pay for performance on patient outcomes**. *N Engl J Med*. 2012; **366**(17): 1606–1615.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Khalife J, Rafeh N, Makouk J, et al.: **Hospital Contracting Reforms: The Lebanese Ministry of Public Health Experience**. *Health Systems & Reform*. 2017; **3**(1): 34–41.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Kristensen SR, Meacock R, Turner AJ, et al.: **Long-term effect of hospital pay for performance on mortality in England**. *N Engl J Med*. 2014; **371**(6): 540–548.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Kronfol N, Khalife J, Romanos J, et al.: **Observations on the Ministry of Public Health program of support to the hospitalization of patients in Lebanon**. *J Med Liban*. 2014; **62**(1): 33–39.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Kruk M, Freedman L: **Assessing Health System Performance in Developing Countries: A Review of the Literature**. 2008.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Lichtig L: **Hospital Information System for Case Mix Management**. New York: John Wiley and Sons Press. 1986.
- Lindenauer PK, Remus D, Roman S, et al.: **Public reporting and pay for performance in hospital quality improvement**. *N Engl J Med*. 2007; **356**(5): 486–496.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Lopez AD, Shibuya K, Rao C, et al.: **Chronic obstructive pulmonary disease: current burden and future projections**. *Eur Respir J*. 2006; **27**(2): 397–412.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Lopez Bernal J, Soumerai S, Gasparrini A: **A methodological framework for model selection in interrupted time series studies**. *J Clin Epidemiol*. 2018; **103**: 82–91.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Macinko J, Dourado I, Aquino R, et al.: **Major Expansion Of Primary Care In Brazil Linked To Decline In Unnecessary Hospitalization**. *Health Aff (Millwood)*. 2010; **29**(12): 2149–2160.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Mandavira R, Mehta N, Schilder A, et al.: **Effectiveness of UK provider financial incentives on quality of care: a systematic review**. *Br J Gen Pract*. 2017; **67**(664): e800–e815.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Mariotto A, Zeni L, Selle V, et al.: **Utilization of percutaneous transluminal coronary angioplasty for quality assurance in health care from 1983 to 1996**. *Int J Technol Assess Health Care*. 1999; **15**(3): 473–479.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Mendez CM, Harrington DW, Christenson P, et al.: **Impact of hospital variables on case mix index as a marker of disease severity**. *Popul Health Manag*. 2014; **17**(1): 28–34.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Naci H, Soumerai SB: **History Bias, Study Design, and the Unfulfilled Promise of Pay-for-Performance Policies in Health Care**. *Prev Chronic Dis*. 2016; **13**: E82.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Park YT, Lee J, Lee J: **Association between Health Information Technology and Case Mix Index**. *Healthc Inform Res*. 2017; **23**(4): 322–327.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Penfold RB, Zhang F: **Use of interrupted time series analysis in evaluating health care quality improvements**. *Acad Pediatr*. 2013; **13**(6): Suppl: S38–44.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Radu CP, Chiriac DN, Vladescu C: **Changing patient classification system for hospital reimbursement in Romania**. *Croat Med J*. 2010; **51**(3): 250–258.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Roberts M, Hsiao W, Berman P, et al.: **Getting Health Reform Right: A Guide to Improving Performance and Equity**. 2004; 4(4).
[Publisher Full Text](#)
- Ryan AM, Burgess JF Jr, Pesko MF, et al.: **The early effects of Medicare's mandatory hospital pay-for-performance program**. *Health Serv Res*. 2015; **50**(1): 81–97.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Services CFMM: **Details for title: FY 2011 Final Rule Data Files**. 2011.
[Reference Source](#)
- Smith R, Hanson K: **Health systems in low- and middle-income countries: an economic and policy perspective**. Oxford, Oxford University Press. 2012.
[Publisher Full Text](#)
- Soucat A, Dale E, Mathauer I, et al.: **Pay-for-Performance Debate: Not Seeing the Forest for the Trees**. *Health Systems & Reform*. 2017; **3**(2): 74–79.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Stranges E, Friedman B: **Potentially Preventable Hospitalization Rates Declined for Older Adults, 2003-2007: Statistical Brief #83**. *Healthcare Cost*

and Utilization Project (HCUP) Statistical Briefs. Rockville (MD). 2006.

[PubMed Abstract](#)

Sukul D, Hoffman GJ, Nuliyalu U, *et al.*: **Association Between Medicare Policy Reforms and Changes in Hospitalized Medicare Beneficiaries' Severity of Illness.** *JAMA Netw Open.* 2019; **2**(5): e193290.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Sutton M, Nikolova S, Boaden R, *et al.*: **Reduced mortality with hospital pay for performance in England.** *N Engl J Med.* 2012; **367**(19): 1821–8.

[PubMed Abstract](#) | [Publisher Full Text](#)

Tashobya CK, da Silveira VC, Ssengooba F, *et al.*: **Health systems performance assessment in low-income countries: learning from international experiences.** *Global Health.* 2014; **10**: 5.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

To T, Feldman W, Young W, *et al.*: **Hospitalization rates of children with gastroenteritis in Ontario.** *Can J Public Health.* 1996; **87**(1): 62–5.

[PubMed Abstract](#)

United Nations, D. o. E. a. S. A., Population Division: **World Population Prospects: The 2017 Revision.** DVD Edition. 2017.

[Reference Source](#)

Van Herck P, De Smedt D, Annemans L, *et al.*: **Systematic review: Effects, design choices, and context of pay-for-performance in health care.** *BMC Health Serv Res.* 2010; **10**: 247.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Wadhera RK, Joynt Maddox KE, Wasfy JH, *et al.*: **Association of the Hospital Readmissions Reduction Program With Mortality Among Medicare Beneficiaries Hospitalized for Heart Failure, Acute Myocardial Infarction, and Pneumonia.** *JAMA.* 2018; **320**(24): 2542–2552.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Wasfy JH, Zigler CM, Choirat C, *et al.*: **Readmission Rates After Passage of the Hospital Readmissions Reduction Program: A Pre-Post Analysis.** *Ann Intern*

Med. 2017; **166**(5): 324–331.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Weiss AJ, Elixhauser A: **Trends in Operating Room Procedures in U.S. Hospitals, 2001– 2011.** *HCUP Statistical Brief #171.* Rockville, MD., Agency for Healthcare Research and Quality. 2014.

[PubMed Abstract](#)

Werner RM, Kolstad JT, Stuart EA, *et al.*: **The effect of pay-for-performance in hospitals: lessons for quality improvement.** *Health Aff (Millwood).* 2011; **30**(4): 690–698.

[PubMed Abstract](#) | [Publisher Full Text](#)

Witter S, Fretheim A, Kessy FL, *et al.*: **Paying for performance to improve the delivery of health interventions in low- and middle-income countries.** *Cochrane Database Syst Rev.* 2012; (2): CD007899.

[PubMed Abstract](#) | [Publisher Full Text](#)

World Health Organization: **International Statistical Classification of Diseases and Related Health Problems.** 10th Revision 2019. 2006.

[Reference Source](#)

World Health Organization: **Systems thinking for health systems strengthening.** D. de Savigny and T. Adam. Geneva, Alliance for Health Policy and Systems Research, World Health Organization. 2009.

[Reference Source](#)

World Health Organization: **Health policy and systems research: a methodology reader.** L. Gilson. Geneva, Alliance for Health Policy and Systems Research, World Health Organization. 2012.

[Reference Source](#)

World Health Organization: **FluNet, Global Influenza Surveillance and Response System (GISRS).** 2019.

[Reference Source](#)

Yang CM, Reinke W: **Feasibility and validity of International Classification of Diseases based case mix indices.** *BMC Health Serv Res.* 2006; **6**: 125.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Open Peer Review

Current Peer Review Status:  

Version 2

Reviewer Report 18 December 2020

<https://doi.org/10.21956/wellcomeopenres.18000.r41802>

© 2020 Shan L. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Linghan Shan

Department of Social Medicine, School of Public Health, Harbin Medical University, Harbin, Heilongjiang, China

Dear author, I think this is a high-quality study that can provide evidence for many policy researchers and policy makers.

Is the work clearly and accurately presented and does it cite the current literature?

Yes

Is the study design appropriate and is the work technically sound?

Yes

Are sufficient details of methods and analysis provided to allow replication by others?

Yes

If applicable, is the statistical analysis and its interpretation appropriate?

I cannot comment. A qualified statistician is required.

Are all the source data underlying the results available to ensure full reproducibility?

No source data required

Are the conclusions drawn adequately supported by the results?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Health policy assessment; health insurance; health system reform

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Version 1

Reviewer Report 10 August 2020

<https://doi.org/10.21956/wellcomeopenres.17338.r39557>

© 2020 Shan L. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**Linghan Shan**

Department of Social Medicine, School of Public Health, Harbin Medical University, Harbin, Heilongjiang, China

Thank you for writing such a valuable and enlightening article, it is of an acceptable scientific standard, however, I have a number of small changes to the article, as follows:

1. In methods it was said that "We applied a systematic approach to assess the impact of the new policy on hospital performance." What is the framework of the systematic approach, what are the differences from previous studies, or what is the basis for it? I see that your explanation of "systematic approach" is scattered in the article. I hope it can be more clearly expressed.
2. In the Introduction it was said "The MoPH is the largest public payer, covering hospitalization for about 52% of Lebanese, who otherwise lack any insurance coverage for hospitalization (Ammar, 2009)." What is the coverage now? I want to know this information because it is directly related to the impact of payment reform on hospitals. High coverage means that the payer has strong bargaining power, so the reform of payment mode will bring greater binding force to hospitals, otherwise, the binding force will be small.
3. I would like to see more details about the pre-reform payment mechanism, restrictive conditions of reform (such as you said "lack of local Diagnosis Related Groups (DRGs) on which most casemix systems rely on"), characteristics, advantages and disadvantages of the reform.
4. I see that your suggestions are all about recommending this reform measure. Do you have any relevant suggestions for the sustainable development of this reform measure? Is there a need for more supporting policies to further improve efficiency? Does this policy reform need further improvement?

Is the work clearly and accurately presented and does it cite the current literature?

Yes

Is the study design appropriate and is the work technically sound?

Yes

Are sufficient details of methods and analysis provided to allow replication by others?

Yes

If applicable, is the statistical analysis and its interpretation appropriate?

I cannot comment. A qualified statistician is required.

Are all the source data underlying the results available to ensure full reproducibility?

No source data required

Are the conclusions drawn adequately supported by the results?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Health policy assessment; health insurance; health system reform

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 06 Oct 2020

Jade Khalife, Faculty of Medicine at Lund University, Lund, Sweden

We are grateful for the valuable feedback received from the reviewer, which we have used to improve the manuscript. Kindly find below our responses on a point-by-point basis.

- In methods it was said that "We applied a systematic approach to assess the impact of the new policy on hospital performance." What is the framework of the systematic approach, what are the differences from previous studies, or what is the basis for it? I see that your explanation of "systematic approach" is scattered in the article. I hope it can be more clearly expressed.

We have clarified our approach at the beginning of the discussion section to define what we mean by systematic approach. Our approach involved at a first stage examining all hospitalizations by case type, lengths of stay and hospital ownerships for casemix index changes. This was followed by a second stage for identifying and quantifying which diagnoses and procedures explain changes identified during the first stage.

- In the Introduction it was said "The MoPH is the largest public payer, covering hospitalization for about 52% of Lebanese, who otherwise lack any insurance coverage for hospitalization (Ammar, 2009)." What is the coverage now? I want to know this information because it is directly related to the impact of payment reform on hospitals. High coverage means that the payer has strong bargaining power, so the reform of payment mode will bring greater binding force to hospitals, otherwise, the binding force will be small.

The coverage in 2014 is expected to be similar (52%), according to the Ministry of Public Health. A more updated estimate is not available. However, this figure is expected to have increased since about 2018, as a consequence of increasing unemployment, which results in more citizens having coverage transferred from the National Social Security Fund to the Ministry of Public Health.

- I would like to see more details about the pre-reform payment mechanism, restrictive conditions of reform (such as you said “lack of local Diagnosis Related Groups (DRGs) on which most casemix systems rely on”), characteristics, advantages and disadvantages of the reform.

We have added a paragraph under ‘Policy reform’ to provide more detailed background information on the pre-reform period. The reform was used to categorize hospitals into the three reimbursement categories, however the medical cases remained reimbursed through fee-for-service, and surgical cases through a pre-defined flat-rate. The initiative had to solely utilize the reimbursement-performance link, but not hospital budgets, due to historical and political constraints. In the interest of space, we refer the reader to the following article where we have previously elaborated on some aspects:

<https://www.tandfonline.com/doi/full/10.1080/23288604.2016.1272979>; ref.25, Khalife et al. 2017.

- I see that your suggestions are all about recommending this reform measure. Do you have any relevant suggestions for the sustainable development of this reform measure? Is there a need for more supporting policies to further improve efficiency? Does this policy reform need further improvement?

We have added some text under the conclusion towards this. The commitment of the MoPH to maintain the reimbursement-performance linkage is necessary, including ongoing evaluations and development. This would likely involve annual or bi-annual generation of hospital performance results, based on performance measures that have been proven beneficial and meaningful for inclusion.

Competing Interests: No competing interests were disclosed.

Reviewer Report 08 June 2020

<https://doi.org/10.21956/wellcomeopenres.17338.r38761>

© 2020 Kahur K. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Kristiina Kahur 

Nordic Casemix Centre, Helsinki, Finland

Thank you for this very interesting research which is somehow unique on that field. Below are some comments and reflections which raised while reading the article:

- In conclusions it was said that “This increase was mainly attributed to decreased unnecessary hospitalizations and was accompanied by improved medical discharge coding practices”. I am wondering how was the improvement of coding assessed and is improved coding necessarily referring to increase of effectiveness?
- It was said in the article that the reimbursement was linked to a composite hospital total

performance score (TPS) which means that the higher the CMI the higher the TPS. CMI is explicitly related to coding quality. It is known from different researches that coding activity is often incentivized by payments, thus this can incentivize the hospitals to increase the coding activity and perhaps up-coding in some cases which probably happened also in Lebanon. How much this could have been a case?

- Related to previous comment - in the article it was mentioned that “providers do not have access to the weights used in casemix evaluation”. This might limit the gaming and up-coding indeed. But on other hand it also makes the system non-transparent and weights used for calculation the CMI will be as a black box within the system which might have negative impact on improvement of coding quality.
- Given that there is no DRG system in place in Lebanon it somehow limited the development of CMI by using DRG cost-weights which would have been a conventional way when DRG system would have been used. The chosen approach was compensating the missing DRG cost-weights. However, it did not use the information/variables which would potentially have had impact on weights and CMI calculation. E.g. variable like age was not available for this study (as indicated in Data sources chapter) which would have had impact on weights in some cases. Also information about co-morbidities and complications was not taken into account which might be cost driver and thus, may have had impact on weights and CMI. It was also interesting that the medical cases were divided based on LOS whereas the surgical ones were not. In conventional cost-weight calculation also the cost/LOS outliers are taken into account which was not systematically the case in this study. Having DRG system in place, the calculations of CMI would be perhaps more consistent and internationally comparable by taking into account all possible variables a casemix system is using and which were not taken into account in this study.
- Given that five-year cost averages was used for calculation of CMI, how (if at all) was the inflation and other factors influencing the cost throughout the years taken into account? In other words, was the weight of one particular ICD or procedure code the same throughout the five-year period?
- It was mentioned in the article that the weight was calculated for each ICD (and procedure code). In table 4 it remains unclear if the separate weights were calculated e.g. for diagnosis code J44.1 and J44. Or was the separate weight calculated for codes J20-J20.9 and J20. However, table 5 which refers to diagnosis codes with the greatest change in terms of CMI, shows different values e.g. for code P22 and P22.0. This is somehow confusing because in general, subcodes of three-character diagnosis code should be clinically and economically meaningful and should have one weight, i.e. codes P22 and P22.0 should be similar in terms of clinical severity and cost. In general (and this really depends on national coding guidelines), if a three-character diagnosis code has subcodes, then the subcode as more specific one should be used instead of three-character code.

Is the work clearly and accurately presented and does it cite the current literature?

Yes

Is the study design appropriate and is the work technically sound?

Yes

Are sufficient details of methods and analysis provided to allow replication by others?

Yes

If applicable, is the statistical analysis and its interpretation appropriate?

I cannot comment. A qualified statistician is required.

Are all the source data underlying the results available to ensure full reproducibility?

No source data required

Are the conclusions drawn adequately supported by the results?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Casemix systems' implementation, maintenance and development. Primary classifications. Coding quality. Contracting health care providers. Payment methods. Health financing reforms.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Author Response 06 Oct 2020

Jade Khalife, Faculty of Medicine at Lund University, Lund, Sweden

We are grateful for the valuable feedback received from the reviewer, which we have used to improve the manuscript. Kindly find below our responses on a point-by-point basis.

- In conclusions it was said that "This increase was mainly attributed to decreased unnecessary hospitalizations and was accompanied by improved medical discharge coding practices". I am wondering how was the improvement of coding assessed and is improved coding necessarily referring to increase of effectiveness?

Improved coding was assessed based on the changes in the coding of chemotherapy cases under the chemotherapy code (Z51.1), particularly breast cancer and leukemia, and the accompanied changes in the coding of neoplasms (C00-D49); this explained the greatest change on short-stay CMI. That these were indeed chemotherapy sessions, rather than hospitalization for investigation, was confirmed by their short length of stay (same/one-day) and their hospitalization cost.

Effectively, this meant that hospitals were correctly coding chemotherapy sessions. Furthermore, we also found a shift to full coding digits and more specific codes, among short and medium-stay cases, specifically among neoplasms, pneumonia, COPD, respiratory distress of newborn, and diarrhea and gastroenteritis.

Our findings regarding improved coding were used to explain increased casemix index among short-stay cases, specifically the level changes. These do not refer to increased effectiveness, but would reflect improved code validity for use in the system. Improved effectiveness in our investigation refers to decreased unnecessary hospitalizations. We have clarified this in the

'conclusions' section.

- It was said in the article that the reimbursement was linked to a composite hospital total performance score (TPS) which means that the higher the CMI the higher the TPS. CMI is explicitly related to coding quality. It is known from different researches that coding activity is often incentivized by payments, thus this can incentivize the hospitals to increase the coding activity and perhaps up-coding in some cases which probably happened also in Lebanon. How much this could have been a case?

Although up-coding is generally a recognized issue in the Lebanese healthcare setting. However, given that providers were unaware of the TPS prior to the intervention in August 2014, we would not expect that there would have been an increase in up-coding above what are background levels. This similarly applies to increased coding activity, but only within surgical and mixed cases, where multiple-coding is used. Nonetheless, we acknowledge this factor will likely have to be more closely considered in future evaluations, with providers having become more aware of the incentive potential of up-coding, and other potential gaps to game the system.

- Related to previous comment - in the article it was mentioned that "providers do not have access to the weights used in casemix evaluation". This might limit the gaming and up-coding indeed. But on other hand it also makes the system non-transparent and weights used for calculation the CMI will be as a black box within the system which might have negative impact on improvement of coding quality.

The lack of access to the weights used does limit transparency. However, based on past experience and under the existing conditions, this was considered by the Ministry of Public Health as an important feature to limit the abuse/miss-use of hospitalization among some providers. This has been possible to maintain thus far due to the joint commitment of the Syndicate of Private Hospitals towards the P4P initiative. Nevertheless, increasing transparency regarding weights and the TPS process would be important goals for future system development.

- Given that there is no DRG system in place in Lebanon it somehow limited the development of CMI by using DRG cost-weights which would have been a conventional way when DRG system would have been used. The chosen approach was compensating the missing DRG cost-weights. However, it did not use the information/variables which would potentially have had impact on weights and CMI calculation. E.g. variable like age was not available for this study (as indicated in Data sources chapter) which would have had impact on weights in some cases. Also information about co-morbidities and complications was not taken into account which might be cost driver and thus, may have had impact on weights and CMI. It was also interesting that the medical cases were divided based on LOS whereas the surgical ones were not. In conventional cost-weight calculation also the cost/LOS outliers are taken into account which was not systematically the case in this study. Having DRG system in place, the calculations of CMI would be perhaps more consistent and internationally comparable by taking into account all possible variables a casemix system is using and which were not taken into account in this study.

This is largely correct, although it is also pertinent to recognize the existing institutional context within which the CMI-based reimbursement system is developing. We have added a paragraph

under 'strengths and limitations' to address age and comorbidities. We acknowledge that inclusion of age as a variable for casemix index calculation would be more advantageous. Its inclusion has been discussed at length with and within the MoPH. The conclusions of these discussions were that, although some diagnoses and procedures would entail greater complexity or cost for older patients, there was insufficient information for selecting which of these to adjust for, and to what extent, within the Lebanese setting. This remains an important area for future investigations and development of this system. We note that in the 2018 update of the Total Performance Score, the proportion of elderly was captured as a factor, and thus had some influence on determining hospital reimbursement tiers; this was to incentivize hospitals away from avoiding elderly patients.

We also acknowledge the importance of capturing comorbidities within casemix index calculation. Unfortunately, the MoPH and hospitals do not routinely code for comorbidities; the field for these is 'optional', and as such, any entries were unsuitable for use in this study. The MoPH had agreed in principle that this factor should be included, however there were concerns regarding practical implementation and potential for miss-use. Therefore, this awaits a more comprehensive revision of the coding-reimbursement-auditing mechanisms.

Medical cases were divided by length of stay on the basis that they are reimbursed by fee-for-service, and length of stay among these cases has a direct relation to the cost submitted by hospitals for MoPH reimbursement. However, surgical cases are reimbursed on a flat-rate basis, regardless of length of stay. In this investigation we maintained the same categories used by the MoPH to calculate casemix index. However, we acknowledge that development of the system to account for length of stay also among surgical cases would allow a more accurate measurement of case complexity.

- Given that five-year cost averages was used for calculation of CMI, how (if at all) was the inflation and other factors influencing the cost throughout the years taken into account? In other words, was the weight of one particular ICD or procedure code the same throughout the five-year period?

For surgical cases (flat-rate), the MoPH updated the procedure costs in March 2013, which remained in use until the 2018 update. These updates increased the base-rate of all procedures, to account for inflation. We used weights based on the March 2013 update, also retrospectively up to 2011. For medical cases (fee-for-service), the MoPH undertook a hospital-bed update (base-rate), which is only one component of the bill charged by hospitals to the MoPH, but nevertheless represents an internal accounting for inflation. For medical cases, we used the 5-year average per ICD10 code, without further adjustment.

- It was mentioned in the article that the weight was calculated for each ICD (and procedure code). In table 4 it remains unclear if the separate weights were calculated e.g. for diagnosis code J44.1 and J44. Or was the separate weight calculated for codes J20-J20.9 and J20. However, table 5 which refers to diagnosis codes with the greatest change in terms of CMI, shows different values e.g. for code P22 and P22.0. This is somehow confusing because in general, subcodes of three-character diagnosis code should be clinically and economically meaningful and should have one weight, i.e. codes P22 and P22.0 should be similar in terms of clinical severity and cost. In general

(and this really depends on national coding guidelines), if a three-character diagnosis code has subcodes, then the subcode as more specific one should be used instead of three-character code.

Separate weights were calculated for each code, whether three or four-character. Therefore, J44.1 and J44 had different weights (differing by 14%), as did P22 and P22.0. The decision to have separate weights was based on the review of various codes, which in some cases revealed considerable differences between three and four-character codes. Acknowledging that this is likely due to a mix of factors, including imprecise coding practices and actual diagnosis, but lacking information to resolve these differences, we choose to use weights for each three and four-character code. This approach is similar to that used in the calculation of casemix index by the MoPH. The use of three-character codes is also due for phasing out by the MoPH, which will require four-character codes for all hospitalizations.

Competing Interests: No competing interests were disclosed.

Reviewer Response 21 Dec 2020

Kristiina Kahur, Nordic Casemix Centre, Helsinki, Finland

Thanks a lot for clarifications. Good luck!

Competing Interests: No competing interests were disclosed.

Paper 3



Impact of pay-for-performance on hospital readmissions in Lebanon: An ARIMA-based intervention analysis using routine data.

Jade Khalife ¹, Walid Ammar ², Fadi El-Jardali ^{3,4}, Maria Emmelin ¹, Björn Ekman ¹

¹ Social Medicine and Global Health, Department of Clinical Sciences, Malmö (IKVM), Faculty of Medicine, Lund University, Malmö, Sweden.

² Higher Institute of Public Health, Faculty of Medicine, Saint Joseph University of Beirut, Lebanon.

³ Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, Ontario, Canada.

⁴ Department of Health Management and Policy, Faculty of Health Sciences, American University of Beirut, Lebanon.

Corresponding author: Jade Khalife, Malmö (IKVM), Faculty of Medicine, Lund University, Malmö, Sweden.

Abstract

Objective: To estimate the impact of hospital pay-for-performance on readmissions for a set of common conditions.

Design: This retrospective cohort study included all hospitalizations under the coverage of the Ministry of Public Health (MoPH) in Lebanon between 2011 and 2019. We calculated 30-day all-cause readmissions following general, pneumonia, cholecystectomy and stroke cases. We used an interrupted time series design, including the use of AutoRegressive Integrated Moving Average (ARIMA) models.

Setting: This nationwide study was undertaken in Lebanon, using hospitalizations at about 140 private and public hospitals contracted by the MoPH.

Participants: Participants included citizens across all ages under the MoPH's coverage (52% of citizens).

Intervention: The MoPH engaged hospital leaders on the addition of a readmissions component to the ongoing pay-for-performance initiative. Participants included hospital directors and managers, and the leadership of the Syndicate of Private Hospitals.

Main outcome measures: Age-adjusted monthly all-cause readmission rates for each of general, pneumonia, cholecystectomy and stroke cases.

Results: Across 2011-2019, mean monthly readmission rates were 5.91% (S.D. 0.51%) for general readmissions, 4.81% (S.D. 0.90%) for pneumonia, 1.02% (S.D. 0.48%) for cholecystectomy, and 2.31% (S.D. 1.71%) for stroke. Using ARIMA models we found a decrease in mean monthly readmissions in the post-intervention period for cholecystectomy (24.8%; CI 5.1%-44.5%) and stroke (8.5%; CI 1.5%-15.5%). There was no evidence of intervention impact on pneumonia and general readmissions, both overall and among small, medium and large hospitals.

Conclusions: Including readmissions within pay-for-performance has the potential to improve hospital performance and patient outcomes, even in countries with more limited resources. Effects may vary across conditions, indicating the need for careful design and understanding of the particular context, both with respect to implementation and to evaluation of impact.

Key messages

- Reducing hospital readmissions is an important goal for health systems and has increasingly been linked to financial incentives.
- Country-level experiences with pay-for-performance are few, and it is unclear if this could be used to reduce hospital readmissions.
- We found that the addition of readmissions into the pay-for-performance initiative in Lebanon was followed by a decrease in readmissions following stroke and cholecystectomy, but not pneumonia and general cases.
- Including readmissions within pay-for-performance has the potential to improve hospital performance and patient outcomes in limited resource settings, but adjustments are necessary to avoid unintended consequences.

Introduction

The Lebanese Ministry of Public Health (MoPH) established a pay-for-performance (P4P) initiative in 2014, to incentivize hospitals to improve their performance across select components¹. The main components were on patient satisfaction, hospital casemix and accreditation. This determined the total hospital performance score (TPS), which was used to categorize about 140 hospitals across three reimbursement tiers. To incentivize hospitals to reduce readmissions, a new component on hospital readmission rates was announced in 2018, and implemented in the following year. This included readmissions for four conditions: general, pneumonia, stroke and cholecystectomy cases.

Hospital readmission has been commonly used in assessing the performance of hospitals and health systems, usually within a quality of care framework². They also have a direct impact on patients themselves. Hospital readmissions are frequent, costly and sometimes life-threatening³. Several factors may influence the risk of readmission, some of which are not modifiable.

These include factors at the patient, hospital and community levels. The most notable patient characteristics include age, medical condition or procedure, and comorbidities⁴⁻⁸. Hospital treatment quality and discharge instructions (and compliance) have a major role in readmissions, while the role of hospital size, volume and geography are less understood⁸⁻¹². There is also a major role of community-level factors, such as socioeconomic status and poverty^{4,6}.

Linking readmissions to financial incentives

Reducing readmissions has been considered an important health policy goal, with some countries setting national-level policies, including Denmark, England, Germany and the United States¹³. In England and the US, these have been linked to financial incentives, within a P4P framework.

The most prominent national-level experience has been the Hospital Readmission Reduction Program (HRRP) in the United States, enacted by the Patient Protection and Affordable Care Act of 2010¹⁴. HRRP's first iteration targeted readmissions following acute myocardial infarction, heart failure and pneumonia. Penalties were applied from 2012 onwards on hospitals with higher than average readmission rates, initially at 1% of hospital reimbursement costs, and subsequently reaching up to 3%. The program later expanded to include Chronic Obstructive Pulmonary Disease (COPD), coronary artery bypass graft surgery, and hip and knee replacement surgery. Evaluations of the first HRRP iteration were generally favorable, with decreased readmission rates across all three conditions, hundreds of millions of dollars in healthcare savings, and no increased mortality¹⁵⁻¹⁷. However, some of the early decreases were not sustained. Some studies suggested HRRP may have led to unintended consequences such as increased post-discharge mortality for heart failure and pneumonia^{14,18}. This has led to considerable debate among researchers and policymakers, as well as important lessons learned¹⁹.

In contrast to the HRRP, the Hip Fracture Best Practice Tariff (BPT) in England had a simpler P4P design, focusing on a few high-value

outcomes. Implementation of the BPT was followed with decreased mortality among elderly patients with hip fracture, as well as the reversal of a previously increasing trend of readmissions²⁰.

Limited resource countries' challenges in tackling readmissions

Despite increased literature in the past decade on incentivizing readmission reduction in the US, similar national-level initiatives are uncommon. Contextual differences may be expectedly be greater for countries with more limited resources to systematically collect data for tracking readmissions. This may include absence of computerized records at hospital level, weak medical coding capacity and absence of a harmonized database at regional or national levels.

Purpose of this paper

The purpose of this article is to estimate the impact of P4P on readmissions of general, pneumonia, stroke and cholecystectomy cases. In particular, we compare readmission rates before and after the announcement of the P4P component on readmissions. We also analyze heterogeneous changes across hospital sizes, and explain the rationale and algorithms used to determine readmissions.

Intervention

The MoPH engaged hospital leaders to inform them on the addition of a readmissions component to the ongoing pay-for-performance initiative. This was undertaken through two stakeholder engagement events, in January and March 2018, which centered on the policy and the technical levels respectively. Participants included hospital directors, quality managers, financial managers and the leadership of the Syndicate of Private Hospitals. Discussions involved the selection, calculation and weighing of performance components, including a new component on readmissions.

Methods

Overall study design

We use an interrupted time series (ITS) design with an intervention analysis. This was necessarily determined by the availability of multi-year data from the MoPH hospitalization database coupled with the inability to randomize the intervention (MoPH legal regulations). In such circumstances, a valid inference of the causal relationship between intervention and outcome requires a well-designed time series experiment^{21p.2}. We use a historical control, by categorizing pre- and post-intervention data-points within the same population.

Study setting

Lebanon is an Eastern Mediterranean country with a population of about 6.8 million people, including two million refugees. The MoPH contracts about 140 private and public hospitals to provide coverage for 52% of citizens, who otherwise lack formal health insurance. This has been further described elsewhere²².

Data collection and preparation

Administrative data can accurately capture all-cause readmissions²³. Data was extracted from the MoPH hospitalization database by the Information Technologies Department. This included all cases under MoPH coverage between January 2011 and December 2019. The extracted data was provided in an annually segregated manner by case type (medical, surgical or intensive-care). Each file contained anonymized patient identifiers, with fields including admission and discharge dates, hospitalization record number, case identifier, age, sex, hospital code, medical code (ICD10) and/or surgical procedure code (CPT), and procedure order.

Algorithms for readmission calculation

We developed case definitions for each of the readmission indicators, based on a review of the literature. Using these, algorithms were coded for each of the calculation of age-adjusted readmission rates across the four conditions, using Stata v.16 software. Monthly readmission rates at national level were calculated for 11 months per year, excluding December (see

supplement A). A total of 99 monthly data points were generated, including 80 pre-intervention data points and 19 post-intervention data points (after March 2018). To improve interpretability, linear interpolation was used to populate December values between 2011 and 2018.

For all conditions, a readmission was identified as a patient having been readmitted within 30 days from previous hospitalization discharge, regardless of readmission cause (i.e. any-cause), unless otherwise specified (see supplement A). General cases included all medical and surgical hospitalizations, unless otherwise specified. Patients with multiple readmissions per year had only the first readmission counted as such, to limit the effect of outliers due to patients with high comorbidities.

Age adjustment of readmission rates

We used direct adjustment for age on the crude readmission rates. The 2015 denominator population was used as the standard reference for readmission data across 2011-2019. Each month of 2015 was used as the standard population for the corresponding month of other years. This was undertaken since monthly readmission rates were desired, and we expect the proportion of ages to vary across different months according to seasonal disease patterns. We used six age groups: 0-5 years; 6-20 years; 21-40 years; 41-60 years; 61-80 years; and ≥ 81 years. In effect, the monthly age-adjusted rate was the weighted average of the age-specific (crude) rates. This allowed us to reduce the potential confounding effect of age.

Statistical approach and analysis

Our main analytical tool involved AutoRegressive Integrated Moving Average (ARIMA) models. The work of George Box and Gwilym Jenkins in the 1970s popularized ARIMA models, whose precursors included the

ARMA models of Peter Whittle (1951). They have had increasing use over the past few decades, with some minor development of the analytical approaches and increasing emphasis on model usefulness and interpretability ^{21p.20}. In recent years ARIMA models have had considerably increased use within the health sciences.

In this study, the total monthly data points generated qualified this as a medium-length time series. This length increases reliability for analysis of intervention impact and allows for a seasonal component. The risk of model misspecification, and of type I and type II errors is inversely proportional to time series length ^{21p.264}.

A total of 10 ARIMA models were developed: one for each of the four readmission conditions, and across three hospital sizes for each of general and pneumonia readmissions. Hospitals were categorized as small (<50 beds), medium (50-100 beds) or large (>100 beds). We used an iterative identify-estimate-diagnose process, evaluating several models before narrowing down on each final model ^{21p.19}. We elaborate on the analytical process steps in table 1. We opted not to use forecasted-to-observed difference, to avoid the errors associated with this approach ^{21p.167}. For comparison purposes, we estimated a single-group interrupted time series analysis with Newey ordinary least-squares (OLS) regression, including seasonality adjustment. Two-sided statistical significance was set at $p < 0.05$.

Ethical considerations

The research protocol approval was granted by the Institutional Review Board (IRB) at the American University of Beirut (ID: FHS.FE.21). The requirement for patient consent was waived by the IRB.

Table 1: Analytical process steps.

Step	Description
1	A priori, we expected the intervention to (potentially) have an immediate impact on the scale of weeks, more likely resulting in a level change in readmission rates, but possibly slope and/or pulse changes. We did not identify other policies or interventions that may have otherwise affected our outcome measure.
2	We generated descriptive statistics, and identified eight gaps in the data (expected) for each December throughout 2011-2019. This was due to MoPH IT data extraction limitations. We used interpolation to compute the readmission rates for these months, based on the average of the previous and subsequent month.
3	We plotted monthly data points to visualize the data, and to support the identification of trends and outliers. Subsequent steps 4 to 6 used the pre-intervention data-points only.
4	To determine stationarity of the series, we assessed if variance was changing over time (heteroscedasticity). Failure to transform a non-stationary series would be a threat to statistical conclusion validity, and renders unusable the subsequent steps and any conclusions ^{21p.61} . We first used the Breusch-Pagan/Cook-Weisberg test. This was followed by the Dickey-Fuller test for unit root (non-stationarity). If a stationarity was found, then a differencing order would be applied to the data and iterative diagnostics repeated till a stationary series is produced.
5	We tabulated and visualized autocorrelation and partial autocorrelation functions (ACF and PACF), to assess autocorrelation and stationarity, as well as inform the appropriate AR and/or MA terms of our final model. Since seasonality is expected, by the nature of the data, seasonal ARIMA models on the order of 12 months were pursued.
6	We used a statistical command (autoarima) as a starting point to suggest models, following by individual model evaluation, using information developed from the previous steps. We pursued the most parsimonious model, dropping non-significant parameters while also relying on the Bayesian information criterion (BIC) generated. This was to avoid having unnecessary parameters resulting in increased variance and an over-fit model. The most practically useful model was chosen
7	We created the intervention variables for level, ramp and pulse changes, as there was uncertainty regarding what impact type to expect. A priori, a level change was considered more likely. A pulse change was defined as a three-month change (1) which then reverted to baseline form (0). The model was estimated, with parsimony again pursued. A visualization of the data (from step 3) was also used to support interpretation.
8	We ran diagnostics on the model, to ensure that the model met the assumptions upon which our analytical approach was based, specifically by comparing the residuals to white noise, and then testing for normality and independence. Residual plots were made for autocorrelations, partial autocorrelations, and a scatter plot to ensure non-heteroscedasticity. For normality testing, we used kernel density estimate plot, followed by quantile and standardized normal probability plots (Q-Q plot and P-P plot). If a normal distribution is approximated, then the residuals are considered to be no different than white noise. We then used the Kolmogorov-Smirnov to compare the outcome variable to a normal distribution. Testing for independence first involved a scatter-plot of residuals by time variable, followed by the Ljung-Box test on the residuals. Once normality and independence criteria were met, we considered the model to be statistically adequate, and the model selection algorithm is concluded. If any criteria were not met, the iterative process of model selection was repeated.

Table 2: Descriptive statistics of 30-day readmissions for four conditions, 2011-2019.

		General	Pneumonia	Cholecystectomy	Stroke
Admissions, total		1,333,691	70,585	26,820	13,370
Readmissions, total		80,080	3,569	681	876
Monthly readmission rate	Mean	5.91%	4.81%	2.42%	6.48%
	S.D.	0.51%	0.90%	1.02%	2.31%
	Min.	4.65%	3.13%	0.48%	1.71%
	Max.	7.29%	7.65%	4.89%	14.21%

Results

The study population fulfilling the inclusion and exclusion criteria varied by readmission type (see table 2). General readmissions included about 1.33 million hospitalizations across 2011-2019, and 80,080 readmissions. Index hospitalizations were lowest among stroke cases. Mean monthly readmission rates ranged across conditions between 2.42% and 6.48%. The number of hospitals having different readmissions varied by year, but was generally similar across 2011-2019 (see supplement B). About twice as many hospitals had pneumonia and general readmissions than those that had cholecystectomy and stroke readmissions.

Seasonality was observable across readmissions, which further justified the a priori decision to use seasonal ARIMA models (see figure 1). Readmission rates data was stationary across all four conditions. This was unsurprising, since one of the criteria used in selecting these conditions for the P4P intervention was that they displayed a generally stable long-term trend, to allow a better evaluation of change. The final ARIMA models are shown in tables 3-5. A first-order seasonal difference was applied for each of pneumonia, cholecystectomy and stroke models,

as this provided the most practically useful model, in line with our analytical approach ^{21p.83}.

The first iterations of the intervention models included all three changes (level, ramp, pulse), which we subsequently narrowed down by parsimony and data visualization. The final intervention models included a level change alone. Independence and normality criteria were met by all final models, as assessed in the post-estimation diagnostic tests and plots (see tables 3-5 and supplement C).

Using the ARIMA intervention models, there was no evidence of intervention impact on general and pneumonia readmissions. However, a level change was found in both cholecystectomy and stroke readmissions following the intervention. Mean monthly cholecystectomy readmissions decreased by about 24.8% (5.1%-44.5%) and stroke readmissions decreased by about 8.5% (1.5%-15.5%). Tables 3-5 summarize the results of all final ARIMA models. Validation using Newey-OLS regression found a decreased level change for cholecystectomy readmissions by about 52.7% (6.3%-99.0%, $p=0.026$), but no change was found among stroke, pneumonia or general readmission.

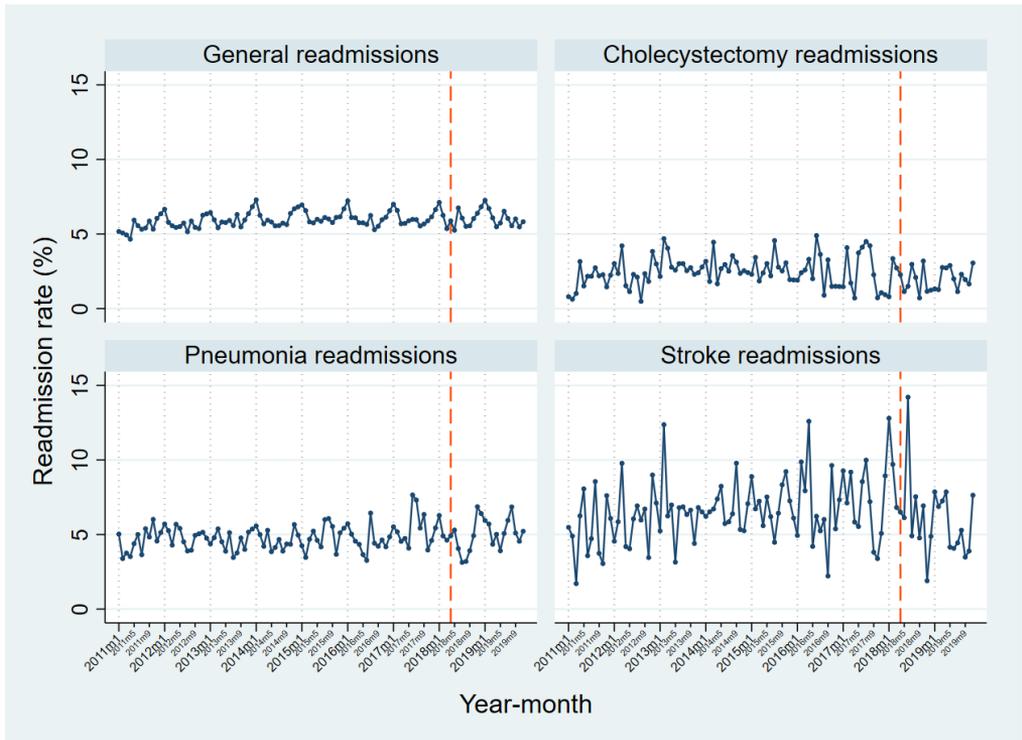


Figure 1: 30-day readmission rates for general readmissions, pneumonia, cholecystectomy and stroke, 2011-2019. A dashed line represents the anticipated impact time point, lagging after the hospital engagement events.

Table 3: Final ARIMA models and results across four readmission types, 2011-2019.

	General		Pneumonia		Cholecystectomy		Stroke	
Model	(1,0,0) (1,0,0) ₁₂		(1,0,0) (1,1,0) ₁₂		(1,0,0) (1,1,0) ₁₂		(0,0,1) (0,1,1) ₁₂	
BIC	79.1		215.9		252.2		364.4	
Level coefficient, p, (95%CI)	0.256	0.075	-	0.658	0.714	-	-	-
		(-0.026 to 0.537)	0.154	(-0.837 to 0.528)	0.048	(-1.420 to -0.008)	1.637	(-2.907 to -0.367)
Constant	5.825	<0.001	0.081	0.520	0.084	(-0.134 to 0.302)	0.274	(0.063 to 0.485)
	0.348	<0.001	0.270	0.037	0.071	(-0.184 to 0.326)	-	-
SAR	0.664	<0.001	0.597	<0.001	0.502	(-0.708 to -0.296)	-	-
		(0.510 to 0.817)	-	(-0.774 to -0.420)	-	-	-	-
MA	-	-	-	-	-	-	0.056	0.693
		-	-	-	-	-	-	(-0.331 to 0.220)
SMA	-	-	-	-	-	-	0.870	<0.001
		-	-	-	-	-	-	(-1.190 to -0.550)
Sigma	0.344	<0.001	0.973	<0.001	1.188	(1.018 to 1.358)	2.281	<0.001
		(0.299 to 0.388)	-	(0.866 to 1.080)	-	-	-	(1.924 to 2.638)
Log likelihood	-41.1	-	134.9	-	152.9	-	221.1	-
Kolmogorov-Smirnov test	-	0.480	0.899	-	-	-	-	-
		-	-	-	-	-	-	0.389
Ljung-Box test	-	0.806	0.739	-	-	-	-	0.900
		-	-	-	-	-	-	-

BIC: Bayesian information criterion, (S)AR: (seasonal) autoregressive term, (S)MA: (seasonal) moving average term.

Table 4: Final ARIMA models for general readmissions among small, medium and large hospitals, 2011-2019.

General readmissions	Small hospitals		Medium hospitals		Large hospitals	
	(1,0,0) (1,0,0) ₁₂		(1,0,0) (1,0,0) ₁₂		(1,0,0) (1,0,0) ₁₂	
Model						
BIC	114.2		115.7		235.7	
Level coefficient, p, (95% CI)	0.246	0.168	-0.044	0.801	0.289	0.331
		(-0.104 to 0.596)		(-0.384 to 0.297)		(-0.294 to 0.872)
Constant	5.951	<0.001	5.727	<0.001	5.442	<0.001
		(5.689 to 6.211)		(5.510 to 5.944)		(5.156 to 5.728)
AR	0.299	0.001	0.369	<0.001	0.228	0.020
		(0.128 to 0.470)		(0.201 to 0.538)		(0.036 to 0.419)
SAR	0.667	<0.001	0.209	0.089	0.157	0.179
		(0.523 to 0.811)		(-0.032 to 0.449)		(-0.072 to 0.386)
MA	-	-	-	-	-	-
SMA	-	-	-	-	-	-
Sigma	0.403	<0.001	0.473	<0.001	0.838	<0.001
		(0.340 to 0.465)		(0.411 to 0.535)		(0.717 to 0.958)
Log likelihood	-58.1	-	-72.1	-	-133.0	-
Kolmogorov-Smirnov test	-	0.430	0.993	-	-	0.366
Ljung-Box test	-	0.965	0.54	-	-	0.918

BIC: Bayesian information criterion, (S)AR: (seasonal) autoregressive term, (S)MA: (seasonal) moving average term.

Table 5: Final ARIMA models for pneumonia readmissions among small, medium and large hospitals, 2011-2019.

Pneumonia	Small hospitals		Medium hospitals		Large hospitals	
Model	(1,0,0) (1,0,0) ₁₂		(1,0,0) (1,1,0) ₁₂		(0,0,1) (0,1,1) ₁₂	
BIC	267.6		342.8		468.8	
Level coefficient, ρ , (95% CI)	0.145	0.670 (-0.521 to 0.811)	-0.411	0.547 (-1.751 to 0.928)	-0.326	0.872 (-4.299 to 3.647)
Constant	4.767	<0.001 (4.452 to 5.082)	0.107	0.563 (-0.255 to 0.468)	0.163	0.592 (-0.434 to 0.760)
AR	0.205	0.055 (-0.004 to 0.415)	0.085	0.411 (-0.118 to 0.288)	-	-
SAR	0.114	0.224 (-0.070 to 0.298)	-0.531	<0.001 (-0.722 to -0.338)	-	-
MA	-	-	-	-	0.177	0.031 (0.016 to 0.338)
SMA	-	-	-	-	-0.790	<0.001 (-1.018 to -0.561)
Sigma	1.036	<0.001 (0.873 to 1.199)	2.098	<0.001 (1.798 to 2.399)	1.188	<0.001 (1.018 to 1.358)
Log likelihood	-155.7	-	-207.2	-	-281.3	-
Kolmogorov-Smirnov test	-	0.910	0.755	-	-	0.029
Ljung-Box test	-	0.872	0.851	-	-	0.750

BIC: Bayesian information criterion, (S)AR: (seasonal) autoregressive term, (S)MA: (seasonal) moving average term.

Discussion

Our study suggests that the addition of readmissions to hospital pay-for-performance resulted in a decrease of stroke and cholecystectomy readmissions. We found no evidence of impact on pneumonia and general case readmissions, across all hospital sizes.

Several factors may have contributed to our findings. These include intervention design, context and characteristics of patients and hospitals. Intervention design has a major role in determining outcomes. A narrative review of the US Medicare programs structured this under four domains: program scope (broad, narrow), performance (relative, absolute), awards (achievement, improvement, both), and incentive (reward, penalty, both)²⁴. The Lebanese MoPH's P4P model had a mixed scope, using relative hospital performance, and rewarding achievement. Hospitals were incentivized to improve performance to achieve higher reimbursement tiers. This reimbursement-performance linkage has previously led to improved hospital performance in 2014, specifically by decreasing unnecessary hospitalization and improving coding practices¹. Readmissions were a new component added to the P4P model in 2018. Each of the four readmissions types would equally contribute to the 2% of the total performance score. The readmission component may have been too small to incentivize hospitals to decrease certain readmissions, since the major TPS components were on patient satisfaction (20%), hospital accreditation (30%) and casemix (45%).

The number of hospitals that had cholecystectomy and stroke readmissions was about half of those that had pneumonia and general readmissions. Improvement across a few hospitals may be more easily reflected in national-level outcomes of the first group. Hospitalization volume per condition may also have a similar effect, as general cases were orders of magnitude greater than other conditions. In our study, the decrease in cholecystectomy and stroke readmission was not reflected at the scale of general readmissions.

This may be due to a dilution effect, or mixed changes in non-targeted readmissions.

Readmission risk is also influenced by the medical complexity. Persons undergoing a relatively less complex surgery such as cholecystectomy are generally expected to be younger and healthier than those hospitalized with pneumonia or stroke. Patient characteristics such as comorbidities are not modifiable by hospitals. Therefore, the potential improvement space would vary across readmission types. Here we note the wider variation in readmission rates of stroke and cholecystectomy.

Broad outcome-based payments generally have more positive effects on quality and cost, compared to narrow models²⁵. Using a similar logic, the P4P's inclusion of general readmissions was intended to incentivize hospitals to address processes that could impact readmissions across a wide range of conditions. However, our study found no evidence of such change. This suggests that hospitals were unable and/or uninterested to more broadly address readmissions. Being broader than the three other readmission types, general readmissions are likely more challenging to address. Hospitals may have also considered the more specific readmissions as 'low-hanging fruits'.

We also found no evidence of change in pneumonia and general readmissions, across small, medium and large hospitals. Such segregated analysis was not undertaken for the lower-volume conditions, due to decreased reliability. Previous studies in the US have found a mixed or weak associations between hospital size and readmission rate^{10, 26, 27}. Some have suggested that increased bed capacity induces demand and makes readmissions easier, thus leading to more care, but not necessarily better care²⁶. Hospital volume is related to hospital size, and has generally stronger associations with readmission rates. Higher volume is typically related to better outcomes, although some studies have found the opposite relation²⁸. Among surgical conditions, higher volume has been associated with lower readmission rates, possible due to larger hospitals having more systematic approaches to determine when to discharge patients and

prevent readmissions¹⁰. In the Lebanese context, we expected hospital size to be more relevant than volume, because the MoPH also contracts with the largest hospitals, but for a small volume of typically more complex patients.

Strengths and limitations

Country-level experiences linking readmissions to financial incentives are few, and studies are predominantly reliant on the US experience. Our study contributes to this limited evidence base, and from a limited-resource country setting. We have developed readmission algorithms for four conditions, using routine administrative data spanning nine years and involving about 140 hospitals. Our analytical approach involved an established iterative stepwise process. Our use of stable medium-length time-series increases the validity of statistical conclusions reached. We were able to risk-adjust for age, which is the predominant risk factor for readmissions, but not for comorbidities, due to the lack of such data. The role of the latter was minimized by including only one readmission per patient per year.

We cannot determine the extent to which lower readmissions genuinely reflect better quality of care. While we consider it probable that improved care resulted in decreased readmissions, other contributing factors are plausible. Some hospitals may have intentionally reduced access to hospitalization, either by treating in emergency rooms (ER) or otherwise refusing to re-hospitalize. The MoPH does not offer ER coverage for its beneficiaries and lacks the data to verify such practice. Although we are unable to estimate this factor's contribution, we do not expect this to have had a major role, as decreases in readmissions were limited to two of the four conditions. There have been mixed findings of unintended consequences of the US HRRP experience. Earlier reports failed to find an association of hospital visits with readmission rates, while subsequent ones found decreased readmissions were accompanied by increased emergency and observation room stays^{16, 29}. For example, up to 80% of the decrease in heart failure readmissions may have been accounted for by increased emergency and observation

room stays¹⁹. It is difficult to establish causal relations in such complex contexts, due to study designs and inadequate statistical approaches. One solution would be to include the entire spectrum of hospital visits within a single outcome measure¹⁹.

For outcome-based P4P initiatives, it is important to not only estimate impact, but also to explain why it occurred (or not)^{30, 31}. This would involve investigation of the causal pathways involved, particularly in complex interventions³². As is typical of hospital P4P, for the Lebanese MoPH initiative the hospital functions as a 'black box', into which we have little insight regarding the motivations, decisions and actions of managers and health professionals. Understanding the factors involved in mediating impact may allow us to better design P4P interventions, and also explain the differential impact observed across the readmission types included³².

We are unable to assess any changes in post-discharge mortality, which may be another consequence of hospitals avoiding readmissions. Some studies of the US HRRP found an increase in post-discharge mortality among heart failure and pneumonia patients, but not those with acute myocardial infarction¹⁸. Other studies found increased mortality only among those hospitalized for heart failure¹⁴. An overly strong incentive for reducing readmissions has been noted, with the penalty for hospitals having excess readmissions being 15 times greater than those having excess mortality¹⁴. Increasing mortality for heart failure remains not fully understood, but this has led to recommendations that readmission and mortality be combined¹⁹.

Competing risk models have been recently proposed as a solution to some unintended consequences. This is defined as an event whose occurrence impedes the occurrence of the primary outcome of interest³³. For example, a readmissions competing risk model would keep any deaths that occur in the calculation denominator, rather than remove them as is standard in current metrics. For high-risk groups or conditions, such models would have greater relevance, given the more frequent mortality.

Policy implications

Based on our findings and reflections using the wider literature, several recommendations can be offered for the P4P initiative in Lebanon. ER visits should be captured in the MoPH hospitalization data, and readmissions and mortality data should be linked. Readmission outcomes should be widened to include ER visits, and either combined to capture mortality, or within a competing risk model. The MoPH may consider alternating different readmission types within P4P, while excluding high-risk conditions or those more prevalent in more vulnerable groups (e.g. elderly). The focus on readmissions may be exchanged with processes directly related to them, such as hospital-to-community patient transitions.

Acknowledgements

We are grateful to Abeer Al Halabi and Elise Barakat, who as project research assistants supported the validation of the outputs from the readmission algorithms. We are grateful to following individuals at the Lebanese MoPH: Rita Freiha, Hilda Harb and Jihad Makouk (feedback and insight); and Rabiaa Rachid and Jenny Romanos (data extraction from hospitalization database); and to Anne Mills (London School of Hygiene and Tropical Medicine, project senior advisor).

References

1. Khalife J, Ammar W, Emmelin M, El-Jardali F, Ekman B. Hospital performance and payment: impact of integrating pay-for-performance on healthcare effectiveness in Lebanon. *Wellcome Open Research*. 2020;5(95).
2. Donabedian A. The quality of care. How can it be assessed? *JAMA*. 1988;260(12):1743-8.

Conclusion

Engaging hospitals on pay-for-performance was followed by a decrease in stroke and cholecystectomy readmissions. We did not find a change in pneumonia and general readmissions, regardless of hospital size. Including readmissions within P4P has the potential to improve patient outcomes and system efficiency. Encompassing all hospital visit types and capturing patient mortality would help mitigate unintended consequences. It is conceivable that the next generation of P4P may use a competing risk design, and exchange complex outcomes with processes known to favorably influence them.

Competing interests

The authors have no competing interests to declare.

Funding

This work was supported by the Wellcome Trust [204020] and a project grant [MR/N015916/1] from the Joint Health Systems Research Initiative (DFID/MRC/Wellcome Trust/ESRC). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

3. Jencks SF, Williams MV, Coleman EA. Rehospitalizations among patients in the Medicare fee-for-service program. *N Engl J Med*. 2009;360(14):1418-28.
4. Gohil SK, Datta R, Cao C, Phelan MJ, Nguyen V, Rowther AA, et al. Impact of Hospital Population Case-Mix, Including Poverty, on Hospital All-Cause and Infection-Related 30-Day Readmission Rates. *Clin Infect Dis*. 2015;61(8):1235-43.

5. Barnett ML, Hsu J, McWilliams JM. Patient Characteristics and Differences in Hospital Readmission Rates. *JAMA Intern Med.* 2015;175(11):1803-12.
6. Calvillo-King L, Arnold D, Eubank KJ, Lo M, Yunyongying P, Stieglitz H, et al. Impact of social factors on risk of readmission or mortality in pneumonia and heart failure: systematic review. *J Gen Intern Med.* 2013;28(2):269-82.
7. Friebel R, Hauck K, Aylin P, Steventon A. National trends in emergency readmission rates: a longitudinal analysis of administrative data for England between 2006 and 2016. *BMJ Open.* 2018;8(3):e020325-e.
8. Roshanghalb A, Mazzali C, Lettieri E, Paganoni AM, Bettle A. Stability over time of the "hospital effect" on 30-day unplanned readmissions: Evidence from administrative data. *Health Policy.* 2021;125(10):1393-7.
9. Horwitz LI, Partovian C, Lin Z, Grady JN, Herrin J, Conover M, et al. Development and use of an administrative claims measure for profiling hospital-wide performance on 30-day unplanned readmission. *Annals of internal medicine.* 2014;161(10 Suppl):S66-S75.
10. Tsai TC, Orav EJ, Joynt KE. Disparities in surgical 30-day readmission rates for Medicare beneficiaries by race and site of care. *Ann Surg.* 2014;259(6):1086-90.
11. Fekri O, Manukyan E, Klazinga N. Associations between hospital deaths (HSMR), readmission and length of stay (LOS): a longitudinal assessment of performance results and facility characteristics of teaching and large-sized hospitals in Canada between 2013-2014 and 2017-2018. *BMJ Open.* 2021;11(2):e041648.
12. Lindenauer PK, Bernheim SM, Grady JN, Lin Z, Wang Y, Wang Y, et al. The performance of US hospitals as reflected in risk-standardized 30-day mortality and readmission rates for medicare beneficiaries with pneumonia. *J Hosp Med.* 2010;5(6):E12-8.
13. Kristensen SR, Bech M, Quentin W. A roadmap for comparing readmission policies with application to Denmark, England, Germany and the United States. *Health Policy.* 2015;119(3):264-73.
14. Gupta A, Fonarow GC. The Hospital Readmissions Reduction Program-learning from failure of a healthcare policy. *Eur J Heart Fail.* 2018;20(8):1169-74.
15. Desai NR, Ross JS, Kwon JY, Herrin J, Dharmarajan K, Bernheim SM, et al. Association Between Hospital Penalty Status Under the Hospital Readmission Reduction Program and Readmission Rates for Target and Nontarget Conditions. *JAMA.* 2016;316(24):2647-56.
16. Zuckerman RB, Sheingold SH, Orav EJ, Ruhter J, Epstein AM. Readmissions, Observation, and the Hospital Readmissions Reduction Program. *New England Journal of Medicine.* 2016;374(16):1543-51.
17. Dharmarajan K, Wang Y, Lin Z, Normand ST, Ross JS, Horwitz LI, et al. Association of Changing Hospital Readmission Rates With Mortality Rates After Hospital Discharge. *Jama.* 2017;318(3):270-8.
18. Wadhwa RK, Joynt Maddox KE, Wasfy JH, Haneuse S, Shen C, Yeh RW. Association of the Hospital Readmissions Reduction Program With Mortality Among Medicare Beneficiaries Hospitalized for Heart Failure, Acute Myocardial Infarction, and Pneumonia. *Jama.* 2018;320(24):2542-52.
19. Psofka MA, Fonarow GC, Allen LA, Joynt Maddox KE, Fiuzat M, Heidenreich P, et al. The Hospital Readmissions Reduction Program: Nationwide Perspectives and Recommendations: A JACC: Heart Failure Position Paper. *JACC Heart Fail.* 2020;8(1):1-11.
20. Metcalfe D, Zogg CK, Judge A, Perry DC, Gabbe B, Willett K, et al. Pay for performance and hip fracture outcomes: an interrupted time series and difference-in-differences analysis in England and Scotland. *Bone Joint J.* 2019;101-b(8):1015-23.
21. McCleary R, McDowall D, Bartos B. Design and analysis of time series experiments. New York, NY: Oxford University Press; 2017.

22. Khalife J, Rafeh N, Makouk J, El-Jardali F, Ekman B, Kronfol N, et al. Hospital Contracting Reforms: The Lebanese Ministry of Public Health Experience. *Health Systems & Reform*. 2017;3(1):34-41.
23. Sellers MM, Merkow RP, Halverson A, Hinami K, Kelz RR, Bentrem DJ, et al. Validation of new readmission data in the American College of Surgeons National Surgical Quality Improvement Program. *J Am Coll Surg*. 2013;216(3):420-7.
24. Joynt Maddox KE, Sen AP, Samson LW, Zuckerman RB, DeLew N, Epstein AM. Elements of Program Design in Medicare's Value-based and Alternative Payment Models: a Narrative Review. *J Gen Intern Med*. 2017;32(11):1249-54.
25. Vlaanderen FP, Tanke MA, Bloem BR, Faber MJ, Eijkenaar F, Schut FT, et al. Design and effects of outcome-based payment models in healthcare: a systematic review. *Eur J Health Econ*. 2019;20(2):217-32.
26. Fisher ES, Wennberg JE, Stukel TA, Skinner JS, Sharp SM, Freeman JL, et al. Associations among hospital capacity, utilization, and mortality of US Medicare beneficiaries, controlling for sociodemographic factors. *Health Serv Res*. 2000;34(6):1351-62.
27. Horwitz LI, Bernheim SM, Ross JS, Herrin J, Grady JN, Krumholz HM, et al. Hospital Characteristics Associated With Risk-standardized Readmission Rates. *Med Care*. 2017;55(5):528-34.
28. Horwitz LI, Lin Z, Herrin J, Bernheim S, Drye EE, Krumholz HM, et al. Association of hospital volume with readmission rates: a retrospective cross-sectional study. *BMJ*. 2015;350:h447.
29. Wadhwa RK, Joynt Maddox KE, Kazi DS, Shen C, Yeh RW. Hospital revisits within 30 days after discharge for medical conditions targeted by the Hospital Readmissions Reduction Program in the United States: national retrospective analysis. *BMJ*. 2019;366:l4563.
30. Renmans D, Holvoet N, Orach CG, Criel B. Opening the 'black box' of performance-based financing in low- and lower middle-income countries: a review of the literature. *Health Policy and Planning*. 2016;31(9):1297-309.
31. de Walque D, Robyn PJ, Saidou H, Sorgho G, Steenland M. Looking into the performance-based financing black box: evidence from an impact evaluation in the health sector in Cameroon. *Health Policy and Planning*. 2021;36(6):835-47.
32. Anselmi L, Binyaruka P, Borghi J. Understanding causal pathways within health systems policy evaluation through mediation analysis: an application to payment for performance (P4P) in Tanzania. *Implementation Science*. 2017;12(1):10.
33. Marang-van de Mheen PJ, Putter H, Bastiaannet E, Bottle A. Competing risks in quality and safety research: a framework to guide choice of analysis and improve reporting. *BMJ Quality & Safety*. 2021;30(12):1031-7.

Supplement A

Data development details

The case definitions for each of the readmission measures specified the inclusion criteria (diagnosis and procedure codes), and exclusion criteria (cardiac catheterization, lithotripsy, renal dialysis, chemotherapy, radiotherapy, malignancy, obstetric cases, motor vehicle accidents, blood transfusion, palliative care and transfer/same-day cases).

Since the data was provided with an annual timeframe by the MoPH, admissions in December had been artificially limited to an end-of-year discharge date. The calculation of monthly readmissions was therefore made for 11 months per year (January to November), including December cases only for identifying readmissions, and not for index hospitalization (i.e. for the numerator, not the denominator). This allowed the time opportunity for November cases to be readmitted in December.

ARIMA models description

ARIMA models include three components. An AutoRegressive (AR) component that indicates that an outcome variable is dependent on the previous values of this same term, and a stochastic term (randomness). A Moving Average (MA) component represents the influence of past error terms on the current error term; this is not to be confused with moving average of an outcome variable). In essence, an AR process remembers past realizations, while an MA process remembers past shocks. The Integrated (I) component indicates a transformation (integration) function which may be used to normalize the time series. The combination of these three components forms an ARIMA model.

All models require that the data be stationary. Where normality is violated (i.e. non-stationary), a normalization transformation of the data is necessary, to meet the underlying model assumptions. This pre-requisite is essential, because the analytical mechanism requires that a time series process operate identically in the future as it has in the past.

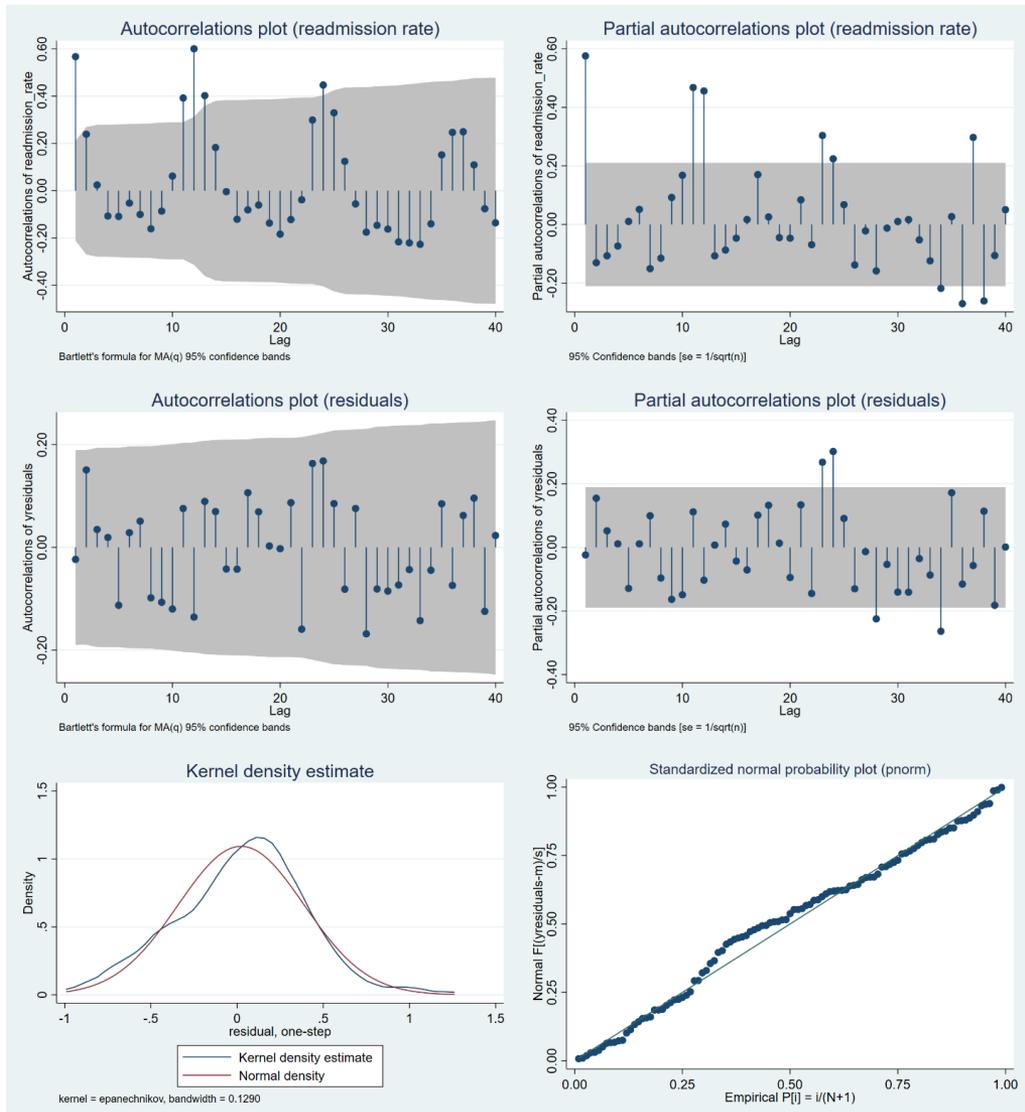
Supplement B

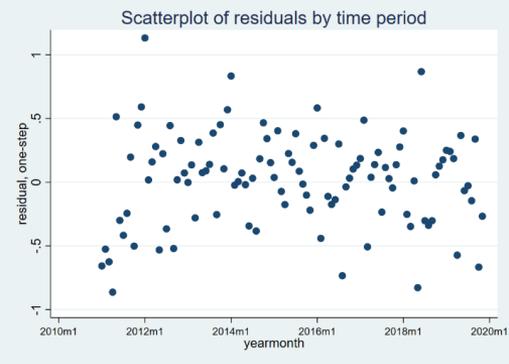
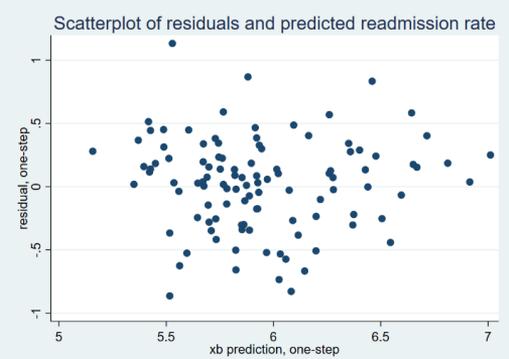
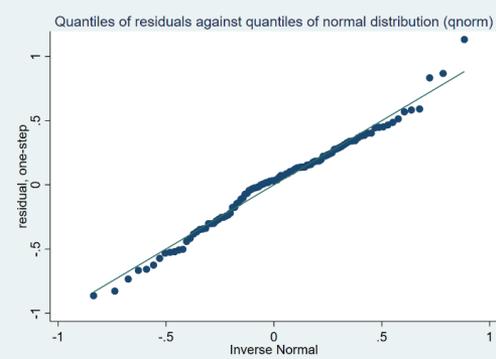
Table 6: Hospitals with readmissions, by hospital size, 2011-2019.

Condition	Size	2011	2012	2013	2014	2015	2016	2017	2018	2019
General cases	<50 beds	40	40	44	48	48	51	51	55	53
	50-100 beds	54	55	53	54	55	54	56	56	55
	101-200 beds	28	28	28	28	28	28	28	28	28
	>200 beds	8	8	8	8	8	8	8	8	8
	Total	130	131	133	138	139	141	143	147	144
Pneumonia	<50 beds	19	26	24	27	28	30	28	27	30
	50-100 beds	31	33	36	37	36	37	33	38	40
	101-200 beds	20	19	21	22	19	20	21	22	21
	>200 beds	5	5	5	6	6	6	4	6	5
	Total	75	83	86	92	89	93	86	93	96
Cholecystectomy	<50 beds	11	8	14	13	12	16	10	8	13
	50-100 beds	17	23	25	18	21	15	24	14	18
	101-200 beds	9	12	13	15	12	12	13	11	7
	>200 beds	4	2	1	4	3	2	2	2	2
	Total	41	45	53	50	48	45	49	35	40
Stroke	<50 beds	5	6	11	10	9	7	8	19	8
	50-100 beds	22	22	25	20	20	25	25	22	22
	101-200 beds	13	16	13	12	12	11	9	11	12
	>200 beds	4	1	3	3	4	3	1	2	4
	Total	44	45	52	45	45	46	43	54	46

Supplement C

Diagnostic plots used in the analytical process are included below, using general readmissions data. This includes autocorrelation and partial autocorrelation plots of dependent variable and of residuals; kernel density plot; P-P plot and Q-Q plot; scatterplot of residuals and time period; and scatterplot of residuals and model dependent variable.





Paper 4



RESEARCH ARTICLE

Exploring patient perspectives: A qualitative inquiry into healthcare perceptions, experiences and satisfaction in Lebanon

Jade Khalife^{1*}, Björn Ekman¹, Walid Ammar², Fadi El-Jardali^{3,4}, Abeer Al Halabi⁴, Elise Barakat⁴, Maria Emmelin¹

1 Social Medicine and Global Health, Department of Clinical Sciences, Faculty of Medicine, Lund University, Malmö, Sweden, **2** Higher Institute of Public Health, Faculty of Medicine, Saint Joseph University of Beirut, Beirut, Lebanon, **3** Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, Ontario, Canada, **4** Department of Health Management and Policy, Faculty of Health Sciences, American University of Beirut, Beirut, Lebanon

* jade.khalife@med.lu.se



Abstract

OPEN ACCESS

Citation: Khalife J, Ekman B, Ammar W, El-Jardali F, Al Halabi A, Barakat E, et al. (2023) Exploring patient perspectives: A qualitative inquiry into healthcare perceptions, experiences and satisfaction in Lebanon. *PLoS ONE* 18(8): e0280665. <https://doi.org/10.1371/journal.pone.0280665>

Editor: Mabel Aoun, Faculty of Medicine, Saint-Joseph University, LEBANON

Received: January 4, 2023

Accepted: August 5, 2023

Published: August 17, 2023

Copyright: © 2023 Khalife et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: Data cannot be shared publicly, to ensure participant confidentiality. Data are available from the Institutional Review Board at the American University of Beirut (contact via irb@aub.edu.lb) for researchers who meet the criteria for access to confidential data.

Funding: This work was supported by the Wellcome Trust [204020] and a project grant [MR/N015916/1] from the Joint Health Systems

Background

Patient perspectives have received increasing importance within health systems over the past four decades. Measures of patient experience and satisfaction are commonly used. However, these measures do not capture all the information that is available through engaging with patients. An improved understanding of the various types of patient perspectives and the distinctions between them is needed. The lack of such knowledge limits the usefulness of including patient perspectives as components within pay-for-performance initiatives. This study aimed to explore patient perspectives on hospital care in Lebanon. It also aimed to contribute insights that may improve the national pay-for-performance initiative and to the knowledge on engaging patients towards person-centered health systems.

Methods

We conducted a qualitative study using focus group discussions with persons recently discharged after hospitalization under the coverage of the Lebanese Ministry of Public Health. This study was implemented in 2017 and involved 42 participants across eight focus groups. Qualitative content analysis was used to analyze the information provided by participants.

Results

Five overall themes supported by 17 categories were identified, capturing the meaning of the participants' perspectives: health is everything; being turned into second class citizens; money and personal connections make all the difference; wanting to be treated with dignity and respect; and tolerating letdown, for the sake of right treatment. The most frequently prioritized statement in a ranking exercise regarding patient satisfaction was regular contact with the patient's doctor.

Research Initiative (DFID/MRC/Wellcome Trust/ESRC); JK, WA, FJ, BE, ME. <https://wellcome.org/>. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Conclusions

Patient perspectives include more than what is traditionally incorporated in measures of patient satisfaction and experience. Patient valuing of health and their perceptions on each of the health system, and access and quality of care should also be taken into account. Hospital pay-for-performance initiatives can be made more responsive through a broader consideration of these perspectives. More broadly, health systems would benefit from wider engagement of patients. We propose a framework relating patient perspectives to value-based healthcare and health system performance.

Introduction

Understanding how people perceive healthcare is important for improving people's own well-being, as well as health system quality and effectiveness. Perceptions may involve a wide spectrum of interactions with healthcare, such as pre-hospitalization preparations to hospital discharge. People's perceptions may be influenced by prior direct or indirect healthcare interactions, their personal perceptions of health, their expectations, as well as the overall health system and social context.

The perspective of patients within health systems has increasingly been recognized over the past four decades. This is exemplified by several landmark reports, though the translation into action has been gradual. Griffith's Report in 1983 urged the inclusion of public opinion and perception of healthcare for the UK National Health Service reforms [1]. In 2001, the US Institute of Medicine (IOM) highlighted patient-centeredness as one of the six dimensions of quality of care. In 2018, the IOM also called for an expansion towards person-centeredness, whereby the care provided is "respectful of and responsive to individual preferences, needs and values" [2, 3].

At the 69th World Health Assembly (2016) member states adopted the Framework on Integrated People-centered Health Services, whose vision emphasized the patients' role in defining their needs and co-producing health services [4]. A 2018 joint report by the World Health Organization, World Bank and the Organisation for Economic Co-operation and Development (OECD) called patient-centeredness the core of quality, and called for the engagement of people and communities in service design, delivery and assessment [5]. Indeed, patient-centeredness has also been called "the doorway to all qualities", not merely one dimension among others [6]. Moreover, the patient perspective has also been included as the personal value pillar in a re-definition of value-based healthcare [7]. Altogether, these developments represent a modern-day return to a fundamental aspect of early Hippocratic medicine: accompanying the patient and meeting their individual goals.

Understanding the patient's perspective is an essential precursor to patient-centeredness. Improved understanding helps in ensuring that they can be "full partners in the service delivery design and governance and in improving their own health" [8]. Patient perspectives have been most commonly addressed through measures of patient experience and satisfaction. Patient experience and satisfaction are both recognized as multi-dimensional, but are ambiguous and under-theorized [9]. They also do not capture all the information available from engaging patients, such as the value patients attach to health. Therefore, despite their wide application patient perspectives are not well understood.

One application of patient perspectives has been as a component within pay-for-performance (P4P) initiatives, which have become widespread in healthcare over the past two decades. The earliest national initiatives were launched in the United Kingdom (UK) and the

United States (US). In the UK, the Quality and Outcomes Framework incentivized the measurement of patient satisfaction by practitioners [10]. In the US, a national survey by the Centers for Medicare and Medicaid Services is used to capture the patients' perspective on care received, and since 2012 the results are linked to incentive payments within the Hospital Value-Based Purchasing program [11].

The largest sub-national P4P initiative preceding those in the UK and US was that in the state of São Paulo, Brazil, which created Social Organizations in Health (OSS) in 1998 to manage hospitals [12]. The OSS model included a patient satisfaction component (complaints and completion of satisfaction surveys) for the performance component used to set global hospital budgets, alongside volume targets.

There is a need for greater clarity on the different types of patient perspectives available, as well as on the distinction between them. The lack of such knowledge limits the usefulness of patient perspectives in pay-for-performance initiatives, potentially leading to unintended consequences [9]. More broadly, it also impedes the movement towards person-centered health systems.

In Lebanon, patient satisfaction and experience measures have been included in the hospital P4P initiative since 2014, which is used for setting hospital reimbursement tiers by the Ministry of Public Health (MoPH) [13]. In 2019, these measures comprised a fifth of the total performance score set by the MoPH. The inclusion and weight given to this was intended to represent the ministry's commitment to capturing patients' perspectives, and incentivize hospitals to improve their performance. However, considering the aforementioned limitations of patient experience and satisfaction, it is unknown to what extent this represents patients' true perspectives. Studying how patients experience and perceive hospital care in Lebanon would contribute to the overall knowledge on engaging patients towards person-centered health systems, and help improve P4P design and identification of impacts.

The aim of this study was to explore how people with experience of being hospitalized perceive the healthcare, focusing on health perceptions, access to care, experiences of hospitalization and satisfaction of care. It also aimed to contribute insights that may improve the design and effectiveness of the national pay-for-performance initiative, and to the broader understanding of engaging patients towards person-centered health systems.

Methods

Overall study design

A qualitative approach was used, since it allows us to gather information directly from the perspective of patients, on a topic that is not well understood [14]. It is based on constructivist view on science, where knowledge is developed in the interaction between the researcher and the informants and where inductive reasoning is central for developing the analysis. Qualitative methods therefore allow a deeper understanding of patients' perspectives, and would not have the limitations associated with structured questions that are characteristic of quantitative surveys [15].

FGDs were chosen, because we were interested in a wide range of views and experiences, and to encourage discussion and explanation of issues. It limits the influence of an interviewer on a respondent's comments, and allows for more spontaneous issues to arise [16]. FGDs rely on the interaction among group participants to encourage information generation on perceptions, attitudes and beliefs, while also allowing a facilitator to probe further when required [16]. This study is reported in accordance with the Standards for Reporting Qualitative Research (SRQR) [17].

Study setting

This study was undertaken in Lebanon, an Eastern Mediterranean country with a population of about 6.8 million people, including about 2 million refugees (most of whom are due to the conflict in neighboring Syria). About 52% of Lebanese citizens lack formal health insurance and are therefore under the coverage of the MoPH for hospitalization services. The MoPH contracts with public and private hospitals throughout the country to provide hospitalization services for its beneficiaries. Hospitals are reimbursed by the MoPH for 85% of the hospitalization bill, and 15% remaining as patient co-payment.

Study population and sampling

We used a sample of adult persons who had been hospitalized (and discharged) under MoPH coverage of the Ministry of Public Health (MoPH) during the preceding 3 months. Persons covered by the MoPH are typically comprised of the poorest or most disadvantaged stratum, as opposed to those having health coverage from other payers (e.g. employees, students, police, army) [13]. The purposive sampling aimed for a maximum variation of experiences in this target group by including both men and women from different age groups and from different geographic areas. To reach the target group we used a national database which includes persons from all regions and hospitals (public and private) which had been hospitalized under the MoPH coverage (comprising about 230,000 hospitalizations per year. Since we did not have specific knowledge about participants, we considered that a random sampling would allow achieving the purposive maximum variation sought for. A computer command was run on the hospitalization database SQL server to develop a list of 500 potential participants, using simple random sampling. Using this list, potential participants were contacted in their sequential order via telephone until we had a broad variation of men and women across a range of ages, residing in different types of regions, to participate in the FGDs.

The final sample involved 42 participants (22 men, 20 women) with a median age of 49 years (range: 25–65 years; see Table 1). All participants belonged to the least privileged half of the citizen population, having lacked formal health insurance. About 21% of persons contacted by telephone accepted the invitation and participated in the discussions. The main reason given for declining participation was due to persons being on vacation, or having other engagements. All participants had had several instances of interaction with healthcare (including at least one hospitalization, by design).

Discussion guide and pile-sorting statements

A discussion guide was developed using open-ended questions, to support the facilitation of the discussions. Based on our reading of the literature and discussion within our research

Table 1. Focus group discussion characteristics.

FGD #	Duration (min)	Participants, n	Sex
1	65	7	Men
2	37	3	Women
3	64	4	Women
4	57	4	Men
5	58	4	Men
6	59	5	Women
7	82	8	Women
8	79	7	Men

<https://doi.org/10.1371/journal.pone.0280665.t001>

team, we developed the following themes for the key questions: the meaning of health; description of local healthcare; characteristics of services received; description of 'good' and of 'bad' hospital stays; information needed upon hospital admission; and factors that would make a person revisit a hospital. A pilot FGD was held in March 2017, to test the discussion guide. The pilot included eight men as participants. The discussion guide was subsequently revised for greater clarity of questions and terms, based on the pilot results. Pilot participants were selected using the same process as that subsequently used for the eight FGDs, but limited to the Beirut region.

To increase our understanding of how patients prioritized different factors regarding their hospitalization, we also prepared a structured pile-sorting exercise. The pile sort is a method that has been used since the 1960s across different fields to investigate how people classify items, and may be considered a combination of quantitative and qualitative techniques [18–20]. One example involves asking participants to sort statements about a phenomenon of interest in piles of more versus less important [21]. In this study, the statements were developed based on our assumptions of what factors may affect the satisfaction and experience of patients, and also using existing literature and survey tools, including patient satisfaction dimensions and items of the US Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS) [22, 23]. The pile sorting exercise was held after the conclusion of the FGDs. Participants were asked to sort 16 statements into two piles (more versus less important). These included topics such as the importance of regular contact with personnel, hospital organization, cleanliness, communication, pain and privacy. One participant was supported by a research assistant who privately read the statements to them, due to being illiterate.

Data collection

The FGDs were held between July and September 2017. The groups were arranged separately for men and women since we assumed that this would create a more open discussion climate. Variability in group size was largely due to some cancellations and rescheduling of participants in the period preceding the discussions. They were all held in a private room at the MoPH headquarters in Beirut. Following an introduction to the study, all participants were asked to provide oral consent and agreed to be audio-recorded using a digital recording device.

The first author (JK) was the discussion facilitator, while two research assistants (AH, EB) noted the group layout and assessed interactions but they did not participate in the discussions. The first author (JK) introduced himself to participants as being involved in the hospital pay-for-performance initiative, which was a collaboration between the MoPH, the American University of Beirut and Lund University. The discussions involved exploring the six thematic areas, with probing questions used to clarify statements or explore them in more depth. The median duration of discussions was 62 minutes (range: 37–82 minutes; see [Table 1](#)). Notes were taken both during and after each FGD. These notes were continuously discussed within the data collection team to inform the next FGD and the forthcoming analysis.

No material or financial compensation was given to participants, but all were offered complimentary transport by taxi from their residence to the discussion site and back, and snack refreshments after the FGDs had ended.

Data analysis

The recordings were transcribed and translated from Lebanese Arabic to English verbatim (AH, EB). During the transcription process participant names and other personal identifiers (e.g. residence) were removed. Accuracy and sense-making of the transcription was iteratively assessed by the two research assistants and the facilitator.

Table 2. An example from the analysis process, moving from text, to code and category.

Text	Code	Category
You have nothing even if you don't have health, even if you own the whole world. (FGD1-P5)	Without health we have nothing	More important than money or wealth
Briefly, health is the whole life, who does not have good health, has nothing because the sick person is always depressed. (FGD3-P1)	Without health we have nothing	
Health is everything, if you have all the money in the world but you have poor health, it means you are poor and you own nothing. (FGD2-P3)	Health is more important than money	
At the end you give priority to health over other needs, this is how I think, for example I buy anything cheap, but I don't buy a cheap medicine to save money, and same for the doctor [...] a person should be frugal on everything except on his health, this is how I think [...]. (FGD3-P3)	Being frugal except with your health	

<https://doi.org/10.1371/journal.pone.0280665.t002>

Qualitative content analysis based on Graneheim & Lundman (2004) was used. The unit of analysis was the focus group discussion transcripts. All transcripts were read several times before coding was begun, to increase the understanding of the issues raised, as well as their depth and breadth. Statements were maintained uncondensed, and then labelled with codes. The statements were analyzed in relation to the specific research questions that also formed the basis for the content areas. Codes sharing communalities were used to construct categories based on expressed and explicit (manifest) meaning. The latent meaning of categories was subsequently used to develop themes. An example of the analytical process is found in [Table 2](#). NVivo 12.0 software was used to support the coding procedure.

Trustworthiness and methodological considerations

The 'trustworthiness' of our research was based on careful consideration of data credibility, dependability and transferability [24].

Credibility was improved by selecting a maximum variation of persons who had been recently discharged followed hospitalization under MoPH coverage. This sample purposefully included men and women, from different age groups and representing both rural and urban areas. Credibility was further enhanced by the first author (and moderator) having had experience with the MoPH in patient satisfaction and its P4P initiative. In addition, the moderator and research assistants (AH, EB) held debriefing sessions following each FGD [14]. The MoPH was not involved in the analysis of the data developed.

Dependability was improved by using a discussion guide, which ensured core questions were addressed, while allowing discussions to differently explore further topics.

To facilitate transferability, we have described the context of the study, the selection of participants and their general background, as well as the data collection and analytical process. We also made wide use of quotations where deemed appropriate. It is thus up to the readers to assess the relevance of the results in other settings.

Ethical considerations

The research protocol approval was granted by the Institutional Review Board (IRB) at the American University of Beirut (ID: FHS.FE.21). All participants were given oral information at the initial telephone invitation (by AH, EB) and again on-site at the introduction to the FGDs (by JK). This information included the purpose of the study; the fully voluntary nature of their participation and right of refusal; that the decision/participation would not be

associated with or affect their MoPH coverage in any manner; that all material would be handled confidentially; and no results would be presented allowing participant identification. Participants were given the choice to allow the discussions to be audio-recorded or not (all accepted recording). Participants were also asked to respect that “what’s said in the group stays in the group”, since the researcher could only promise confidentiality on behalf of the research group.

Results

The analysis resulted in five overall themes reflecting the underlying meaning of the discussions, supported by 17 categories giving the more manifest level of the interpretation. The overall results are presented in Fig 1 in relation to five content areas, based on the focus of the study. The first theme “Health is everything” indicates the high value that participants put on health. The second theme “Being turned into second class citizens” illustrates the inequity participants perceived that characterized the health system in general, while “Money and personal connections (*wasta*) make all the difference” specifies how this inequity is manifested in the actual access to healthcare access. The fourth theme “Wanting to be treated with dignity and respect” relates directly to what participants value most when assessing hospital care while the fifth theme “Tolerating letdown, for the sake of right treatment” implies what takes ultimately precedence is the medical results.

The section below presents the themes as headings in bold, while categories are in italics in the running text. Where relevant, quotes are included in italics, to support the analysis using the participants’ own words.

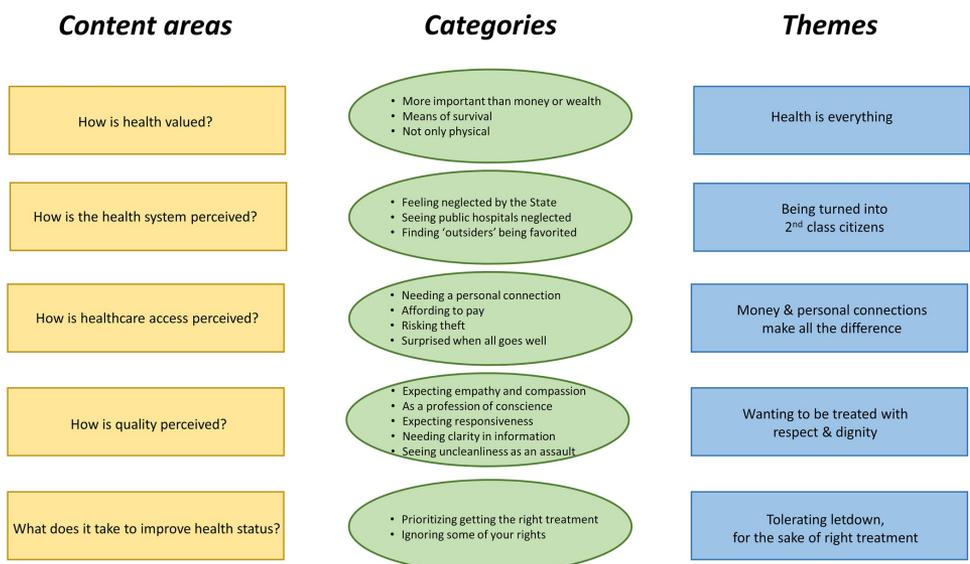


Fig 1. Overview of the main findings including content areas, categories and themes.

<https://doi.org/10.1371/journal.pone.0280665.g001>

Theme 1: Health is everything

This theme illustrates the valuing of health. Participants described health as everything, more important than anything else in life. Health was portrayed as **more important than money or wealth**, and they would exchange anything they own to regain their health. Health formed the crucial foundation for life. It was also considered a priority above other goods or services, whereby one may be frugal with other items, but not with health services or medications.

“Health is literally all life, if you don’t have health you don’t have a life; that’s it. Regardless of what your disease is, even if you only have headache, this is pain, and no one feels the pain except the patient himself.”

(FGD1 Men-P3)

“You give priority to health over other needs, this is how I think, for example I buy anything cheap, but I don’t buy a cheap medicine to save money, and same for the doctor.”

(FGD3, Women-P3)

Being healthy was not only for one’s own self, but was also important to be able to fulfill one’s role in the family and in the community. This meant supporting those dependent on you, and as a **means of survival** to be able to work and provide.

“Health is everything, I am a carpenter and I am paid on a daily basis, I have stopped working since a month and a half. Health is the basis of our existence, if we are not healthy we cannot work or do anything else.”

(FGD4 Men-P3)

Participants valued **not only physical health**, but also psychological and emotional well-being, which are important wherever a person is, be it their workplace or with their family. If you are not healthy, you also lose self-confidence. Although very important, health was also often neglected, as people get busied by other things in life and forget to take care of themselves, until they need a hospital.

“A person runs and endures, endures, endures and endures, and when time comes to relax, he finds himself unproductive because of his health, or could not anymore live the way he expected to live.”

(FGD1 Men-P4)

Theme 2: Being turned into second-class citizens

This theme relates to how the health system is perceived. Participants described how they felt like second-class citizens in their own country. Participants recognized that **“some hospitals are not for us”**. Some attributed this to which hospitals they had access to, while others perceived they were taken advantage of by hospitals, due to their educational status. They also reported that very often the first question they would be asked at the emergency room was whether they are under the coverage of private insurance or the NSSF. Participants recognized that those with greater financial means had much better access to healthcare than the poor.

“The hospitals in Lebanon are classified into classes, if you tell someone you are going to [well-known hospital X], they tell you ‘this is not for you’; this is the way they reply.”

(FGD1 Men-P1)

Participants **felt neglected by the state**, with citizens under the coverage of other payers (non-MoPH) having better healthcare access and services. They highlighted such preferential behavior particularly being given to those covered by the NSSF and private insurance. This was perceived from the first moment of interaction between patients and hospitals; usually the hospital admission desk, but even at the emergency rooms.

“My parents are old and don’t have NSSF nor insurance. [My father was ill and] the hospital asked for 650 USD, and I don’t have money. I borrowed the money from someone. This is how we live in Lebanon.”

(FGD8 Men-P6)

“Whenever any patient arrives to a hospital, he must be admitted to the ER, but the first question they ask you is: are you insured? Covered by NSSF? Army? If you have any of these they admit you, if you tell them you are a normal citizen they keep you on the stretcher.”

(FGD1 Men-P4)

A common perception was that hospitals had less respect for patients covered by the MoPH, as well as the ministry itself. They described the ministry as being weak, and they shared the excuses they would be given by hospitals for refusing hospitalizations, such as no beds being available. Participants thought that the MoPH should more strongly advocate for and safeguard the rights of the poor, and “make us feel that we are human beings”. They also wanted to see the ministry actively evaluate and regulate hospitals and primary care centers. This included investigating how some wealthy persons are reportedly receiving services under MoPH coverage, while some of the poor do not. While most participants did not have complaints regarding MoPH requirements for hospitalization or treatment, some participants considered themselves unfairly disadvantaged. Participants recognized the adverse role of what they considered was excessive bureaucracy, such as needing to travel to different locations to receive approvals for certain procedures (e.g. to the MoPH headquarters).

“If you want to benefit this country you need to think of cutting down on the processes; the current way is very tiring.”

(FGD4 Men-P4)

Participants reported **seeing public hospitals neglected** and having limited resources devoted to them. This neglect was translated into worse services and patients being treated with less dignity and respect than those going to private hospitals. They considered that public hospitals have a major role in supporting the poor, yet their potential was ignored. This neglect was also transferred to nurses, physicians and other staff at public hospitals.

P4: “If they improve the services of the public hospitals, there will be no need for private ones. We have a public hospital which is the largest hospital in the Middle East [...], but you cannot go there.”

Moderator: “Why?”

P1: "Because you would die!"

P4: "First the treatment is very bad, and the cleanliness is worse!"

(FGD3 Women)

Some participants proposed solutions to improve public hospitals. Some ideas shared were: increasing funding of public hospitals, allocating the better/best doctors to public hospitals ("we always follow the big names"), improving services, and accountability of hospital directors. Participants often avoided primary care centers. They perceived these centers to be under-staffed, and that patients had less time, and with less qualified doctors (compared to private clinics). They recognized that healthcare staff at these centers were underpaid, although they also have a duty because "there are also poor people who must be treated" (FGD5 Men-P1).

Participants also said they **found 'outsiders' being favored** with faster and free treatment. This referred to refugees with coverage from international non-governmental organizations and agencies.

Theme 3: Money and personal connections make all the difference

This theme illustrates how participants perceived access to healthcare. Participants predominantly linked this closely with the ability to have a 'wasta' ('who you know'/personal connection) or have enough money. Participants described that whatever the obstacles to access may be, one who pays directly out-of-pocket or has a personal connection would solve them. **Needing a personal connection** was recognized as a major factor in accessing healthcare, across the services spectrum. This particularly included hospital admission and surgical operations, but also medications for cancer and other chronic conditions.

Who you know was also important for healthcare institutions who are connected to political or religious figures/authorities. Participants considered that this relation functioned in both directions: for healthcare institutions to be secure from accountability, and for figures/authorities to use their influence to facilitate coverage for patients at the healthcare institutions.

P1: "There are connections and some people may be protected by others"

P3: "There are 'wasayet' [personal connections] all over the country"

[...]

P2: What is the relationship between hospitals and politics?

P1: I will tell you: 'hospital X' is for 'politician A', hospital Y' is for 'political party B', hospital Z is for 'religious council C'"

(FGD2 Women)

Another function of a personal connection was to decrease the hospital bill of patients, through the connection with political or religious figures. Participants reported several instances where their use of a personal connection for themselves or relatives resulted in a considerable decrease in their hospital bill: "We have some connections and they got me 20% off" (FGD8 Men-P1). Personal connections were also used to remedy perceived injustice or theft by the hospital through its physicians or administrative personnel: "The papers were signed after 'wasta' that the ministry can cover it" (FGD7 Women-P5). They also recognized that

knowing a connection in the ministry gave them an advantage in confrontations with hospitals. Also, in some instances, the treating physician was a friend of a patient, and they would intervene with a personal connection on behalf of the patient to lower the bill.

While the use of personal connections was considered widespread, participants recognized that this was a negative factor, although it was sometimes unavoidable to resort to. They thought that the health ministry should support them against such practices, and that getting rid of the use of personal connections everywhere would improve both hospital and the health system.

“As long as a person wants to be admitted through ‘wasta’, the hospital will not work properly.”

(FGD3 Women-P3)

Not all examples of connections were perceived negatively. Living near healthcare institution sometimes gave a relation of familiarity between patients and personnel. Some participants reported having very positive interactions with personnel in local hospitals, many of whom were either relatives or neighbors, or came from the same village or town. This was also sometimes reflected in more trustful behavior regarding payment, for example by patients being allowed to ‘pay later’, as they were known to the personnel. However, there were also the fewer instances where a participant reported being taken advantage of by personnel who were relatives or neighbors.

The financial cost of health services was a major concern for participants, specifically **affording to pay**. This also affected the perception and behavior of participants towards healthcare. Hospitals were perceived by participants to pursue money above all: “they want to take as much money as they can from the patients” (FGD4 Men-P4). Patients who are able to pay out-of-pocket would receive the best treatment, and sometimes they can only be hospitalized if they have the money.

“We ask God to stay healthy because we do not have money to pay for healthcare services.”

(FGD7 Women-P2)

Money was seen as a solution to any problem encountered at hospitals, especially if one lacked a personal connection. Patients reported that almost all problems occur either at admission or at the cashier. Sometimes hospitals would claim that no hospital beds are available, to deny admission for those covered by the MoPH. Some participants suggested this was a deceptive practice to allow hospitals to retain more profitable patients covered by other payers, or to compel MoPH-covered patients to pay out-of-pocket. One anticipatory approach mentioned by a participant was to claim to pay out-of-pocket at their initial interaction with hospital personnel, but after confirming bed availability he would then seek health ministry coverage.

“[. . .] I told the nurse my mother is not the daughter of a minister or a president; I cannot pay [out-of-pocket]. Then we took her to another hospital [. . .].”

(FGD3 Women-P4)

Unaffordability led some patients to early hospital discharge or to forego medical tests. Participants recognized this was harmful to their health, but they had no alternative that would allow them to pay for these services. Some participants recounted a family member pretending to be better to be discharged earlier, due to the mounting hospital bill. Others questioned the

utility of doing medical diagnostic tests, reasoning that since they cannot afford treatment, then it was better for their mental health not to know more about their illness. Some participants resorted to selling personal belongings to cover costs of medical tests and treatment, and sometimes had to forego necessary medical testing for themselves or their children, for conditions such as cancer and neurological illnesses.

“When I had breast cancer [...] I couldn’t do regular tests for checkup. I went through very hard times to do the tests and get the treatment. I sold my wedding ring [to get treatment]. The ministry couldn’t cover all the expenses; I reached a very difficult situation.”

(FGD7 Women-P1)

“My mother used to listen to the talk about money, and she’d indicate to want to leave the hospital, pretending that she is feeling well, though she was in a bad condition.”

(FGD1 Men-P1)

Resorting to borrowing from family and relatives was often a necessary measure, when the cost of healthcare was unaffordable for patients, despite the partial coverage by the MoPH. Sometimes this also necessitated a personal connection. But recurring costs such as those for chronic medications remained a burden. Participants emphasized that they have a right to health, but they do not know how to realize these rights.

The unemployed were particularly vulnerable. Without a source of income, these persons could not afford to pay for visits to doctors’ clinics (outpatient), not afford some of their medications for chronic conditions. The safety nets available for hospitalization (e.g. MoPH coverage) were not available for patient follow-up after leaving the hospital. The impact of unaffordable healthcare costs was not limited to patients alone. Family members would be actively engaged in collecting funds to cover hospitalization costs, as well as in gathering and submitting administrative papers for coverage approvals. This sometimes meant skipping university classes or work.

Risking theft when seeking healthcare was emphasized by participants. This was a major concern affecting their perception of hospitals as well as health professionals. Although services were often considered to be of good quality, the lack of information and transparency over hospital bills contributed to a feeling of patient distrust towards hospitals; as though ‘they were stealing from us’. Such practice took on different forms, and could involve different actors. A common complaint from participants was of a doctor or nurse misinforming that a procedure was not covered by the MoPH. Some participants also reported doctors soliciting bribes for signing admission approval papers for patients.

“[...] Then we knew that the ministry does cover the surgery, although the doctor has told us that it does not [...] my father stopped the cheque [...] The papers were signed after ‘wasta’ [...] The doctor had lied to us; why did he do that?”

(FGD7 Women-P5)

Participants noted that it was common to be asked by hospitals or physicians for payment above the MoPH pre-defined co-payment amount, sometimes by several times more. However, many were unaware that this was an illegal practice according to the contracting terms between the ministry and hospitals. In some instances, over-charging on co-payment was not hidden from patients, and hospitals or doctors attempted to justify this. Participants recognized that not all doctors are the same. Some were more helpful than others, in informing

patients of their rights under MoPH coverage and encouraging them to stand up for their rights. Also, some participants were surprised by the large differences in the cost of some surgical procedures between comparable hospitals, which they considered a signal of over-charging. Other participants noted instances where their copayment was high enough to cover most or all of a surgical procedure's cost.

“The nurse did not tell me that I have to do endoscopy and that it is covered by the ministry. I did not know that, so I paid out of pocket 900 USD the first time. So when I went inside, the doctor told me that we have to do the endoscopy, so I did it [. . .] the whole story is that they were not committed to the prices of the ministry. They said that the ministry reserves something and the hospital spends something else [. . .].”

(FGD1 Men-P6)

“I didn't have any problem with the process at the ministry, but they wanted me to pay 3500 USD at the hospital. We disagreed with them, even the physician disagreed with such payment, then they decreased the amount to 2500 USD [after the doctor spoke to the hospital].”

(FGD4 Men-P3)

Some participants reported not being informed of costly tests or procedures in a timely manner. Once hospitalized, patients would find these services unaffordable. A common response of hospital personnel in such instances was that the MoPH reimbursement to hospitals was insufficient to cover hospital costs.

Insufficient information regarding payment meant that patients had a weaker role in their interaction with hospitals. Participants were aware that they were the weaker of the three parties involved (MoPH and hospitals being the other two), and that they would sometimes bear the burden of mistakes made by the hospital. In some instances, this resulted in patients being over-charged. They were also aware of some of the limitations regarding hospital reimbursement from the MoPH. A downstream impact of these may be further over-charging on patient co-payments.

“Every patient admitted under the ministry's coverage doesn't know how much they are expected to pay [. . .]. The ministry delays its payments to hospitals, so [hospitals] want to benefit from another source.”

(FGD4 Men-P3)

Some participants noted that the actual payment and invoice amounts can differ, but they would be obliged to accept it to receive treatment. In some instances when leaving the hospital, participants would find themselves placed in an uncomfortable and embarrassing position by hospital personnel requesting payment for tests or services they claimed were not covered by the MoPH. Participants recognized the importance of speaking up about their challenges for healthcare access, especially hospitalization and medication costs. While they appreciated having coverage from the MoPH, it was far from sufficient for some.

Many participants also reported positive interactions with healthcare, whether with the ministry or hospitals. They were **surprised when all goes well**, especially when the administrative process for admission approval went smoothly; upon receiving coverage by the MoPH (85%) for their hospitalization costs; as well during hospitalization. Participants did not have to resort to neither a personal connection nor over-charges for their healthcare services in these instances.

Some participants reported being denied surgery or hospitalization under the coverage of private insurance companies they were subscribed to, and then being surprised to have such services under MoPH coverage. Such interactions affected the perception and trust of participants towards the MoPH. In some instances this was preceded by being let down due to exclusions by private insurance. Although this resulted in positive perceptions towards the health ministry, participants doubted they could rely on other ministries for delivering on other services; “no one cares about us regarding other issues” (FGD7 Women-P5).

Theme 4: Wanting to be treated with dignity and respect

This theme portrays how patients perceive the quality of care received at hospitals, expressed as wanting to be treated with dignity and respect, often implicitly but also explicitly. It also illustrates how patients view the health profession in itself. Reflecting on their past experiences, participants acknowledged both positive and negative interactions. These were not necessarily tied to the bio-physical outcome of treatment, but they did have an impact on how participants perceived hospitals.

Expecting empathy and compassion during their interactions with hospital personnel was very important. This included the manner of communication between health professionals and patients. The ease of obtaining admission approval and navigating administrative steps once hospitalized were crucial. Negative interactions had strong impressions on patients and their feelings of self-worth. They also had differing perceptions of private and public hospitals, with the former being considered to treat patients with more dignity and respect.

“They are putting me in an endless circle, and ultimately I am not getting anything out of it. How can I get admitted to a hospital with dignity?”

(FGD1 Men-P2)

“There is a huge difference between health services in public hospitals and private hospitals. Private hospitals treat patients with dignity and respect, which you can't see in public hospitals.”

(FGD8 Men-P4)

When asked what factors result in a positive experience during a hospital stay participants often referred to past interactions when healthcare personnel had been compassionate in their behavior towards them, especially when they had been in pain. The positive demeanor and care of nurses was particularly recalled by some participants.

Participants also recounted instances when doctors decreased patient co-payments (e.g. changing hospitals, personal connection), which influenced their perception of their doctor as a compassionate one. In one instance, a participant shared his story of being operated on and followed up by a physician without being asked for any payment.

“[The doctor said I needed surgery] and he visited me in the morning, and noon, and night. [...] He doesn't take a Lira from me. He knew I am poor and suffering. I told him I'm a farmer [...] He said 'my brother, this is helping someone in need.’”

(FGD8 Men-P4)

Seeing health practitioners belonging to a **profession of conscience** was also an expression of patients' desire to be treated with dignity and respect. They considered that being humane was the most important attribute of a doctor or nurse. Participants also noted that all

personnel working in healthcare institutions should be bound by the purpose of the institution, which they regarded as to help those in need. This was also specified for administrative personnel, including those in hospital admission and cashier roles.

“Humanity is the most important thing to be found at hospitals”

(FGD2 Women-P1)

This also meant that the over-arching priority should be ensuring every person’s ability to receive care, regardless of the ability to pay. This was considered as a fundamental right for humans, despite many being deprived of it.

Participants were also **expecting responsiveness** from staff while hospitalized, and reported both positive and negative experiences. There was understanding regarding the long working hours and challenging conditions that hospital staff worked under, particularly for nurses. Some suggested this to be a reason why responsiveness was sometimes lacking. More ‘difficult’ or demanding patients were also considered to be a challenge for staff. However, maintaining patient-centeredness was considered a necessity.

“When the nurse is in this profession, he must be expecting what he will face, he must not get annoyed and he must be patient. When the patient is at the hospital, he isn’t going to be faking it, he will really be in pain. This is why he will be nagging; because no one nags for no reason.”

(FGD7 Women-P5)

Staff responsiveness was particularly important when a patient was in pain, or needed aid to use the toilet. The lack of responsiveness in such situations led to strongly negative experiences by patients.

Participants highly valued the time personnel devoted to them. It was especially important to have enough time with the doctor, in order to have a clear explanation of their medical condition and treatment options, as well as to get responses to their questions. Doctors that did not make time for their patients were perceived as unresponsive and arrogant, regardless of their reasons.

Moderator: “How would you differentiate between a good and humane doctor, and a bad one?”

P4: “When he provides you with information, as I told you. Because my doctor’s clinic is so busy, if I want to ask him a question he says ‘there is no need to know about these things, I know about them’; this annoyed me.” [..]

P5: “The doctor is good when he gives you from his time, even though sometimes he is in a rush, but he has to make you relaxed, to explain your condition to you.”

(FGD7 Women)

It was also important to have a doctor you are comfortable with. Participants highlighted the importance of having a doctor they could rely on. ‘Following the doctor’ was how participants largely explained their decision to visit or re-visit a hospital. This sometimes included situations where they were not comfortable with the hospital. Participants were aware that countries would have a mix of better and worse doctors, and that none could be right or perfect all the time. They also acknowledged that chance also plays a role in whether you find a good doctor or not, as well as the importance of hearing the recommendation of friends or relatives before choosing a doctor.

Needing clarity in the information provided by both hospitals and the MoPH was important to participants. From hospitals, they expected more clarity on the treatment options and how long their stay may be. They highlighted the need clearer information from both the MoPH and hospitals regarding the amount for co-payment. More broadly, participants wanted to be more aware of their rights through the MoPH, and thus more empowered to defend themselves.

“The ministry must improve citizens’ awareness [...] We should know which hospitals we cannot be admitted to, the services and benefits we can get [...] When we are aware, we can fight for our rights.”

(FGD 7 Women-P5)

Participants emphasized that information clarity is perhaps even more important in health settings than in other (non-health) services since as a patient one is more vulnerable and dependent on others.

More broadly, participants thought it was important to know which hospitals were better performers; the location and medications accessible from medication dispensaries and primary care centers; and the cost (or co-payment) of surgical procedures under MoPH coverage. Regarding the latter, participants found it illogical and a lapse of accountability that co-payments could not be pre-determined and committed to. Many were unaware of the difference between a deposit receipt and their hospitalization bill, as well as their right to have a detailed hospital bill. There was minimal information provided at discharge regarding their bill, most of which was verbal, not written, with the exception of the receipt.

Cleanliness was an important issue for participants, particularly regarding toilets and bed-sheets. For some, this was more important than the medical treatment. Participants described some of their experiences and how they **saw uncleanliness as an assault** and as a danger to themselves.

“I was worried of getting a virus there; the toilets are not clean, such things makes me worried of getting an infection [...] hygiene and toilets are very important. I care about such things more than treatment [...] cleanliness is the most important factor.”

(FGD7 Women-P5)

Cleanliness was also something visible that patients sometimes directly associated with quality of treatment. Participants mentioned cues they would use to assess cleanliness. These included spotting spider webs on ceilings, and the frequency and timing of cleaning staff work. Cleanliness left a strong impression of hospitals among participants; “you see and sense cleanliness” (FGD4 Men-P3).

The responsiveness of nurses and cleaning staff after an incident was also important. Participants with a positive experiences appreciated being attended to quickly, while others recalled negative experiences after having to wait a few hours for a change of bedsheets soiled by their surgical drainage.

“They treated me well [...] my legs were swollen and I couldn’t step on the floor; I couldn’t use the toilet. They used to be next to me after one minute of ringing the bell to clean me up and wash me.”

(FGD5 Men-P3)

Theme 5: Tolerating letdown, for the sake of right treatment

This theme illustrates what participants see it takes to improve one's health status. They underscored that their purpose in being in a hospital was **to get the right treatment**, and that they generally prioritized this above all else. This was also a major reason why a patient would consider re-visiting a hospital. Some would accept being in a less-favored hospital, if it meant they could 'follow their doctor' and get appropriately treated.

"If nurses have a disagreement, it would affect you [...] Sometimes nurses are very nice, it depends on your chance [...] At the end we say it is fine and we thank God there is a hospital that admits us."

(FGD4 Men-P4)

Some participants noted that they tolerated some delays or behaviors, because they empathized with healthcare personnel's working conditions. Participants also suggested personnel should have shorter working schedules and more rest time.

Getting the right treatment sometimes involved **ignoring some of your rights**. Participants would sometimes be compelled to not voice their concerns or displeasure towards personnel, because they did not want to compromise on their treatment outcomes. In some instances, participants would try to overlook negative behaviors or incidents, and focus on having their health status improved.

"I ignore lots of things, you can say that I ignore 40–50% of my rights, the most important thing is to get the treatment."

(FGD4 Men-P4)

Pile sorting exercise

When asked directly to rank the 16 pre-formulated statements on patient satisfaction, the most frequently prioritized statement was regular contact with one's doctor, with nearly all participants considering this to be more important. This was followed by clear medication and care instructions at discharge, room cleanliness, shared decision-making, and good hospital organization (see Table 3). The least frequently prioritized statements were regarding food, privacy and ability to discuss fears and anxieties. However, it was notable that even these statements were still considered 'more important' by about half of the participants.

Statements were similarly prioritized by women and men participants, with few exceptions. Women more frequently considered it more important to be able to discuss any fears and anxieties (13:8), and to have a clear receipt from the hospital (16:12). Men more frequently considered it more important to have regular contact with their nurse (16:11), and to receive pain relief if in need (16:12).

The results of the sorting exercise reflected the wide range of factors that patients consider important during their hospitalization and affects their satisfaction. This provided 'patient satisfaction' as a sixth patient perspective, following the previously identified five perspectives. The most prioritized statements related to issues that were commonly raised by participants during the FGDs. Specifically, this applied across statements *a* to *k*. A notable exception to this was having clear instructions at discharge (*b*), which suggests a missed opportunity to engage discussion participants on this topic.

Table 3. Results from the pile sorting (n = 42).

Item	Statement	More important	Less important
a	To have regular contact with my doctor	41	1
b	To have clear medication and care instructions at discharge	36	6
c	To be in a clean hospital room	35	7
d	To be involved in any decision related to my health, and to have discussed the treatment plans	35	7
e	To have a clear and efficient hospital admission process	34	8
f	To be in a hospital that is well organized	33	9
g	To be treated with dignity and respect	32	10
h	To feel that communication is good and agreement among the medical personnel, regarding treatment plans	31	11
i	To receive pain relief if I am in pain	28	14
j	To be in a hospital with a good reputation	28	14
k	To have a clear receipt from the hospital	28	14
l	To have regular contact with my nurse	27	15
m	To have all my questions and concerns addressed	23	19
n	To have privacy during discussions with medical personnel	23	19
o	To be able to discuss any fears and anxieties	21	21
p	To have good quality food	15	27

<https://doi.org/10.1371/journal.pone.0280665.t003>

Discussion

In this qualitative study, we explored patient perspectives on hospital care in Lebanon. All participants had been recently discharged from a hospital under the insurance coverage of the Ministry of Public Health, which typically includes the poorest or most disadvantaged stratum of Lebanese. Our findings add to the limited international evidence base on patient perspectives, and how patients can evaluate their care [25]. A key message of this research is that patients appear to have a clear idea of what is needed to improve their healthcare access and experience.

Summary of findings

Health was above everything and necessary for living and having a disease need not be accompanied by dis-ease, if well managed. The participants held the health profession as one of conscience, where all its practitioners are held to a standard. Despite the prioritization of health, they saw that being under the coverage of MoPH meant that they were turned into second-class citizens and felt neglected by the State. They wanted to see more invested in public hospitals, which they considered have been neglected, despite being a national cornerstone for supporting the ill. The results also indicate that money and personal connections can overcome any barriers to healthcare, and that this group of patients were vulnerable to over-charging of co-payments. How to pay for one's healthcare was a major concern, causing worry for financial debt.

Participants also described positive experiences with healthcare, which were accompanied by feelings of surprise and even pride. This may reflect the wider resonance of negative experiences and barriers in shaping public perception. Health practitioners and institutions are not homogenous, and a patient may have the 'good fortune' to interact with those that are both professional and humane. During hospitalization patients wanted to be treated with dignity and respect. They expected personnel to be empathetic, compassionate and responsive to their needs. Cleanliness was important in affecting patients' perception of hospitals. They wanted to have sufficient time with their doctors to better understand their health status and treatment.

Patients wanted to know their rights to be able to break down barriers to healthcare, particularly knowing which services they were entitled to. They also wanted clearer information regarding their hospitalization bill and co-payment. Patients also thought it was important to know which are the better performing hospitals. Patients were sometimes let down, but they tolerated much of this, including compromising on their rights. They did this for the sake of getting treatment to improve their health.

Patients placed a distinctly high importance on the contact with their doctor, as well as hospital cleanliness. A wide range of factors were important for patients' hospital experience and satisfaction. Some were considered more important than others, particularly having clear information at discharge, room cleanliness and shared decision-making. Also, patients' satisfaction was not only determined by their interactions and surroundings, but also by their worries regarding payment.

How our findings relate to other studies

Many of our findings were in agreement with previous studies investigating what patients perceived to matter to them. Patients want to be treated by health professionals who are humane, informative, available and not money-driven [26–30].

The purpose of being hospitalized is to receive appropriate medical treatment or diagnosis. Nevertheless, our findings suggest that other factors may be equally or more important to some patients, such as having humane personnel and hospital cleanliness. This was not unusual, considering reports such as a 2018 England survey, which found that twice as many people would prioritize compassion over medical outcome, than those who would not [31]. While not detracting from the primacy of the medical outcome, this underlines the importance of compassion.

The importance of responsiveness was not limited to patient perceptions alone. It has also been linked to incidence of hospital-acquired infections, with poor responsiveness possibly acting as a symptom of wider hospital problems [32]. Participants referred to particular positive or negative experiences repeatedly, often to respond to very different issues. This suggests that discrete experiences can have a major role in shaping patient perceptions of both the hospital and the overall health system. This was also in agreement with research supporting the prominent impact of patient perceptions of care [33, 34].

Our study contributes towards untangling the patient experience from satisfaction. 'Patient satisfaction' has been the predominant term used to encompass the patient's perspective since the 1970s, originating from consumerist theories as an analogue to 'customer satisfaction'. In a healthcare context, however, satisfaction has been challenging to define. Our findings suggest that while satisfaction is important, it is a downstream result of other factors such as being comfortable with all or most aspects of care. Fundamentally, satisfaction is an emotion that refers to how patients feel. It is therefore not entirely explainable through objective reasoning. Most theories that attempted to explain satisfaction revolved around the relation with expectation [35]. Nevertheless, it is important to note that expectancy theories were insufficient, and expectation has been found to explain only a minor share of the variation in satisfaction reports [35]. Therefore, satisfaction should not be used alone, as a solitary reductionist measure, while ignoring other patient perspectives.

Patient experience is a term often used interchangeably with patient satisfaction. However, despite some overlap, the two terms are not the same. Patient experience may be defined as the sum of all interactions that influence perceptions of patients. Our findings concur with other research in that this usually reflects the perception of the quality of care experienced by the patient [36]. Our findings also revealed that the patient experience also had a wider influence,

such on perceptions regarding healthcare access or overall health system. What precisely is included in perception of quality is less clear, but can be ascertained through the questions included in different measurement tools.

How this may be useful for pay-for-performance

There are numerous tools designed to quantify the patient's perception of quality in hospitals, some of which have been used in different pay-for-performance initiatives. A few are considered of high-quality, such as the Ethiopian Patient Experiences with Inpatient Care (I-PAHC), Indian Patient Perceptions of Quality (PPQ), and the US Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS) [37]. Such tools are predominantly developed using expert opinion, rather than being informed directly by patient perspectives. As such, to improve the validity of these tools, it is important to use qualitative research conducted with patients themselves [30]. Quantitative and qualitative investigations are complementary. Although qualitative investigations make comparisons and generalizations difficult, their strength lies in being based on the patients' own words [15]. Our research contributes information that may be used in improving the validity of quantitative tools used to evaluate patient perspectives.

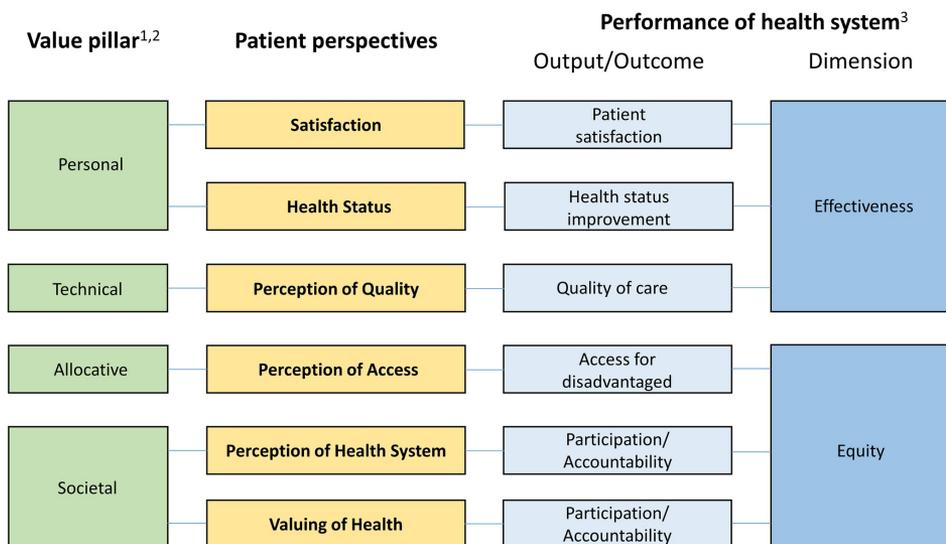
In the Lebanese context, the tool currently used by the MoPH for its P4P patient-related component was a locally adapted and abbreviated version of the US HCAHPS. By design, this focused predominantly on perception of quality. Our research identified issues that are partly or entirely uncaptured by this tool. Some of these related to perception of quality (e.g. information clarity), but also to other patient perspectives (e.g. risk of theft). It may therefore be useful for the MoPH to consider revision of this tool to enhance its validity. We also note that a preliminary analysis of this data was used to modify some items of the MoPH tool for the 2018 patient survey (e.g. time spent with personnel, discharge information). Our findings also suggest several practical actions that can be undertaken to improve patient healthcare access and experience, such as those relating to co-payment and over-charging.

More widely, our findings may also be useful towards developing a Lebanese patient-centered health system. The high value patients attach to health is not necessarily reflected in national governance and spending. Though patients shared numerous positive experiences, our findings suggest there is much space for development of the health system, particularly towards supporting public hospitals and increased accountability of health institutions and personnel. Healthcare cannot always deliver on cure, but it should also address wider patient perspectives and not only health status. This is perhaps best expressed in the oft-cited aphorism adopted by the physician Edward Livingston Trudeau for the Saranac Lake sanatorium: "to cure sometimes, to relieve often, to comfort always".

How this relates to value-based care and health systems

We also sought to examine how patient perspectives relate to value-based care and person-centered health systems. We had identified six distinct but related patient perspectives. While many health systems include some measures of perception of quality, others are largely uncaptured, namely the valuing of health, perception of access and perception of overall health system. Some systems additionally measure health status, including functionality, often using patient reported outcome measures. We note that these are absent at most Lebanese hospitals.

Given the absence of a framework to relate the six patient perspectives to health system performance and value-based healthcare, we developed the framework shown in Fig 2. This uses the value pillars recently proposed by the WHO EU Health Observatory and the European Commission, together with the Kruk and Freedman framework for health systems performance [7, 38, 39]. The content areas in Fig 1 link to the patient perspectives in Fig 2.



¹ Smith PC, et al (2020). Building on value-based health care: Towards a health system perspective. European Observatory Policy Briefs. World Health Organization.

² European Commission (2019). Defining Value in 'Value-Based Healthcare'. Report of the Expert Panel on effective ways of investing in Health.

³ Kruk M., Freedman L. (2008). Assessing Health System Performance in Developing Countries: A Review of the Literature; p. 263-276.

Fig 2. Relating patient perspectives to value-based care and health systems performance.

<https://doi.org/10.1371/journal.pone.0280665.g002>

Besides demonstrating the non-interchangeability of patient satisfaction and experience, we also suggest this framework to clarify the current and potential role that patients can have within health systems. Engaging with patients need not be limited to measurement of quality (indirectly), satisfaction and health status. It may also involve assessing healthcare access and the overall health system, thus contributing towards decreasing inequity within populations. How patients value health is important to capture, as it relates to both accountability and participation within a health system.

Value-based care has traditionally been focused on technical value (i.e. cost-effectiveness). More recent initiatives have in addition proposed allocative, societal and personal values [7, 39]. These generally concern the equitable resource distribution within a population (allocative value); the contribution of healthcare towards solidarity, connectedness and social cohesion (societal value); and patient-centeredness (personal value). Relating these values to different patient perspectives allows value-based programs (including pay-for-performance) to more widely incentivize values including allocative and societal values which are largely unaddressed.

Limitations

Participants shared a wide range of opinions, many of which was critical, as well as very personal stories. This suggests that they felt comfortable to speak freely. However, while some participants did share critical opinions of the ministry, we cannot rule out potential impact of the

host location (MoPH headquarters). It is possible that some FGD participants influenced the responses of fellow participants, for example by unintentionally inhibiting others from sharing sensitive experiences or by expressing dominant opinions. Our study design, being qualitative in nature, cannot be used to generalize broadly. However, our findings offer a range of issues regarding patient perspectives, many of which are relevant for better engaging patients in Lebanon and elsewhere.

Conclusion

Patient perspectives include more than patient experience and satisfaction. In addition to traditional measures, patients may also be engaged in their valuing of health, and perceptions on healthcare access and quality of care.

The drive towards patient or people-centered health systems should incorporate a wider consideration of patient perspectives. Pay-for-performance initiatives can also be more responsive and better align patient and provider interests by adopting a broader consideration of patient perspectives.

This study specifically highlights the importance of health to people in Lebanon, and the need to prioritize health services to match with people's expectations. Patients want to be treated with dignity and respect by hospitals throughout their healthcare journey. Addressing inequity should include curbing the influence of personal connections and greater protection against financial exploitation by providers. The standardization or unification of coverage among healthcare payers would circumvent patients being turned into "second-class citizens".

Hippocratic medicine was centered on accompanying the patient and meeting their individual goals. Health systems aiming to return to this fundamental aspect should more widely engage patients for their perspectives, and incorporate these perspectives within clinical practice and health system design.

Supporting information

S1 Fig. Demarcating where patient experience occurs in relation to the six patient perspectives.

(TIF)

S1 Appendix.

(DOCX)

Acknowledgments

We are very grateful to all the participants in the focus group discussions, without whom this research would not have been possible. We are also grateful to following individuals at the Lebanese MoPH: Rita Freiha, Hilda Harb and Jihad Makouk (feedback and insight); and Rabiaa Rachid and Jenny Romanos (data extraction for sampling); and to Anne Mills (London School of Hygiene and Tropical Medicine, project senior advisor).

Author Contributions

Conceptualization: Jade Khalife, Björn Ekman, Maria Emmelin.

Formal analysis: Jade Khalife.

Funding acquisition: Jade Khalife, Björn Ekman, Walid Ammar, Fadi El-Jardali, Maria Emmelin.

Investigation: Jade Khalife, Abeer Al Halabi, Elise Barakat.

Methodology: Jade Khalife, Maria Emmelin.

Project administration: Jade Khalife.

Resources: Walid Ammar, Fadi El-Jardali.

Supervision: Björn Ekman, Walid Ammar, Fadi El-Jardali, Maria Emmelin.

Validation: Jade Khalife, Björn Ekman, Abeer Al Halabi, Elise Barakat, Maria Emmelin.

Visualization: Jade Khalife.

Writing – original draft: Jade Khalife.

Writing – review & editing: Jade Khalife, Björn Ekman, Walid Ammar, Fadi El-Jardali, Abeer Al Halabi, Elise Barakat, Maria Emmelin.

References

1. Griffiths R, Bett M, Blyth J, Bailey B. NHS Management Inquiry. London: 1983.
2. Institute of Medicine Committee on Quality of Health Care in A. Crossing the Quality Chasm: A New Health System for the 21st Century. Washington (DC): National Academies Press (US); 2001.
3. National Academies of Sciences E, Medicine, Health, Medicine D, Board on Health Care S, Board on Global H, et al. Crossing the Global Quality Chasm: Improving Health Care Worldwide. Washington (DC): National Academies Press (US); 2018.
4. World Health Organization. Framework on integrated, people-centred health services: report by the Secretariat. Geneva: World Health Assembly, 2016.
5. World Health Organization, Organisation for Economic Co-operation and Development, The World Bank. Delivering quality health services: a global imperative for universal health coverage. Geneva: 2018.
6. Berwick DM, editor. Speakers and moderators at the Policy Forum on the Future of Health: statement by Donald M. Berwick. 2017; Paris: Organisation for Economic Cooperation and Development.
7. Smith PC, Sagan A, Siciliani L, Panteli D, McKee M, Soucat A, et al. European Observatory Policy Briefs. Building on value-based health care: Towards a health system perspective. Copenhagen (Denmark): World Health Organization; 2020.
8. World Health Organization. The world health report: health systems financing: the path to universal coverage. Geneva, 2010.
9. Ng JHY, Luk BHK. Patient satisfaction: Concept analysis in the healthcare context. *Patient Educ Couns*. 2019; 102(4):790–6. Epub 2018/11/28. <https://doi.org/10.1016/j.pec.2018.11.013> PMID: 30477906
10. Gillam SJ, Siriwardena AN, Steel N. Pay-for-performance in the United Kingdom: impact of the quality and outcomes framework: a systematic review. *Ann Fam Med*. 2012; 10(5):461–8. Epub 2012/09/12. <https://doi.org/10.1370/afm.1377> PMID: 22966110
11. Joynt Maddox KE, Sen AP, Samson LW, Zuckerman RB, DeLew N, Epstein AM. Elements of Program Design in Medicare's Value-based and Alternative Payment Models: a Narrative Review. *J Gen Intern Med*. 2017; 32(11):1249–54. Epub 2017/07/19. <https://doi.org/10.1007/s11606-017-4125-8> PMID: 28717900
12. Cashin C, Chi Y-LS, P. C., Borowitz M, Thomson S. Paying for Performance in Health Care: implications for health system performance and accountability. England: World Health Organization; 2014.
13. Khalife J, Ammar W, Emmelin M, El-Jardali F, Ekman B. Hospital performance and payment: impact of integrating pay-for-performance on healthcare effectiveness in Lebanon. *Wellcome Open Research*. 2020; 5(95). <https://doi.org/10.12688/wellcomeopenres.15810.2> PMID: 33437874
14. Dahlgren L, Hällgren Graneheim U, Sahlén KG, Emmelin M, Winkvist A. Qualitative methodology for international public health. Umeå: Umeå universitet; 2019.
15. LaVela SG, A S. Evaluation and measurement of patient experience. *Patient Experience Journal*. 2014; 1(1):28–36.
16. Hennick M. International Focus Group Research: A Handbook for the Health and Social Sciences. New York: Cambridge University Press; 2007.
17. O'Brien BC, Harris IB, Beckman TJ, Reed DA, Cook DA. Standards for Reporting Qualitative Research: A Synthesis of Recommendations. *Academic Medicine*. 2014; 89(9):1245–51. <https://doi.org/10.1097/ACM.0000000000000388> PMID: 24979285

18. Burton ML, Romney AK. A multidimensional representation of role terms. *American Ethnologist*. 1975; 2(3):397–407. <https://doi.org/10.1525/ae.1975.2.3.02a00020>
19. Shipstone EI. Some variables affecting pattern conception. *Psychological Monographs: General and Applied*. 1960; 74, No. 17(Whole No. 504).
20. Yeh HW, Gajewski BJ, Perdue DG, Cully A, Cully L, Greiner KA, et al. Sorting it Out: Pile Sorting as a Mixed Methodology for Exploring Barriers to Cancer Screening. *Qual Quant*. 2014; 48(5):2569–87. Epub 2014/08/22. <https://doi.org/10.1007/s11135-013-9908-3> PMID: 25143659
21. Quintiliani LM, Campbell MK, Haines PS, Webber KH. The Use of the Pile Sort Method in Identifying Groups of Healthful Lifestyle Behaviors among Female Community College Students. *Journal of the American Dietetic Association*. 2008; 108(9):1503–7. <https://doi.org/10.1016/j.jada.2008.06.428> PMID: 18755323
22. Fitzpatrick R. Surveys of patients satisfaction: I—Important general considerations. *Bmj*. 1991; 302(6781):887–9. Epub 1991/04/13. <https://doi.org/10.1136/bmj.302.6781.887> PMID: 1821624
23. Giordano LA, Elliott MN, Goldstein E, Lehrman WG, Spencer PA. Development, implementation, and public reporting of the HCAHPS survey. *Med Care Res Rev*. 2010; 67(1):27–37. Epub 2009/07/30. <https://doi.org/10.1177/1077558709341065> PMID: 19638641
24. Graneheim UH, Lundman B. Qualitative content analysis in nursing research: concepts, procedures and measures to achieve trustworthiness. *Nurse Educ Today*. 2004; 24(2):105–12. Epub 2004/02/11. <https://doi.org/10.1016/j.nedt.2003.10.001> PMID: 14769454
25. Batbaatar E, Dorjdagva J, Luvsannyam A, Savino MM, Amenta P. Determinants of patient satisfaction: a systematic review. *Perspectives in Public Health*. 2016; 137(2):89–101. <https://doi.org/10.1177/1757913916634136> PMID: 27004489
26. Deledda G, Moretti F, Rimondini M, Zimmermann C. How patients want their doctor to communicate. A literature review on primary care patients' perspective. *Patient Education and Counseling*. 2013; 90(3):297–306. <https://doi.org/10.1016/j.pec.2012.05.005> PMID: 22709720
27. Luthy C, Cedraschi C, Perrin E, Allaz AF. How do patients define "good" and "bad" doctors? *Swiss Med Wkly*. 2005; 135(5–6):82–6. Epub 2005/02/25. <https://doi.org/10.4414/smw.2005.10839> PMID: 15729612
28. Borracci RA, Álvarez Gallesio JM, Ciembrone G, Matayoshi C, Rossi F, Cabrera S. What patients consider to be a good doctor, and what doctors consider to be a good patient. *Revista médica de Chile*. 2020; 148:930–8. <https://doi.org/10.4067/S0034-98872020000700930> PMID: 33399677
29. Quirk M, Mazor K, Haley H-L, Philbin M, Fischer M, Sullivan K, et al. How patients perceive a doctor's caring attitude. *Patient Education and Counseling*. 2008; 72(3):359–66. <https://doi.org/10.1016/j.pec.2008.05.022> PMID: 18684582
30. Wensing M, Jung HP, Mainz J, Olesen F, Grol R. A systematic review of the literature on patient priorities for general practice care. Part 1: Description of the research domain. *Soc Sci Med*. 1998; 47(10):1573–88. Epub 1998/11/21. [https://doi.org/10.1016/s0277-9536\(98\)00222-6](https://doi.org/10.1016/s0277-9536(98)00222-6) PMID: 9823053
31. HealthWatch-England. What do people want from the NHS and social care in the future. 2018.
32. Saman DM, Kavanagh KT, Johnson B, Lutfiyya MN. Can Inpatient Hospital Experiences Predict Central Line-Associated Bloodstream Infections? *PLOS ONE*. 2013; 8(4):e61097. <https://doi.org/10.1371/journal.pone.0061097> PMID: 23577195
33. Rapport F, Hibbert P, Baysari M, Long JC, Seah R, Zheng WY, et al. What do patients really want? An in-depth examination of patient experience in four Australian hospitals. *BMC Health Serv Res*. 2019; 19(1):38. Epub 2019/01/17. <https://doi.org/10.1186/s12913-019-3881-z> PMID: 30646962
34. Schoenfelder T, Klewer J, Kugler J. Determinants of patient satisfaction: a study among 39 hospitals in an in-patient setting in Germany. *Int J Qual Health Care*. 2011; 23(5):503–9. Epub 2011/07/01. <https://doi.org/10.1093/intqhc/mzr038> PMID: 21715557
35. Batbaatar E, Dorjdagva J, Luvsannyam A, Amenta P. Conceptualisation of patient satisfaction: a systematic narrative literature review. *Perspectives in Public Health*. 2015; 135(5):243–50. Epub 2015/07/19. <https://doi.org/10.1177/1757913915594196> PMID: 26187638
36. Berkowitz B. The Patient Experience and Patient Satisfaction: Measurement of a Complex Dynamic. *Online J Issues Nurs*. 2016; 21(1):1. Epub 2016/11/18. <https://doi.org/10.3912/OJIN.Vol21No01Man01> PMID: 27852212
37. Beattie M, Murphy DJ, Atherton I, Lauder W. Instruments to measure patient experience of healthcare quality in hospitals: a systematic review. *Syst Rev*. 2015; 4:97. Epub 2015/07/24. <https://doi.org/10.1186/s13643-015-0089-0> PMID: 26202326
38. Kruk M, Freedman L. Assessing Health System Performance in Developing Countries: A Review of the Literature 2008. 263–76 p.
39. European Commission. Defining Value in 'Value-Based Healthcare'. Report of the Expert Panel on effective ways of investing in Health. 2019.

The Development and Evaluation of Hospital Pay-for-Performance in Lebanon

Jade has a background in medicine, health policy and management, and epidemiology. His past experiences include being involved with the Ministry of Public Health (MoPH) in Lebanon, from 2009 to 2020, primarily on health system initiatives regarding hospitals and primary care centers. The purpose of this thesis was to describe the development and evaluate the impact of hospital pay-for-performance (P4P) in Lebanon, and ultimately to contribute to improved design and implementation of value-based healthcare, particularly in limited resource settings. This thesis uses a mixed methods approach, including interrupted time series analysis and qualitative investigations. The findings reveal several positive impacts of pay-for-performance in Lebanon, including the improvement of the relation between hospitals and the MoPH, and providing a tool for continuous development of the health system. The 2014 and 2018 P4P interventions improved system effectiveness and related patient outcomes, by decreasing unnecessary hospitalizations and decreasing some types of readmissions. Patients in Lebanon highly valued health and supported improving public hospitals and measures to counter the influence of personal connections and money. Health systems can more widely engage people for their perspectives, and patients can have a fundamental role in shaping the values and functions of a health system.

