THE EFFECTIVENESS OF THE PAY-FOR-PERFORMANCE SCHEME IN HEALTH CARE SYSTEMS WITH UNIVERSAL HEALTH COVERAGE: A SYSTEMATIC REVIEW

NITHIMAR SERMSUTI-ANUWAT

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR
THE DEGREE OF MASTER OF SCIENCE (PUBLIC HEALTH)
PROGRAM IN HEALTH ADMINISTRATION
FACULTY OF GRADUATE STUDIES
MAHIDOL UNIVERSITY
2014

Thesis
entitled

THE EFFECTIVENESS OF THE PAY-FOR-PERFORMANCE SCHEME IN HEALTH CARE SYSTEMS WITH UNIVERSAL HEALTH COVERAGE: A SYSTEMATIC REVIEW

.....
Miss Nithimar Sermsuti-anuwat
Candidate

.....
Asst. Prof. Pirudee Pavananunt,
Ph.D.(Development Administration)
Major advisor

.....
Lect. Suthee Rattanamongkolgul,
Ph.D.(Epidemiology)
Co-advisor

.....
Mr. Suwit Wibulpolprasert,
M.D., Diploma Thai Board of Preventive Medicine
Co-advisor

.....
Prof. Banchong Mahaisavariya,
M.D., Diploma Thai Board of Orthopedics
Dean
Faculty of Graduate Studies
Mahidol University

.....
Assoc. Prof. Surachart Na Nongkhai,
Psy.D.(Management)
Acting Program Director
Master of Science (Public Health)
Program in Health Administration
Faculty of Public Health
Mahidol University

Thesis
entitled

THE EFFECTIVENESS OF THE PAY-FOR-PERFORMANCE SCHEME IN HEALTH CARE SYSTEMS WITH UNIVERSAL HEALTH COVERAGE: A SYSTEMATIC REVIEW

was submitted to the Faculty of Graduate Studies, Mahidol University
for the degree of Master of Science (Public Health)
Program in Health Administration
on
November 27, 2014

.....
Miss Nithimar Sermsuti-anuwat
Candidate

.....
Asst. Prof. Wittaya Tonsuwannont,
Ph.D.(Pharmacology)
Chair

.....
Asst. Prof. Pirudee Pavananunt,
Ph.D.(Development Administration)
Member

.....
Mr. Suwit Wibulpolprasert,
M.D., Diploma Thai Board of Preventive Medicine
Member

.....
Lect. Suthee Rattanamongkolgul,
Ph.D.(Epidemiology)
Member

.....
Prof. Banchong Mahaisavariya,
M.D., Diploma Thai Board of Orthopedics
Dean
Faculty of Graduate Studies
Mahidol University

.....
Assoc. Prof. Prayoon Fongsatitkul,
Ph.D.(Environment Engineering)
Dean
Faculty of Public Health
Mahidol University

ACKNOWLEDGEMENTS

I would like to express my sincere thanks to my supervisor, Asst. Prof. Dr.Pirudee Pavananunt for her valuable suggestions and guidance during the course of research execution and thesis writing without which this research would not have been completed. Many thanks to Dr.Suthee Rattanamongkolgul, my co-advisor, for his helpful and kind comments at various points of this research project. To Dr.Suwit Wibulpolprasert, also my co-supervisor, goes a warm appreciation for his constantly overseeing this research work and valuable discussions. I would also like to take this opportunity to extend my gratitude to the Chairman of my oral examination; Asst. Prof. Dr.Wittaya Tonsuwannont for his valuable time and his comments, which help make this thesis effectively concluded.

Last but not the least, thanks to my parents and my sister's family for their supports and assistance.

Nithimar Sermsuti-anuwat

THE EFFECTIVENESS OF THE PAY-FOR-PERFORMANCE SCHEME IN HEALTH CARE SYSTEMS WITH UNIVERSAL HEALTH COVERAGE: A SYSTEMATIC REVIEW

NITHIMAR SERMSUTI-ANUWAT 5437328 PHPH/M

M.Sc.(PUBLIC HEALTH) PROGRAM IN HEALTH ADMINISTRATION

THESIS ADVISORY COMMITTEE: PIRUDEE PAVANANUNT, Ph.D.(DEVELOPMENT ADMINISTRATION), SUTHEE RATTANAMONGKOLGUL, Ph.D.(EPIDEMIOLOGY), SUWIT WIBULPOLPRASERT, M.D., DIP. THAI BOARD OF PREVENTIVE MEDICINE

ABSTRACT

The main purpose of this systematic review was to examine the effectiveness of pay-for-performance (P4P) for improving health service quality and the accessibility of care in the universal health coverage (UHC) context.

Searches were carried out in five electronic databases: Cochrane Library, MEDLINE, PUBMED, EBSCO, and CINAHL. Only the research papers published in English between 2000 and 2013 were included in this review based on the following inclusion criteria: (1) the studies were conducted in countries that provided UHC; (2) P4P were implemented on the supply side; (3) the quality performances of the outcomes were reported. The quality of the studies was then assessed by using the modified version of The Newcastle-Ottawa Scale (NOS).

The electronic search obtained 2,264 publications of which 23 papers met all the inclusion criteria. Most of the studies reported that the achievement of service quality outcome reached the set targets in the period of P4P implementation rather than in the non-P4P period. The P4P scheme implemented in the context of the lower baseline of performance showed more improvements than at the higher baseline. In addition, the P4P scheme could enhance a steady increase in both quality of services and accessibility in the first three years of implementation; after this, the growth rates declined, but still showed improvements.

This study suggests that pay-per-performance (P4P) is probably the most efficient incentive scheme for the low productive health care service areas. Furthermore, the effectiveness of P4P lasts for a reliable period; after that a new incentive scheme may be considered to boost the outcomes.

KEY WORD: PAY-FOR-PERFORMANCE / UNIVERSAL HEALTH COVERAGE

153 pages

ประสิทธิผลของการจ่ายค่าตอบแทนตามผลปฏิบัติงานในระบบสุขภาพที่มีหลักประกันสุขภาพถ้วนหน้า: การทบทวนวรรณกรรมอย่างเป็นระบบ

THE EFFECTIVENESS OF THE PAY-FOR-PERFORMANCE SCHEME IN HEALTH CARE SYSTEMS WITH UNIVERSAL HEALTH COVERAGE: A SYSTEMATIC REVIEW

นิธิมา เสริมสุธีอนุวัฒน์ 5437328 PHPH/M

วท.ม.(สาธารณสุขศาสตร์) สาขาวิชาเอกบริหารสาธารณสุข

คณะกรรมการที่ปรึกษาวิทยานิพนธ์: ภิฤดี ภวนานันท์, Ph.D.(DEVELOPMENT ADMINISTRATION), สุธีร์ รัตนะมงคลกุล, Ph.D.(EPIDEMIOLOGY), สุวิทย์ วิบุลผลประเสริฐ, M.D., DIP. THAI BOARD OF PREVENTIVE MEDICINE

บทคัดย่อ

วัตถุประสงค์หลักของการทบทวนวรรณกรรมอย่างเป็นระบบนี้คือ เพื่อศึกษาประสิทธิผลของการจ่ายค่าตอบแทนตามผลปฏิบัติงาน (pay-for-performance) ต่อการเพิ่มคุณภาพในการรักษาและการเพิ่มการเข้าถึงการให้บริการด้านสุขภาพ ภายใต้บริบทของระบบสุขภาพที่มีหลักประกันสุขภาพถ้วนหน้า โดยการสืบค้นงานวิจัยที่เกี่ยวข้องจากฐานข้อมูลอิเล็คทรอนิคส์ ดังนี้ Cochrane Library, MEDLINE, PUBMED, EBSCO และ CINAHL ขอบเขตงานวิจัยที่นำมาทบทวนวรรณกรรม เป็นงานวิจัยภาษาอังกฤษ ตีพิมพ์ระหว่างปี ค.ศ.2000-2013 ตามเกณฑ์คัดเข้า (Inclusion criteria) คือ (1) เป็นงานวิจัยที่ทำการศึกษาภายใต้บริบทของระบบสุขภาพที่มีหลักประกันสุขภาพถ้วนหน้า (2) มีการการจ่ายค่าตอบแทนตามผลปฏิบัติงานให้แก่ผู้ให้บริการด้านสุขภาพ (3) มีการรายงานผลการปฏิบัติงานเชิงคุณภาพ การประเมินคุณภาพของงานวิจัยใช้แบบประเมิน Newcastle-Ottawa Scale (NOS)

ผลการสืบค้นได้งานวิจัยจากฐานข้อมูลจำนวน 2,264 เรื่อง โดยมี 23 เรื่อง ตรงตามเกณฑ์การคัดเข้า พบว่างานวิจัยส่วนใหญ่รายงานว่า การจ่ายค่าตอบแทนตามผลปฏิบัติงาน สามารถเพิ่มคุณภาพในการรักษาและเพิ่มการเข้าถึงการให้บริการด้านสุขภาพ

งานวิจัยนี้พบว่าการใช้การจ่ายค่าตอบแทนตามผลปฏิบัติงาน (pay-for-performance) ในงานบริการสุขภาพที่มีผลประสิทธิผลต่ำ จะช่วยเพิ่มประสิทธิผลได้ดีกว่าในงานที่มีประสิทธิผลดีอยู่แล้ว และพบว่าประสิทธิผลจะเพิ่มขึ้นภายในระยะเวลา 3 ปีแรก หลังจากนั้นประสิทธิผลจะลดลง แต่ยังคงสูงกว่าก่อนนำการจ่ายค่าตอบแทนตามผลปฏิบัติงาน (pay-for-performance) มาใช้ จึงจำเป็นต้องมีการประเมินประสิทธิผลของงานอย่างต่อเนื่อง

153 pages

CONTENTS

		Page
ACKNOWLEDGEMENTS……		**iii**
ABSTRACT (ENGLISH) ……..…		**iv**
ABSTRACT (THAI) ……..……....		**v**
LIST OF TABLES…….…		**viii**
LIST OF FIGURES……		**ix**
CHAPTER I	**INTRODUCTION……..…**	**1**
1.1	Background and Rationale.....	1
1.2	Research Questions.....	4
1.3	Research Objectives.....	4
1.4	Conceptual Framework.....	4
1.5	Operational Definitions.....	5
CHAPTER II	**LITERATURE REVIEW…….. **	**6**
2.1	Universal health coverage (UHC).....	6
2.2	Pay-for-performance (P4P).....	9
2.3	Systematic review.....	19
CHAPTER III	**METHODOLOGY……**	**28**
3.1	Study design: Systematic review guidelines.....	28
3.2	Scope of the study.....	28
3.3	Study time period.....	28
3.4	Data sources.....	28
3.5	Search methods for identification of primary studies.....	29
3.6	The full search strategies.....	30
3.7	Criteria for considering studies for this review.....	37
3.8	Data extraction.....	38
3.9	Quality assessment.....	39
3.10	Data syntheses.....	39

CONTENTS (cont.)

		Page
3.11	Ethical consideration.....	43
CHAPTER IV	**RESULTS……**	**44**
4.1	Result of the searches.....	44
4.2	Characteristics of the included studies.....	56
4.3	Subgroup analysis.....	61
CHAPTER V	**DISCUSSION……..….**	**86**
5.1	Research findings.....	86
5.2	The limitation of the study.....	90
CHAPTER VI	**CONCLUSION……**	**91**
6.1	Conclusion.....	91
6.2	Recommendations.....	91
REFERENCES……		**94**
APPENDICES……..……..……		**106**
BIOGRAPHY……..…..		**153**

LIST OF TABLES

Table		Page
2.1	Pay-for-performance (P4P) definition…….	10
2.2	The positive effects of pay-for-performance (P4P) incentive in healthcare systems with universal health coverage……..	13
2.3	The partial effects of pay-for-performance (P4P) incentive in healthcare systems with universal health coverage……..	16
2.4	The adverse effects of pay-for-performance (P4P) incentive in healthcare systems with universal health coverage……..	18
2.5	No effects of pay-for-performance (P4P) incentive in healthcare systems with universal health coverage……	18
3.1	The search strategies from CINAHL Plus with Full Text accessed through Mahidol database…….	37
4.1	Characteristics of excluded studies……..……	46
4.2	Characteristics of twenty-three included studies……..……..	47
4.3	The areas of P4P implementation in the included studies……	57
4.4	The quality assessment of the included studies……..…….	58
4.5	Subgroup analysis of the blood pressure targets estimates……..	63
4.6	Subgroup analysis of the cholesterol targets estimates……	71
4.7	Subgroup analysis of pressure recorded estimates…..……	75
4.8	Subgroup analysis of the cholesterol recorded estimates…..……..	80

LIST OF FIGURES

Figures		Page
1.1	Conceptual framework of the effects of pay-for-performance (P4P) payment implementation for the universal health coverage (UHC) achievements	4
2.1	Three broad dimensions of universal health coverage (UHC) ..……..	7
2.2	P4P concept defined by United States Agency for International Development (USAID) …….	9
2.3	Structure of pay-for-performance (P4P) ……..……	11
2.4	Flow chart of a systematic review process (adapted from the Systematic Reviews of Health Promotion and Public Health Interventions: train-the-trainer handbook, Cochrane Health Promotion and Public Health Field) ……...	19
2.5	The heterogeneity and type of data synthesis……...	25
3.1	The search strategies from the Cochrane Library……..…	31
3.2	The search strategies from PubMed…….	32
3.3	The search strategies from Ovid Medline accessed through Mahidol database.……	32
3.4	The search strategies from EBSCO Discovery Service accessed through Mahidol database……..	35
3.5	The search strategies from CINAHL Plus with Full Text accessed through Mahidol database……	37
3.6	Example of forest plots…….	41
3.7	Symmetrical plot in the absence of bias……...	42
3.8	Asymmetrical plot in the presence of reporting bias……	42
3.9	Asymmetrical plot in the presence of bias because some smaller studies (open circles) are of lower methodological quality and therefore produce exaggerated intervention effect estimates……..…	43
4.1	PRISMA flow chart of studies selection process……..…...	44

LIST OF FIGURES (cont.)

Figures		Page
4.2	Subgroup analysis blood pressure target estimates based on target levels	64
4.3	Subgroup analysis blood pressure target estimates based on study designs	65
4.4	Subgroup analysis blood pressure target estimates based on database settings…….……..……..	66
4.5	Subgroup analysis blood pressure target estimates based on diseases	68
4.6	Funnel plot for blood pressure target achievements…….……	69
4.7	Cholesterol target estimates subgroup analysis based on study designs	71
4.8	Cholesterol target estimates subgroup analysis based on database settings…….……..…..…	73
4.9	Subgroup analysis of blood press recorded estimates based on study designs…….……..……..	76
4.10	Subgroup analysis of BP recorded estimates based on database setting	77
4.11	Subgroup analysis of blood pressure recorded estimates based on diseases	78
4.12	Subgroup analysis of cholesterol recorded estimates accessibility based on study designs…….…….	81
4.13	Subgroup analysis of cholesterol record estimates based on database setting…….……..……	83
4.14	The percentage change post-P4P implementation (y-axis) when pre-P4P implementation were baselines (x-axis)…...……..……....	84
4.15	Percentage of diabetes patients achieving quality targets……..…..	85

CHAPTER I
INTRODUCTION

1.1 Background and Rationale

Since the downward economic cycle in the 1970s, it has been widely realized that health care systems in most countries have been inefficiently performed. Resources have been used inappropriately, and the health services are less productive while the cost of health service is increasing at an improper rate [1]. The governments cannot afford to pay for necessary health benefits, and out-of-pocket payments increase. Empirical evidence reported that the obsolescent traditional payment for health care providers was a reason of health care expenditure growth [2]. For example, in a fee-for-service (FFS) system, providers were paid for their volume of services and regardless of health care quality. As a result, many policy makers in various countries have been searching for the solutions that could encourage the quality of care while increasing accessibility and keep the growth rate of health care cost in control. Recently, there emerged a new strategy the pay-for-performance (P4P), which has been implemented globally [3].

P4P incentive system has become well known since the 1990s as a promising incentive scheme for improving quality and efficiency in service achievements of individual personnel and organizations in the health care system. The P4P implementation has covered a wide range of health services from ambulatory and inpatient care to different therapeutic treatments and prescribing. This program offers remunerations or cash bonus to health care providers if they can achieve their performances to some quality indicators. In other words, no performance payment will be made for no result [4]. The P4P has also been introduced to health care systems with universal health coverage (UHC) in many countries such as the United Kingdom, Netherland, Australia, Canada, Japan, Spain and Italy [5]. Including Thailand, various health care organizations have been used alongside the P4P implementations for increasing accessibility and improving quality of clinical health care.

Thailand health care systems can be classified as three health insurance schemes according to different groups of population to which the scheme being applied: (1) The Civil Servants Medical Benefits Scheme (CSMBS) which covers all civil servants and their eligible family members; (2) The Social Security Scheme (SSS) which covers employees who have worked in the formal employment sectors, and (3) Universal Coverage Scheme (UCS) which the Thai government has implemented since 2001. This scheme expands the coverage to the uninsured approximately 18 million people. The UCS is totally financed by general government revenue. The beneficial packages include: both outpatient and inpatient services, dental health service, rehabilitation, health care promotion, disease prevention and medicines. The UCS is expected to cover 75% of the total Thai population with the rest 25% covered by CSMBS and SSS [6].

However, there are a lot of insured people in distance rural provinces still inaccessible to the beneficial health care packages because of the regional disparities and the inadequacy of the health care providers, especially the physicians [7]. In the early stage, the Thai physician reimbursement was the mixture of salary, fee-for-service, capitation, case rates (including diagnosis-related groups), added with the hardship allowance for working in isolated rural areas according to the scheme to enhance motivation and retention of physicians working in rural community hospitals. In 2009, the hardship allowance paid for health care providers was increased, but the rise of this allowance could not lift the service productivity in health care service accordingly during the number of physicians in the hardship areas' increases [8]. It has recognized that the health financing policy has an impact on UHC achievement if the government takes strong and informative roles in regulating, providing and financing the health systems [9]. To this end, in 2013, Thai government launched pay-for-performance (P4P) incentive strategy in order to expand health care productivity, and increase both accessibility and quality of care in the health system in Universal Coverage context, despite the uncertainty about effectiveness of this strategy and a wide range of variations in the results of its implementations due to the different associated factors.

In previous systematic reviews, the following accounts on P4P have been reported: Houle et al. [10] conducted a systematic review to evaluate the effect of P4P

remuneration for targeting health care providers. Most of the studies included were observational study designs. They arrived at a conclusion that the effectiveness of P4P was uncertain. Petersen et al. [11] carried out a systematical analysis to assess the effect of P4P for quality improvement of health care measures. They suggested unclearly that evaluating and monitoring activities were among the important keys for determining the positive and negative effects of P4P on the health care quality. Schatz [12] performed a systematical review to summarize the P4P studies in ambulatory settings at the individual physician and the group physician levels. The study suggested that P4P associated with the quality improvement (but not always) and there were a limited number of good P4P designs. Bruin et al. [13] identified the association of P4P schemes and the implementation of disease management. They described that a number of the P4P incentive programs lacked, and hardly revealed, the effects of P4P on health care quality. Eijkenaar et al. [14] synthesized a systematic review of systematic reviews to arrive at a conclusion that the data on P4P effectiveness, in an overview sense, scattered widely, and more studies were needed.

Former systematic reviews on P4P paid less attention to different health system contexts. McDonald R. and Roland M. [15] suggested that the results of P4P scheme in many countries varied substantially according to the heterogeneous contextual factors and socio-economic conditions. Besides, Herck et al. [16] suggested that the effectiveness of P4P was unpredictable, which could be any of the following forms: positive effects, partial effects, adverse effects and no effect. The successful opportunity of P4P program depended on the design decision and the context of the health system.

Currently, there has been more evidence of the P4P introducing to health care systems, but there still exists some unclear results. Therefore, health care systems, in the context of universal health coverage, concerning Thailand should be cautious about the implementation of P4P incentive scheme.

The purpose of this present review is to clarify the direct association of pay-for-performance (P4P) effectiveness and the achievement of the health care system, in UHC context, in terms of the quality improvement and the accessibility of care.

1.2 Research Questions

1. Does pay-for-performance (P4P) affect the quality improvement in the health care system based on universal health coverage context (UHC)?

2. Does pay-for-performance (P4P) affect the accessibility of care in the health care system based on universal health coverage context (UHC)?

1.3 Research Objectives

1. The primary objective is to summarize the effectiveness of P4P incentive scheme on the quality improvement and the accessibility of care.

2. The secondary objective is to elucidate the conditions associated with the achievement of P4P implementation.

1.4 Conceptual Framework

In Figure 1.1 is shown the conceptual framework of the effectiveness of pay-for-performance (P4P) on the health care systems with universal health coverage in terms of: (1) quality improvement and (2) accessibility of clinical care.

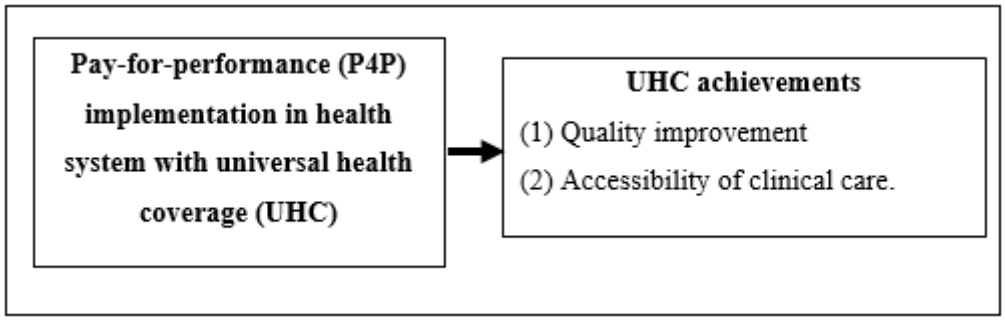


Figure 1.1 Conceptual framework of the effects of pay-for-performance (P4P) payment implementation for the universal health coverage (UHC) achievements.

1.5 Operational Definitions

Universal health coverage (UHC):

Health care systems that exist a legal mandate for provide the same health protection more than 90% of citizens with equal accessibility to the quality health services regardless of individual status and ability to pay, this definition followed the political economy of universal health coverage: Background paper for the global symposium on the health systems research. World Health Organization; 2010 [17].

Pay-for-performance (P4P):

The incentive payment scheme which focuses on the financial incentive that is remunerated to health providers at individual and organization levels when they can achieve quality target measures.

The effectiveness of P4P:

This study focuses the effectiveness of P4P in terms of the quality improvement and the accessibility of care.

The quality improvement:

The improvement of health care referred to the performance measurement of specified quality target indicators in the clinical care services.

The accessibility of clinical care:

A number of individuals accessible to the clinical care treatments and health services that have been recorded.

CHAPTER II
LITERATURE REVIEW

The following literature review is presented in three parts as follows:

2.1 Universal health coverage (UHC)

2.2 Pay-for-performance (P4P)

2.3 Systematic review

2.1 Universal Health coverage (UHC)

By the definition of universal health coverage, all citizens have the same financial protection with equal access to the quality health services, regardless of status or ability to pay. This strategic financial system has been implemented differently according to the socio-economic and political contexts of the countries. The World Health Organization (WHO) recommended that each country should prioritize four key actions to finance UHC: (1) reduce direct payments (out of pocket), (2) maximize mandatory pre-payment, (3) establish broad risk pools, and (4) use general government revenue to cover underprivileged people. The UHC achievement means all people have access to quality health care services such as prevention, promotion, treatment and rehabilitation [18].

2.1.1 Three broad dimensions for UHC

As there had been many countries declaring their commitment to achieving universal health coverage (UHC), the World Health Report [19] suggested the following three characterized dimensions of UHC:

1) The health services are needed (service coverage).

2) The total costs covered through insurance or other risk pooling mechanisms (financial protection).

3) The people who need health care services (population coverage).

Figure 2.1 shows the relationships among these principal dimensions. Population coverage represents the breadth of the health system, while the depth and the height referred to service coverage and financial protection respectively. It can be seen that with a certain amount of pooled funds, a change in any one or more dimensions will affect the others.

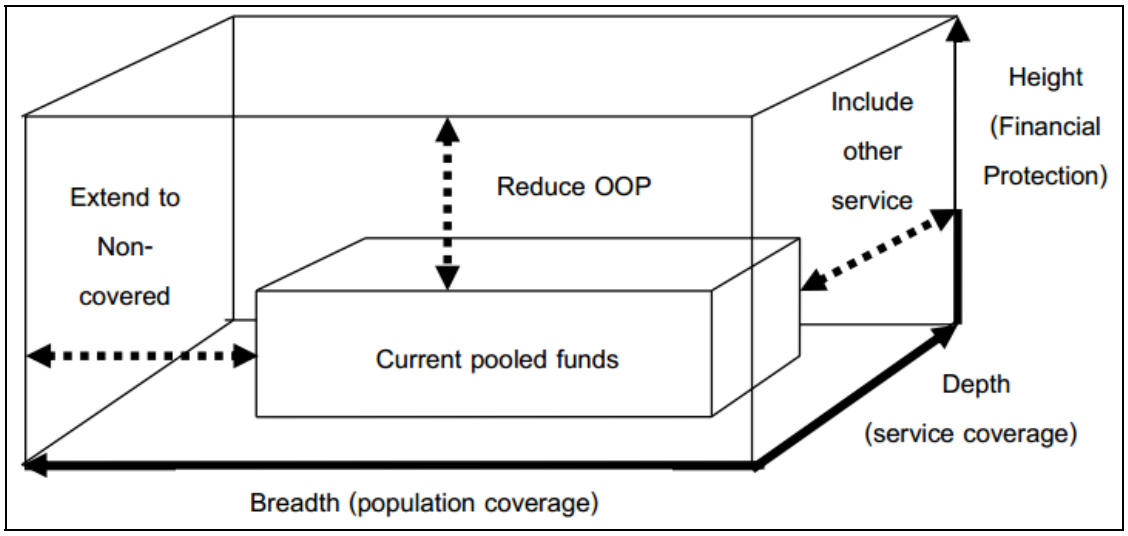


Figure 2.1 Three broad dimensions of universal health coverage (UHC) [19].

These three dimensions of UHC guide each country to reform their health financing systems towards universal health coverage. The health care systems intend to set up policy alternatives concerning benefits to be gained from eliminating critical health care problems they are facing. The benefits from applying the three principal dimensions are the results of their combinations in such a way that are the best fit for the organizational targets, political objectives and the health care expenditures.

2.1.2 The commitments of the UHC system

The UHC system usually commits to focusing on the following purposes: to increase accessibility, quality improvement, cost-containment. These are briefly discussed next:

Accessibility is to the ability to go through and receive services within any health care setting. The health care service accessibility depends on geographical distribution, financial, organizational, quality, and socio-cultural barriers; therefore, an accessibility measurement in terms of service utilization is dependent on the physical availability, affordability and acceptability of services [20].

Quality improvement is a set of procedures for increase quality of health care services. All health care system components are to perform activities following the standard of clinical practices in order to achieve the quality improvement. The resultant findings from the procedure or the process will be analysed to encourage the quality of care for the future treatments.

The quality improvement for health care consisted of six essential elements; effective, efficient, accessible, patient-centred, equitable, and safe. [21].

- **Effective**: providing services that may or may not save costs, but are still worth doing by reducing costs and improving outcomes concurrently.

- **Efficient**: decreasing wastes of equipment, supplies, ideas, and energy.

- **Accessible**: timely delivering and providing health care in a geographical setting where accurate skills and appropriate resources are needed.

- **Patient-centred**: providing care to meet individual patients' preferences, needs, and values.

- **Equitable**: providing same quality care to all individuals without any discrimination in the difference of gender, ethnicity, geographical location, or socioeconomic status.

- **Safe**: relieving injured patients carefully with no more harm.

Cost-containment is a scheme to control or reduce inappropriateness of the expenditures on health care service, which tends to be higher than necessary. The right amount of money should be wisely allocated for health care and should have an impact on the reasonable cost of care to serve [22]. There have

been several studies suggested that the use of different contemporary payment schemes can affect the expenses on health care; therefore, a regular commitment is necessary to achieve the efficiency of these instruments [23, 24].

2.2 Pay-for-performance (P4P)

Pay-for-performance (P4P) is the reward incentives are paid to the eligible beneficiaries who can achieve the specified quality agreements. The recipients included health service providers and the entities responsible for health care in various regions.

2.2.1 Pay-for-performance (P4P) concept

Figure 2.2 depicts the general concept of P4P. This financial scheme has attracted much global attention as a strategy to achieve public health care results.

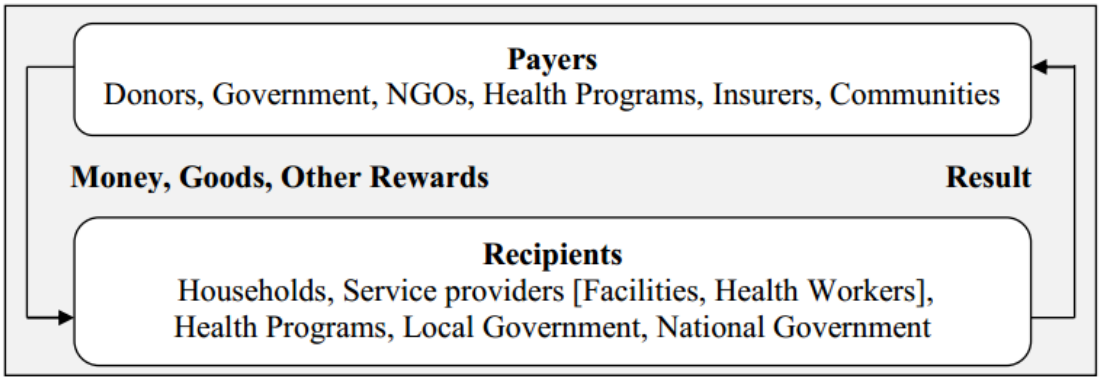


Figure 2.2 P4P concept defined by United States Agency for International Development (USAID) [25].

The physicians in developing countries have been remunerated for their services with salaries. Fixed salaries do not relate to performance; they may not encourage the accessibility of quality care services. In addition, this payment tends to offer low productivity, more absenteeism, inappropriate quality, or less of innovation. These are the main reasons that make P4P become a hopeful key solution [26].

In developed countries, P4P is mainly used for quality improvements. The quality measures can be in terms of structure, process and outcome. The explicit measurements are essential to the achievement of clinical outcomes [27].

2.2.2 Pay-for-performance (P4P) definition

P4P has been defined differently by different organizations. Some of the definitions presented in Table 2.1.

Table 2.1 Pay-for-performance (P4P) definition

Organization	P4P definitions
Agency for Healthcare Research and Quality (AHRQ) [28].	Better quality performance should gain more compensation.
Centers for Medicare & Medicaid Services (CMS) [29].	The payment strategies and other incentives tend to encourage patient-focused and improve high-value quality of care.
WHO [3].	P4P aims to maintain health care efficiency; dynamic, allocation, technology and equity.
United States Agency for International Development (USAID) [25].	P4P incentive rewards for the positive health care result in attainments.
World Bank [30].	They are various incentive payment mechanisms to enhance the performance of the health care system.

In 2003, the P4P movement started in the United States; the first two definitions are from American perspectives, and both focus on quality improvement. The following three definitions take a broader approach and are more regarded with developing countries. These latter three definitions are included both incentives on the supply side to health care providers and on demand-side to the patient, and also concerned about accessibility to health care services. The P4P has been reported to be implemented with different purposes in developed countries and developing countries. In developed countries, they require to improve the quality of health care services

while in developing countries, they always need to improve the equity of access to health care services [31].

2.2.3 P4P implementation in health care system

Although P4P systems can have various designs, they contain similar common elements as shown in Figure 2.3, which include: (1) the measures, (2) the basis for reward and (3) the reward.

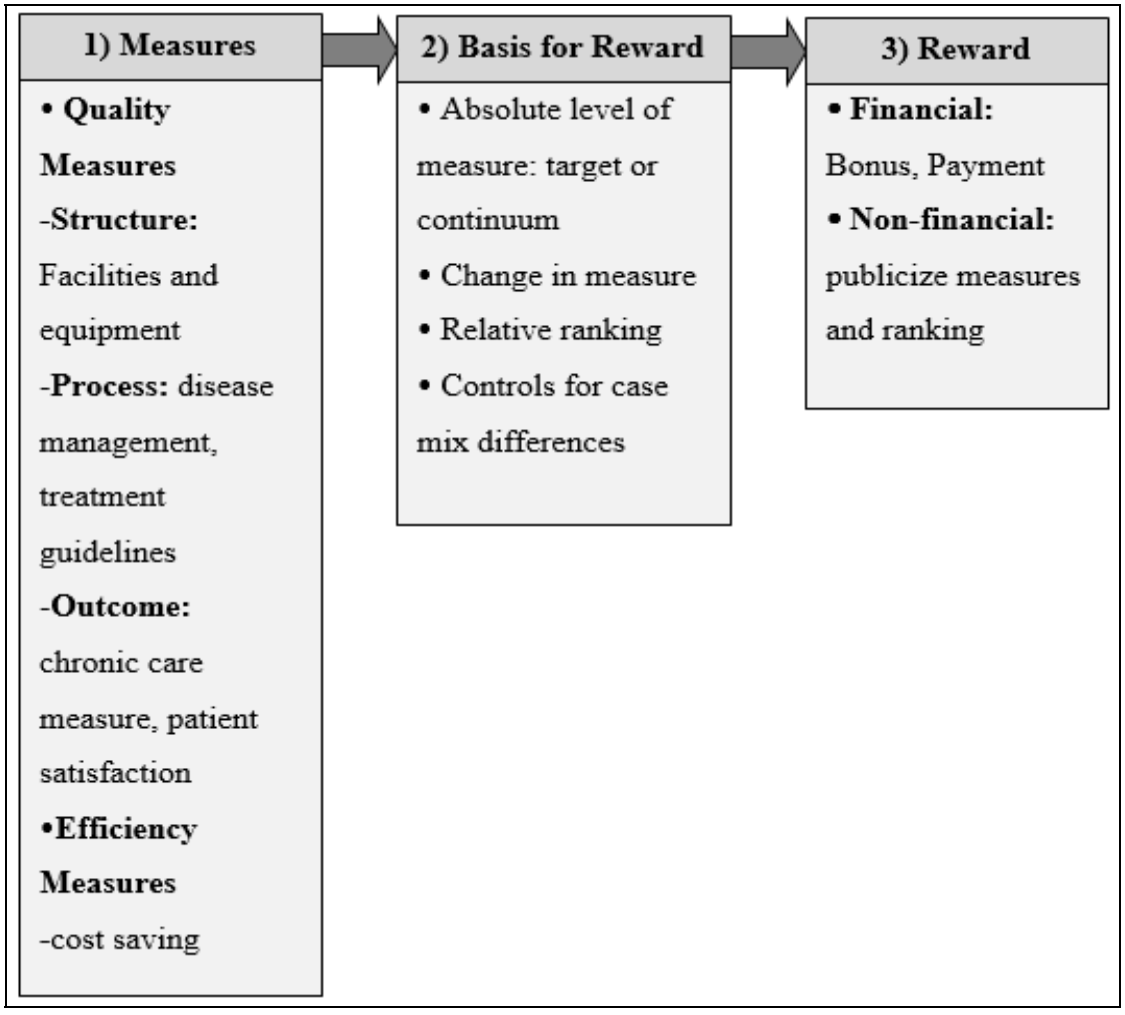


Figure 2.3 Structure of pay-for-performance (P4P) [3]

The first box is the measurement of quality and efficiency. The quality measures consist of three components: structure, process, and outcomes. The structure measures refer to the health care setting, for example, the facility, information

technology, infrastructures, equipment, supplies, pharmaceuticals and human resources. The process measures are regarded as the processes for providing health care services, which include clinical guidelines and recommendations, disease management protocols, specific target indicators and screening rates. The outcome measures are the ultimate results that include mortality or morbidity; they are very difficult to measure. The efficiency measures attend to costs.

The second box is referred to the reasons to pay; the absolute level of measure is paid for continuous quality care completion or for periodic target improvement. Change in the measure is the change in the specified target level. Relative ranking incentive is paid for reaching a given comparable ranking between the providers. Some health care providers have to concern with case mix differences, which requires various controls for the payment.

The third box is the reward incentive; composed of financial or non-financial and the combination of both. Financial compensation is referred to the remuneration payment to the providers for achieving the specified targets. Non-financial reward is referred to the reward that is not in the monetary form and requires a public announcement based on different performance levels.

2.2.4 The effectiveness of P4P

There was a lot of evidence that related to the effectiveness of pay-for-performance (P4P) on the health care system based on universal health coverage (UHC) context. The empirical relevant studies were categorized in four groups; positive effects (Table 2.2), partial effects (Table 2.3), adverse effects (Table 2.4) and no effects (Table 2.5). Most of them were positive and partial effects as the following details.

Table 2.2 The positive effects of pay-for-performance (P4P) incentive in healthcare systems with universal health coverage

Ref/Setting	Positive effects of P4P incentives
Karunaratne K., 2013 [32], the United Kingdom (UK)	The proportion of patients with Chronic kidney disease stage 3-5(CKD3-5) attaining the BP target of 145/80 achieved from 41.5% in the Pre-P4P period to 50% in the post-P4P period.
Simpson C, 2011[33], the UK	The proportion of hypertensive patients controlled blood pressure (<140/90 mmHg) increase from 22% in 2003 to 28% in 2006.
Doran T, 2011 [34], the UK	In the first year of P4P incentive, there were twenty-two significant improvements but remained fourteen in the second year.
Vamos E, 2011 [35], the UK	The smallest and largest practices size, have no significant differences to the blood pressure targets achievement. P4P is appropriate for both large and small practices size to the same positive results.
Cheng S,2012 [36], Taiwan	The tests increased significantly in the P4P program; the improvement declined but still significant.
MacBride-Stewart S, 2007 [37], the UK	The prescribing of relevant drugs increased before the P4P implementation in 2004; the growth existed in the first two years but in the third year declined but still improvement
Doran T, 2008 [38], the UK	The study showed the pay-for-performance programs were useful, providing that the processes were suitably and less adverse effects.
Doran T, 2008 [39], the UK	The financial incentive schemes possible to decrease the inequalities of clinical care accessibility relate to area deprivation.
McGovern M, 2008 [40], the UK	In the first year of P4P period, the patients with diabetes who have an electronic record increased 54.2%. Moreover, the recorded of blood pressure, glycated hemoglobin (HbA1c), cholesterol and serum creatinine improve significantly $p<0.05$.

Table 2.2 The positive effects of pay-for-performance (P4P) incentive in healthcare systems with universal health coverage. (cont.)

Ref/Setting	Positive effects of P4P incentives
Gulliford M, 2007[41], The UK	The proportion of patients who achieving HbA1c ≤ 7.4% improved from 38% pre-P4P to 57% post-P4P implementation.
Millett C, 2007 [42], the UK	The diabetes patients, who were noted a smoking status recorded, substantially improved from 90% in the non-P4P period to 98.8% in P4P period, and the smoking-cessation advice recorded also increased meaningfully from 48.0% to 83.5%, and the prevalence of smoking reduced from 20.0% to 16.2% ($p < 0.001$).
Steel N, 2007 [43], the UK	The six incentive payment indicators significantly achieved from 75% in pre-P4P to 91% in post-P4P with the %change difference = 16% (95%CI = 10 to 22%, $p < 0.01$). The fifteen 'non-incentive indicators' also increased from 53% to 64% with the %change difference = 11% (95%CI = 6 to 15%, $p < 0.01$). The financial incentives also enhanced the quality improvement in the non-incentive conditions.
Millett C, 2007 [44], the UK	After P4P implementation, the proportion of patients in ethnic groups achieving treatment targets for total cholesterol, BP and HbA1c significantly increased.
Scott A, 2013 [45], Australia	In the patients with diabetes, there was potential association of P4P incentives and the guideline processes measurements.
Li Y, 2009 [46], Taiwan	After P4P implementation, there were substantially improvements of the average length of treatment and cure rate on TB program in Taiwan. Besides, the treatment in P4P hospital had better outcomes than the non-P4P hospital considerably.
Kuo R, 2011 [47], Taiwan	After P4P implementation, there were importantly better quality of care on BC-P4P program in Taiwan. Besides, the patients in BC-P4P program had a better outcome significantly.

Table 2.2 The positive effects of pay-for-performance (P4P) incentive in healthcare systems with universal health coverage. (cont.)

Ref/Setting	Positive effects of P4P incentives
Tsai W, 2010 [48], Taiwan	The treatment default rate in the patients with tuberculosis TB decrease from 15.56% before TB-P4P program to 11.37% after TB-P4P program. In addition, the treatment default rate in the patients with TB decrease from 12.7% in non-P4P hospitals to 10.67% in P4P hospitals. The conclusion is P4P program encourage the improvement of the default rate in TB patients.
Lai C, 2013 [49], Taiwan	The guideline-recommended tests/examinations increased significantly in the DM-P4P program for patients with diabetes.
Kontopantelis E, 2013 [50], the UK	Quality of health care recorded increased in the P4P period for all subgroups. In the first year of P4P implementation, the quality grows more than in the non-P4P period. In the third year of P4P incentive, the quality improvement slope was declined but remained significantly improvement.
Colais P, 2013 [51], Italy	The patients with hip fracture that had surgery within 48 hours increased from 11.7% in non-P4P period to 22.2% in P4P period. Besides, the P4P implementation increased the proportion of early hip fracture operations.
Rubinstein A, 2009 [52], Argentina	A multimodal intervention based on pay-for-performance encourage a significant improvement in most clinical effectiveness indicators.
Campbell S, 2009 [53], the UK	In the first two year of the P4P implementation, the quality of care in the patients with asthma and diabetes increased but not for the heart disease patients. However, the quality care improvement in all three conditions was increased in the fourth year.

Table 2.2 The positive effects of pay-for-performance (P4P) incentive in healthcare systems with universal health coverage. (cont.)

Ref/Setting	Positive effects of P4P incentives
Millett C, 2009 [54], the UK	P4P program encouraged the improvement of the intermediate outcome target measures of diabetes in all ethnics, but the results were different.

Table 2.3 The partial effects of pay-for-performance (P4P) incentive in healthcare systems with universal health coverage

Ref/Setting	The partial effects of P4P incentives
Lee J, 2011 [55], the UK	P4P program significantly increasing the improvement of blood pressure control in all ethnics groups in a short period, but there were various results. The reductions of systolic BP of observed in Black and White but not in South Asian.
Calvert M, 2009 [56], the UK	After P4P had been introduced, there were notably increased in intermediate outcome measures and process measures. However, there was no improvement in diabetes patients with >10% HbA1c levels.
McGovern M, 2008 [57], the UK	In P4P period, there were importantly increased of quality indicators recorded and prescribing in patients with Coronary Heart Disease (CHD), but men had more recorded than women.
Hamilton F, 2010 [58], the UK	After P4P implementation, there was the improvement of some intermediate outcome in diabetes; blood pressure, cholesterol, but not in HbA1c domain. Besides, older patients gained more benefits than younger patients.
Simpson CR, 2006 [59], the UK	The quality indicators recorded in the patients with stroke increased from 32.3% in non-P4P period to 52.1%. In P4P period. However, the P4P period the most affluent and the oldest patients (>75 years) tended to have more recording.

Table 2.3 The partial effects of pay-for-performance (P4P) incentive in healthcare systems with universal health coverage. (cont.)

Ref/Setting	The partial effects of P4P incentives
Crawley D, 2009 [60], the UK	In P4P period, there were no significant differences increased of the BP targets achievement in diabetes similar to hypertensive individuals. However, the difference in target BP achievement in CHD between the occupational groups was detected.
Sutton M, 2012 [61], the UK	In P4P period, there were significantly decreased of total mortality for the conditions with an absolute reduction of 1.3 percentage points and a relative reduction of 6%. There was greater decreasing in the patient with pneumonia, but no notably reductions in the heart diseases.
Kirschner K, 2013 [62], the UK	After P4P had been introduced, there was the considerable improvement of the process quality indicators in all chronic conditions. Positive Results in the cardiovascular risk and the asthma, +7.9% and +11.5% respectively. No increasing rate in cervical cancer screening and influenza vaccination. Better improvement in smaller practices.
Campbell S, 2008 [63], the UK	After P4P implementation, there was the increasing of quality target achievement and provider income. However, the relationship between health care provider and the patients tended to change.
Alshamsan R, 2012 [64], the UK	After P4P had introduced, there was the systolic blood pressure improvement in the black and the white patients, but the increasing levels maintained in the black patients only. There was no different on HbA1c in all ethnic groups.
Lee T, 2010 [65], Taiwan	The guideline-recommended tests/examinations recorded increased in the P4P program, but the overall cost of care management was substantially higher.

Table 2.3 The partial effects of pay-for-performance (P4P) incentive in healthcare systems with universal health coverage. (cont.)

Ref/Setting	The partial effects of P4P incentives
Kiran T, 2009 [66], Canada	There was minimal quality improvement of diabetes care at the patient and population level. In addition, the overall cost of care management was considerably higher.

Table 2.4 The adverse effects of pay-for-performance (P4P) incentive in healthcare systems with universal health coverage

Ref/Setting	Adverse effects of P4P incentives
Tsung-Tai C, 2011 [67], Taiwan	After P4P implementation, the older patients with more complex severe conditions or further comorbidity were tended to be excluded from the programs.
Dalton A, 2011 [68], the UK	The South Asian and the Black patients were prone to be excluded from the HbA1c quality indicator than the White patients.
Doran T, 2006 [69], the UK	A few English family practices which had high scores, excluded patients from the P4Pprogram for more income.

Table 2.5 No effects of pay-for-performance (P4P) incentive in healthcare systems with universal health coverage

Ref/Setting	No effects of P4P incentives
Inoue Y, 2011 [70], Japan	There was no different on the return home rate in the Functional Independence Measure (FIM) between Before and after P4P implementation.
Serumaga B, 2011 [71], the UK	There was no effect of P4P on the incidence of myocardial infarction, stroke, renal failure and heart failure. Besides, P4P was not related to mortality in both newly treated subgroups and treatment experienced.

2.3 Systematic review

A systematic review is defined by the Centre for Reviews and Dissemination (CRD) guidance as " review of the evidence on a clearly formulated question that uses systematic and explicit methods to identify, select and critically appraise relevant primary research, and to extract and analyze data from the studies that are include in the review" [72].

The flow chart of a systematic review is presented in Figure 2.4

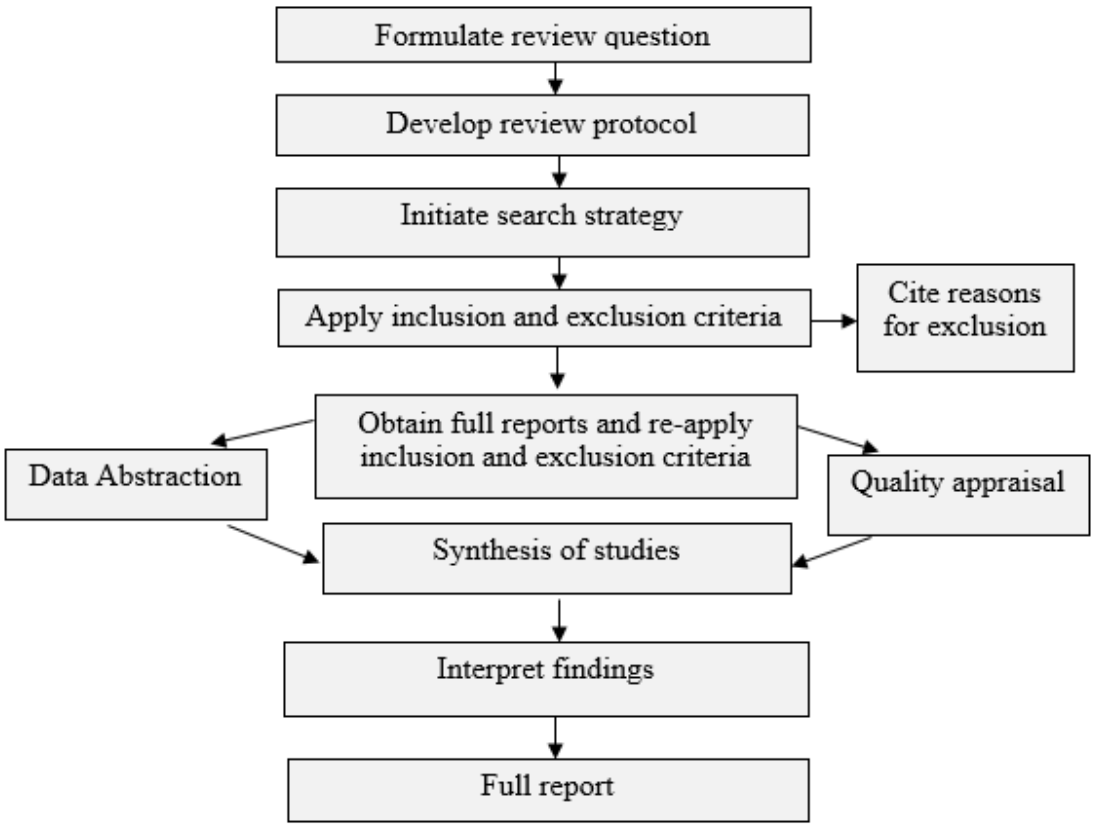


Figure 2.4 Flow chart of a systematic review process (adapted from the Systematic Reviews of Health Promotion and Public Health Interventions: train-the-trainer handbook, Cochrane Health Promotion and Public Health Field) [73].

2.3.1 Processes of a systematic review: The brief description according to the flow chart of a systematic review as presented in Figure 2.4.

- Formulate review question:

The Cochrane handbook [74] suggested that, to perform a systematical analysis have to start with precise review questions. The answers of the explicit questions will provide more appropriate information for decision-making guidance. The review questions can be defined in terms of the participants of population (P), intervention (I), comparator (C) and outcome (O). These components of the review questions, together with the methodological study designs, will be used for define the inclusion criteria of studies selection.

Four components of the review questions were the following;

Population: The people affected by the intervention.

Intervention: The activity of process.

Comparison: The other interventions or the based-line.

Outcome: The ultimate results of health care treatment such as mortality and morbidity.

- Development of a review protocol: The review protocol is a set of methods that the reviewer determined to be used in the review. There were the descriptions about the review questions, inclusion criteria, exclusion criteria, databases, search strategy, study selection, data extraction, quality assessment, data synthesis and statistic methodologies.

- Initiate search strategy: In a systematic review, the reviewer should conduct the relevant studies through the validated electronic databases for decreasing the risk of bias in the review procedure. The searching process should be obviously and recorded. The studies can be searched from a combination of the following method; hand searching, evidence literatures searching, contacting researchers, Internet resources searching.

- Selection of studies: The first step of the selection, all studies from electronic databases have been screened with their titles and abstracts. The studies that related to the inclusion criteria will be added into the review, but the others will be excluded. The second step, more details from the full-text of all relevant studies are necessary to evaluate the eligibility.

- **Cite reasons for exclusion:** After the first step of the selection, some studies were excluded by the seasons of their irrelevant titles and abstracts, the causes of exclusion will be noted.

- **Data Abstraction:** Abstracting is the procedure of identifying and attending the prominent characteristics of the primary studies and removes the inappropriate studies design. Different study designs require distinctive data abstraction forms.

- **Study quality appraisal:** Quality appraisal or quality assessment of any systematical review is a step to category the quality of the included studies. Quality assessment form is a tool to reveal the quality of the eligible studies in two domains; the appropriateness of study design to the review purposes and the risk of bias of the studies.

***Appropriateness of study design*:** Types of study used to assess the effects of interventions can be arranged into the following hierarchy.

- Randomized controlled trials (RCT): RCT is an experimental design that randomize participants. It categorizes the eligible into two (intervention group and control group) or more groups. Each cluster has been introduced the interventions according to the study method with a blind technic, and the results are the comparison of outcome findings between groups.

- Non-randomized studies (NRS): NRS is a study that does not use randomization methodologies to the comparison groups. There are many types of NRS as the following;

• Nonrandomized controlled trial: An experimental study that not random participants in different interventions.

• Quasi-experimental studies: A study is likely to the RCT, but it does not complete a blind technic manipulation.

• Controlled before and after study (CBA): A comparison of the control groups, before and after of the interesting interventions.

• Interrupted time series (ITS): An interrupted time series (ITS) or time series is a series of many observational studies. The same participants or different but likely participants can be observed the trend of intervention outcomes over the interested time-period.

• Cohort study: A study that carries out a group of people over time; a prospective cohort study included participant before any intervention and follows them to the future, a retrospective cohort study allocated participants from previous interested records and follows them.

• Case-control study: A study compares the people with interested outcomes (cases) with the people from the same of population without that outcome (control). In order to investigate the association between the outcome and before an intervention implementation.

• Cross-sectional study: A study that collects interested data at one particular point in time.

The RCT is the best study design to investigate the effect of an intervention. Despite, there is a limited number of public health studies that carried out in RCT design.

The Cochrane handbook [74] suggested that the reviewer may use Non-randomized studies (NRS) when the question of interest cannot be replied by a review of randomized trials.

- There three principle reasons for including NRS in a review.

(1) To evaluate the case for support a randomized trial.

(2) To provide evidence of effects (benefit or harm) of interventions that cannot be randomized

(3) To provide evidence of effects (benefit or harm) that cannot be sufficiently conducted in randomized trials.

- Risk of bias: Bias in a systematic review which comes from the poor quality of the included studies; that may mislead the review from the true effect of interventions. There are eight types of bias, categorized by the Cochrane Health Promotion and Public Health Field's handbook [73] as the following;

- Eight types of bias in health promotion and public health studies;

(1) Selection bias: The participants of the included study should be a true representative of the target population.

(2) Allocation bias: The intervention group and control group should have similar baseline characteristics.

(3) Confounding: A situation where the factors influenced the outcomes, and it related to both the intervention and the outcome.

(4) Blinding of outcome assessors (detection bias): The researcher should not know the status of the participants in both control group and intervention group.

(5) Data collection methods: The primary data should be collected from valid and reliable instruments.

(6) Withdrawals and drop-out (attrition bias): The difference number of the intervention and control groups that drop-out from the study before the end of the intervention.

(7) Intervention integrity: If there were the ineffectiveness results or outcome findings, the researcher should be declared.

(8) Statistical analysis: The appropriate and potential statistical methodologies have the ability to detect the difference between the groups of comparison.

- Tools for assessing risk of bias in non-randomized studies.

The Cochrane handbook [74] suggested that the most useful quality assessment for non-randomized studies (NRS) are the Downs and Black instrument and the Newcastle-Ottawa Scale.

• The Downs and Black instrument was developed by Sara H Downs and Nick Black in 1998 [75]. This tool can be used for evaluate the methodological quality both of randomized and non-randomized studies of health care interventions [76].

• The Newcastle-Ottawa Scale is an ongoing collaboration between the Universities of Newcastle, Australia and Ottawa, Canada since 2008. It was developed to assess the quality of non-randomized studies [77].

- **Synthesis of studies**; there are two types of synthesis as the following**.**

Narrative synthesis

Narrative synthesis is an essential section of a systematic review. The general framework of a narrative synthesis suggested by the Cochrane handbook [74] consists of four compositions as the following; (1) Describe the theory of how the intervention works, why and for whom. (2) Developing a primary synthesis of the included studies effect findings. (3) Explanation of the association of the intervention within and across studies. (4) Evaluating the validity of the synthesis.

Meta-analysis

Meta-analysis is a methodology to reveal the effect of the interesting interventions. In a systematic review, there are the groups of studies that have met the inclusion criteria and quality design appropriated. The reviewer can combine the final similar included studies in order to potential the results.

• Principles of meta-analysis

There are four basic principles of meta-analysis:

(1) In the first step, a summary statistic calculated of each study for ascertained intervention effects.

(2) In the second step, a pool effect estimate is calculated as a weighted average. A weighted average is described as

$$\text{Weighted average} = \frac{\text{sum of (estimate x weight)}}{\text{Sum of weights}} = \frac{\sum YiWi}{\sum Wi}$$

Where *Yi* is the intervention effect estimated in the *i* th study, *Wi* is the weight given to the *i* th study, and the summation is across studies.

(3) The intervention effect estimates between studies are in two assumptions; random-effects model and fixed-effect model.

- Random-effects model is incorporating an assumption that the studies are in various estimate intervention effects, according to a distribution between studies.

- Fixed-effects model is incorporating an assumption, that each study is estimating the same intervention effect.

(4) A confidence interval is used to interpret the significant or insignificant of the pooled estimate, and a p-value communicated the robustness of the evidence versus the null hypothesis of no intervention effect.

• To synthesis the included studies, the reviewers have to select the type of analysis in the review. When there are difference studies with the diverse variables. The synthesis should be a narrative synthesis.

The homogenous studies should be performed a meta-analysis, as showed in Figure 2.5. The dissimilarity of the studies is "heterogeneity."

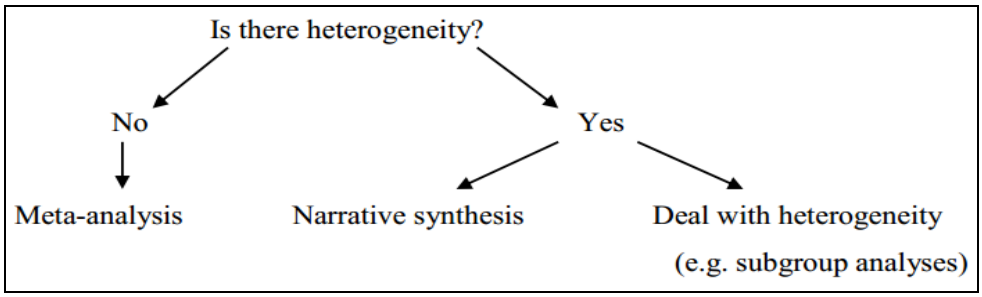


Figure 2.5 The heterogeneity and type of data synthesis

• Heterogeneity in meta-analysis; there are three type of heterogeneity as the following,

(1) Clinical diversity or clinical heterogeneity; dissimilarity in participants, sample size, process of interventions and ascertainment of outcomes between the studies.

(2) Methodological diversity or methodological heterogeneity; dissimilarity in study design and quality of the studies.

(3) Statistical heterogeneity; statistical heterogeneity is the consequence of clinical diversity and/or methodological diversity. There are dissimilarity in the intervention outcomes between the studies.

・Strategies for addressing heterogeneity

(1) Check the data

When the high heterogeneity has been detected, the data may incorrectly recorded into Review Manager (RevMan) [78].

(2) Do not do a meta-analysis; if there is similar in method but difference in results. The meta-analysis can be added into a systematic review. However, the high heterogeneity referred to the more consistency that may be leading to an under-estimated or over-estimated of the intervention effectiveness.

(3) Explore heterogeneity; if there existed the causes of heterogeneity across the studies or the findings. The reviewer should explore the heterogeneity by performing meta-regression or subgroup analysis.

(4) Ignore heterogeneity; if the reviewer determines that there are no different across the studies. Each study has the same intervention effect.

(5) Perform a random-effects meta-analysis; if the reviewer determines that there are different across the studies. Each study has not the same intervention effect and to estimate the intervention effects have to follow the distribution of effect across studies.

(6) Change the effect measure; the included studies always use the different methods. The effect measure is a tool to measure the pool estimate effect of the intervention. As mentioned, there are several effect measures in a meta-analysis. The inappropriate effect measure may affect the level of heterogeneity across the studies as well.

(7) Exclude studies; if the few of included studies have different results from the most studies in subgroups. The heterogeneity may be observed. The reviewer should remove those various results from the analysis.

・Subgroup analyses

Subgroup analysis is an analysis to explore heterogeneity. The reviewer group the similar studies in order to reveal the effects of the intervention. The intervention should affect study factors or patients the same way

in different variables. If the dissimilarity have been detected. They may be the cause of heterogeneity.

- Interpret findings:

The subsequent concerns should be comprised in the discussion and recommendations section of the public health of health promotion systematic review.

- Full report: To summarize the main findings on the effects of the interventions studied in the review in the following formats:

(1) Characteristics of included studies' tables (including 'Risk of bias' tables) present information on individual studies.

(2) Data and analyses (the full set of data tables)

(3) Figures (Forest plots, Funnel plots)

CHAPTER III
METHODOLOGY

3.1 Study design: Systematic review guidelines

This study is based on the two guidelines of systematic reviewing:

1) Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [74].

2) Systematic Reviews CRD's guidance for undertaking reviews in health care [72].

3.2 Scope of the study

This study is limited to the information from the literature being searched from electronic databases and the available health care publications in English from the year of 2000 to March 2014.

3.3 Study time period

The study was performed during June 2013 to June 2014.

3.4 Data sources

The following electronic databases were used for primary studies searching:

The Cochrane Library, PUBMED, MEDLINE, EBSCO host service, CINAHL

3.5 Search methods for identification of primary studies

3.5.1 Formulate review question.

The key components of the search strategy comprise of Medical Subject Heading (MeSH) terms and text words that describe each element of the PICO questions.

There are two research questions in this review;

Question 1: Does pay-for-performance (P4P) affect the quality improvement in the health care system based on universal health coverage (UHC)?

Question 2: Does pay-for-performance (P4P) affect the accessibility of care in the health care system based on universal health coverage (UHC)?

The PICO criteria and defined the question elements as;

Population: The health systems based on universal health coverage context (UHC).

Intervention: Pay-for-performance (P4P) implementation for health care providers.

Comparison: Before P4P period or Non-P4P period

Outcomes:

(1) The performance measurement of the quality improvement.

(2) The performance measurement of health care accessibility.

3.5.2 Search term

Search term were; quality of health care, utilization, health services accessibility, equity for care, quality effectiveness, quality improvement, access to care. They were combined with pay for performance, incentive awards, incentive-based compensation, reimbursement, incentive, physician incentive plans, health provider incentive, reward and bonus.

The following electronic databases were used for primary studies searching; The Cochrane Library, PUBMED, MEDLINE, EBSCO host service, CINAHL.

3.6 The full search strategies

From the Cochrane Library, PUBMED, MEDLINE, EBSCO host service, CINAHL as the following.

3.6.1 The Cochrane Library as presented in Figure 3.1

Search strategies

ID	Search	Hits
#1	MeSH descriptor: [Reimbursement, Incentive] explode all trees	(53)
#2	MeSH descriptor: [Physician Incentive Plans] explode all trees	(13)
#3	pay for performance	(824)
#4	#1 or #2 or #3	(870)
#5	MeSH descriptor: [Quality of Health Care] explode all trees and with qualifier(s): [Utilization - UT]	(246)
#6	equity for care	(231)
#7	quality effectiveness	(24833)
#8	MeSH descriptor: [Health Services Accessibility] explode all trees	(840)
#9	#5 or #6 or #7 or #8	(25836)
#10	#4 and #9	(615)

Search results: 615 text results

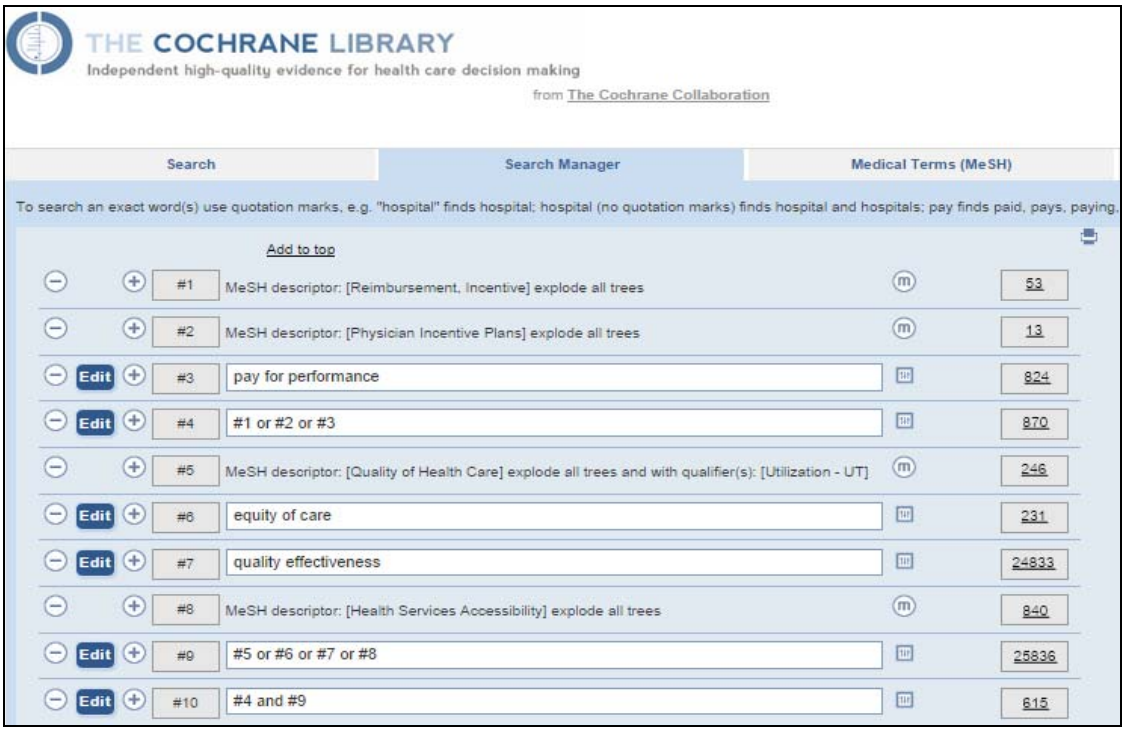


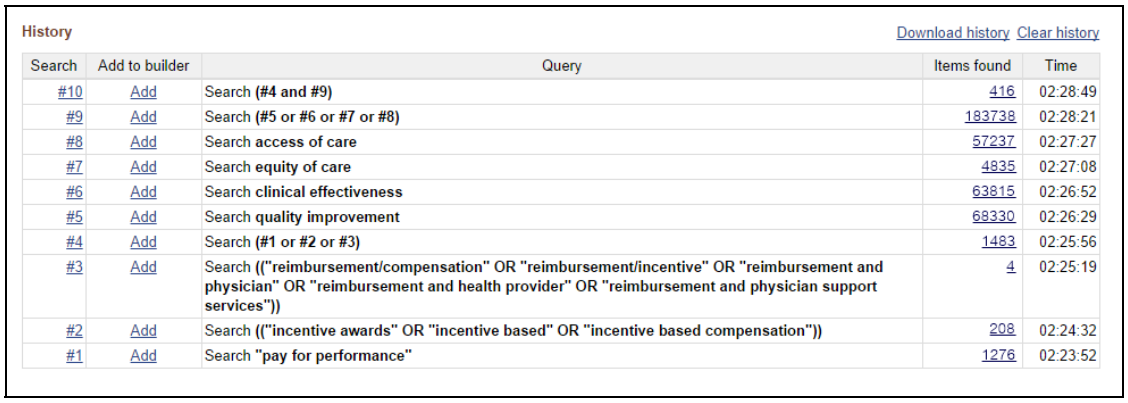
Figure 3.1 The search strategies from the Cochrane Library (www.thecochranelibrary.com)

3.6.2 PubMed as presented in Figure 3.2

Search strategies:

((("pay for performance") OR (("incentive awards" OR "incentive based" OR "incentive based compensation")) OR (("reimbursement/compensation" OR "reimbursement/incentive" OR "reimbursement and physician" OR "reimbursement and health provider" OR "reimbursement and physician support services"))) AND ((quality improvement) OR (clinical effectiveness) OR (equity of care) OR (access of care))

Search results: 416 text results



History

Download history Clear history

Search	Add to builder	Query	Items found	Time
#10	Add	Search **(#4 and #9)**	416	02:28:49
#9	Add	Search **(#5 or #6 or #7 or #8)**	183738	02:28:21
#8	Add	Search **access of care**	57237	02:27:27
#7	Add	Search **equity of care**	4835	02:27:08
#6	Add	Search **clinical effectiveness**	63815	02:26:52
#5	Add	Search **quality improvement**	68330	02:26:29
#4	Add	Search **(#1 or #2 or #3)**	1483	02:25:56
#3	Add	Search **(("reimbursement/compensation" OR "reimbursement/incentive" OR "reimbursement and physician" OR "reimbursement and health provider" OR "reimbursement and physician support services"))**	4	02:25:19
#2	Add	Search **(("incentive awards" OR "incentive based" OR "incentive based compensation"))**	208	02:24:32
#1	Add	Search **"pay for performance"**	1276	02:23:52

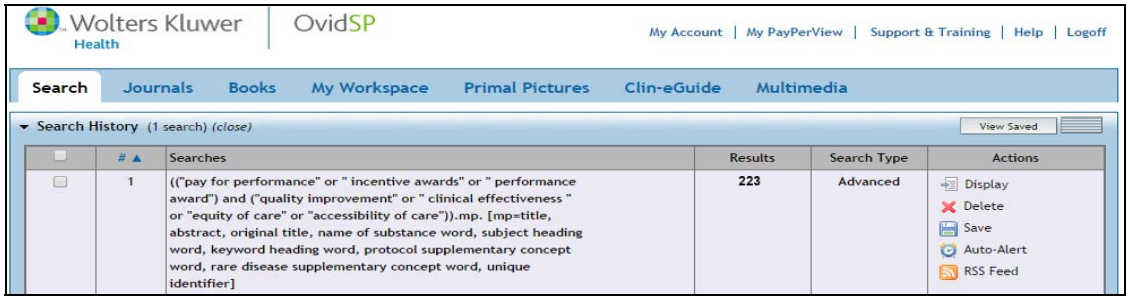
Figure 3.2 The search strategies from PubMed (http://www.ncbi.nlm.nih.gov/pubmed)

3.6.3 Ovid Medline 1946 to Present as presented in Figure 3.3

Search strategies:

((“pay for performance" or" incentive awards" or " performance award") and ("quality improvement" or " clinical effectiveness " or "equity of care" or "accessibility of care")).mp.

Search results: 223 text results



Wolters Kluwer Health | OvidSP

My Account | My PayPerView | Support & Training | Help | Logoff

Search | Journals | Books | My Workspace | Primal Pictures | Clin-eGuide | Multimedia

Search History (1 search) *(close)* View Saved

# ▲	Searches	Results	Search Type	Actions
1	(("pay for performance" or " incentive awards" or " performance award") and ("quality improvement" or " clinical effectiveness " or "equity of care" or "accessibility of care")).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier]	223	Advanced	Display Delete Save Auto-Alert RSS Feed

Figure 3.3 The search strategies from Ovid Medline accessed through Mahidol database

3.6.4 EBSCO Discovery Service as presented in Figure 3.4

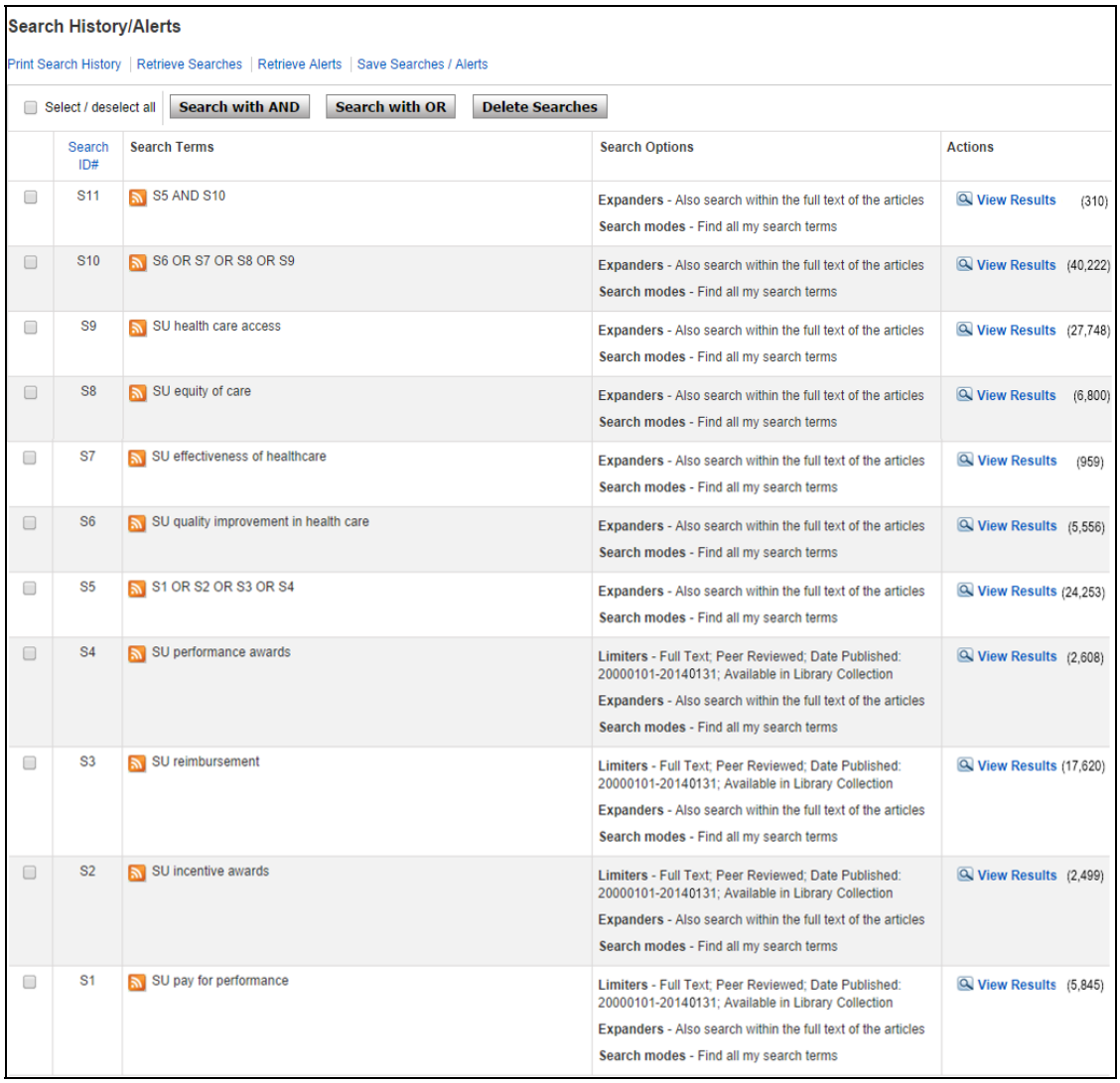
Search strategies:

#	Query	Limiters/ Expanders	Results
S11	S5 AND S10	Expanders - Also search within the full text of the articles Search modes - Find all my search terms	(310)

#	Query	Limiters/ Expanders	Results
S10	S6 OR S7 OR S8 OR S9	Expanders - Also search within the full text of the articles Search modes - Find all my search terms	(40,222)
S9	SU health care access	Expanders - Also search within the full text of the articles Search modes - Find all my search terms	(27,748)
S8	SU equity of care	Expanders - Also search within the full text of the articles Search modes - Find all my search terms	(6,800)
S7	SU effectiveness of healthcare	Expanders - Also search within the full text of the articles Search modes - Find all my search terms	(959)
S6	SU quality improvement in health care	Expanders - Also search within the full text of the articles Search modes - Find all my search terms	(5,556)
S5	S1 OR S2 OR S3 OR S4	Expanders - Also search within the full text of the articles Search modes - Find all my search terms	(24,253)
S4	SU performance awards	Limiters - Full Text; Peer Reviewed; Date Published: 20000101-20140131 Expanders - Apply related words; Also search within the full text of the articles Search modes - Find all my search terms	(2,608)

#	Query	Limiters/ Expanders	Results
S3	SU reimbursement	Limiters - Full Text; Peer Reviewed; Date Published: 20000101-20140131 Expanders - Apply related words; Also search within the full text of the articles Search modes - Find all my search terms	(17,620)
S2	SU incentive awards	Limiters - Full Text; Peer Reviewed; Date Published: 20000101-20140131 Expanders - Apply related words; Also search within the full text of the articles Search modes - Find all my search terms	(2,499)
S1	SU pay for performance	Limiters - Full Text; Peer Reviewed; Date Published: 20000101-20140131 Expanders - Apply related words; Also search within the full text of the articles Search modes - Find all my search terms	(5,845)

Search results: 310 text results



Search History/Alerts

Print Search History | Retrieve Searches | Retrieve Alerts | Save Searches / Alerts

Select / deselect all | Search with AND | Search with OR | Delete Searches

	Search ID#	Search Terms	Search Options	Actions
☐	S11	S5 AND S10	**Expanders** - Also search within the full text of the articles **Search modes** - Find all my search terms	View Results (310)
☐	S10	S6 OR S7 OR S8 OR S9	**Expanders** - Also search within the full text of the articles **Search modes** - Find all my search terms	View Results (40,222)
☐	S9	SU health care access	**Expanders** - Also search within the full text of the articles **Search modes** - Find all my search terms	View Results (27,748)
☐	S8	SU equity of care	**Expanders** - Also search within the full text of the articles **Search modes** - Find all my search terms	View Results (6,800)
☐	S7	SU effectiveness of healthcare	**Expanders** - Also search within the full text of the articles **Search modes** - Find all my search terms	View Results (959)
☐	S6	SU quality improvement in health care	**Expanders** - Also search within the full text of the articles **Search modes** - Find all my search terms	View Results (5,556)
☐	S5	S1 OR S2 OR S3 OR S4	**Expanders** - Also search within the full text of the articles **Search modes** - Find all my search terms	View Results (24,253)
☐	S4	SU performance awards	**Limiters** - Full Text; Peer Reviewed; Date Published: 20000101-20140131; Available in Library Collection **Expanders** - Also search within the full text of the articles **Search modes** - Find all my search terms	View Results (2,608)
☐	S3	SU reimbursement	**Limiters** - Full Text; Peer Reviewed; Date Published: 20000101-20140131; Available in Library Collection **Expanders** - Also search within the full text of the articles **Search modes** - Find all my search terms	View Results (17,620)
☐	S2	SU incentive awards	**Limiters** - Full Text; Peer Reviewed; Date Published: 20000101-20140131; Available in Library Collection **Expanders** - Also search within the full text of the articles **Search modes** - Find all my search terms	View Results (2,499)
☐	S1	SU pay for performance	**Limiters** - Full Text; Peer Reviewed; Date Published: 20000101-20140131; Available in Library Collection **Expanders** - Also search within the full text of the articles **Search modes** - Find all my search terms	View Results (5,845)

Figure 3.4 The search strategies from EBSCO Discovery Service accessed through Mahidol database

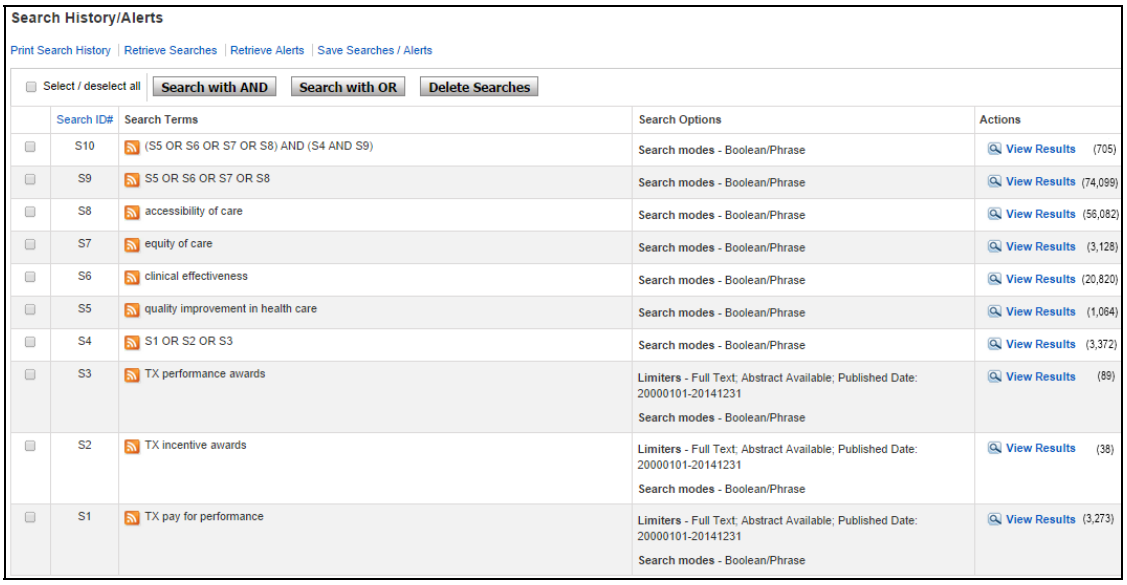
3.6.5 CINAHL Plus with Full Text as presented in Figure 3.5

Search strategies:

#	Query	Limiters/ Expanders	Results
S10	(S5 OR S6 OR S7 OR S8) AND (S4 AND S9)	Search modes - Boolean/Phrase Search Screen - Advanced Search	(705)
S9	S5 OR S6 OR S7 ORS8	Search modes - Boolean/Phrase Search Screen - Advanced Search	(74,099)

#	Query	Limiters/ Expanders	Results
S8	TX accessibility of care	Search modes - Boolean/Phrase Search Screen - Advanced Search	(56,082)
S7	TX equity of care	Search modes - Boolean/Phrase Search Screen - Advanced Search	(3,128)
S6	TX clinical effectiveness	Search modes - Boolean/Phrase Search Screen - Advanced Search	(20,820)
S5	TX quality improvement in health care	Search modes - Boolean/Phrase Search Screen - Advanced Search	(1,064)
S4	S1 OR S2 OR S3	Search modes - Boolean/Phrase Search Screen - Advanced Search	(3,372)
S3	TX performance awards	Limiters - Full Text; Published Date: 20000101-20141231 Search modes - Boolean/Phrase Search Screen - Advanced Search	(89)
S2	TX incentive awards	Limiters - Full Text; Published Date: 20000101-20141231 Search modes - Boolean/Phrase Search Screen - Advanced Search	(38)
S1	TX pay for performance	Limiters - Full Text; Published Date: 20000101-20141231 Search modes - Boolean/Phrase Search Screen - Advanced Search	(3,273)

Search results: 705 text results



Search History/Alerts

Print Search History | Retrieve Searches | Retrieve Alerts | Save Searches / Alerts

Select / deselect all | Search with AND | Search with OR | Delete Searches

	Search ID#	Search Terms	Search Options	Actions
	S10	(S5 OR S6 OR S7 OR S8) AND (S4 AND S9)	Search modes - Boolean/Phrase	View Results (705)
	S9	S5 OR S6 OR S7 OR S8	Search modes - Boolean/Phrase	View Results (74,099)
	S8	accessibility of care	Search modes - Boolean/Phrase	View Results (56,082)
	S7	equity of care	Search modes - Boolean/Phrase	View Results (3,128)
	S6	clinical effectiveness	Search modes - Boolean/Phrase	View Results (20,820)
	S5	quality improvement in health care	Search modes - Boolean/Phrase	View Results (1,064)
	S4	S1 OR S2 OR S3	Search modes - Boolean/Phrase	View Results (3,372)
	S3	TX performance awards	Limiters - Full Text; Abstract Available; Published Date: 20000101-20141231 Search modes - Boolean/Phrase	View Results (89)
	S2	TX incentive awards	Limiters - Full Text; Abstract Available; Published Date: 20000101-20141231 Search modes - Boolean/Phrase	View Results (38)
	S1	TX pay for performance	Limiters - Full Text; Abstract Available; Published Date: 20000101-20141231 Search modes - Boolean/Phrase	View Results (3,273)

Figure 3.5 The search strategies from CINAHL Plus with Full Text accessed through Mahidol database

3.7 Criteria for considering studies for this review

Table 3.1 shows the definition of inclusion and exclusion criteria can be framed in terms of the PICOS; Participants / Population, Intervention(s), Comparator(s), Outcomes and Study design of the included studies. [73]

Table 3.1 The definition of inclusion criteria and exclusion criteria

The PICOS	Inclusion criteria	Exclusion criteria
Participant(s)/ Population	• Studies were conducted in health systems based on universal health coverage context according to the health systems research by the World Health Organization as available in Appendix B	-
Intervention(s)	• At least one year of P4P implementation • period	• The P4P were at the demand side.

Table 3.1 The definition of inclusion criteria and exclusion criteria (cont.)

The PICOS	Inclusion criteria	Exclusion criteria
Comparator(s)	• At least one year non-P4P or before P4P implementation period	-
Outcome(s)	• The outcome findings: (1) The achieving of quality targets. (Quality improvement) (2) The clinical care recorded. (Accessibility of care)	• The irrelevant studies
Study design(s)	The review includes: • Randomized controlled trials • Non-randomized controlled trials; Controlled Clinical Trials, Quasi-experimental • Controlled before-after • Interrupted time series • Observational study; Cohort study, Cross sectional study	The review excludes: • Publications which are not peer reviewed • Publications without full-text • Publications focused on P4P theory or development • The P4P implementation without an evaluation

3.8 Data extraction

The studies were obtained data into a data extraction form (Appendix C) suggested by the Cochrane handbook [74]. In the data extraction form, there were the characteristics of the full-text eligible studies included; the first author's last name, publication year, country, study design, participants, method, interventions, outcomes and results.

3.9 Quality assessment

The eligible full-text studies have been assessed the quality by using the Modified Newcastle-Ottawa scale (NOS) [77] for the observational study design followed a bulletin of the WHO [79].

3.10 Data syntheses

3.10.1 Narrative synthesis

The narrative analysis consists of the findings that the reviewer attends to extract the critical solutions for the review questions. The results from different methodology and quality of studies should be elucidated in the narrative synthesis under the domain of review questions.

3.10.2 Meta-analysis

Meta-analysis had been undertaken followed the suggestion of the CRD [72] and the Cochrane handbook [74]. The included studies which provided separate performance estimates, will be carried out further analyses of the achieving quality targets (Quality improvement) and the recorded of health services (Accessibility to care). All statistical analysis was performed using Review Manager (RevMan) versions 5.3 (The Cochrane Collaboration, Oxford, London, UK) [78].

- **The effect measure**

The effect measure is pooled Mantel-Haenszel odds ratios (OR), in random effect model with 95% Confident Interval (CI), p-values less than 0.05 were considered significant.

The Mantel-Haenszel test is a method to detect an association between two dichotomous independent variables observed in the different circumstantial context.

The ultimate result presents in the form of odds ratio. The odds ratio is the probability of an event (or success) to another of no event (or not success). The odds ratio has to present with a confidence interval (CI). The confidence interval (CI) is a range within a probability degree of assurance (e.g. 95% CI).

The random effect model; the different studies are estimating random intervention effects.

- Heterogeneity explanations

The statistical measures of heterogeneity among studies tested with the I^2 statistic. I^2 values of 0 to 50% might be negligible, I^2 levels 50% to 100% should be considered. When there was statistical heterogeneity, then the possible causes of the dissimilarity should be explored.

The subgroup analysis was performed to explain the three causes of heterogeneity; clinical diversity or clinical heterogeneity, methodological diversity or methodological heterogeneity, statistical heterogeneity. In order to evaluate the variables that affected to the outcome findings.

The Cochrane handbook [74] explained that a forest plot as depicted in Figure 3.6 displays effect estimates and confidence intervals for both individual studies and meta-analyses. Each study is demonstrated by a block at the point estimate of intervention effect with a horizontal line extending either side of the block. The area of the block indicates the weight assigned to that study in the meta-analysis while the horizontal line presents the confidence interval (usually with a 95% level of confidence). The area of the block and the confidence interval contain the same information, but both make different interprets to the graphic. The confidence interval presents the range of intervention effects compatible with the study's result and indicates whether each was individually statistically significant. The size of the block draws the eye towards the studies with larger weight, which dominate the calculation of the pooled result. Heterogeneity can be detected by a forest plot and inspecting the overlap of confidence intervals (CI). When the CI overlap, the studies are more similar or homogeneous. Conversely, if the CI does not overlap; the forest plot presents the heterogeneity.

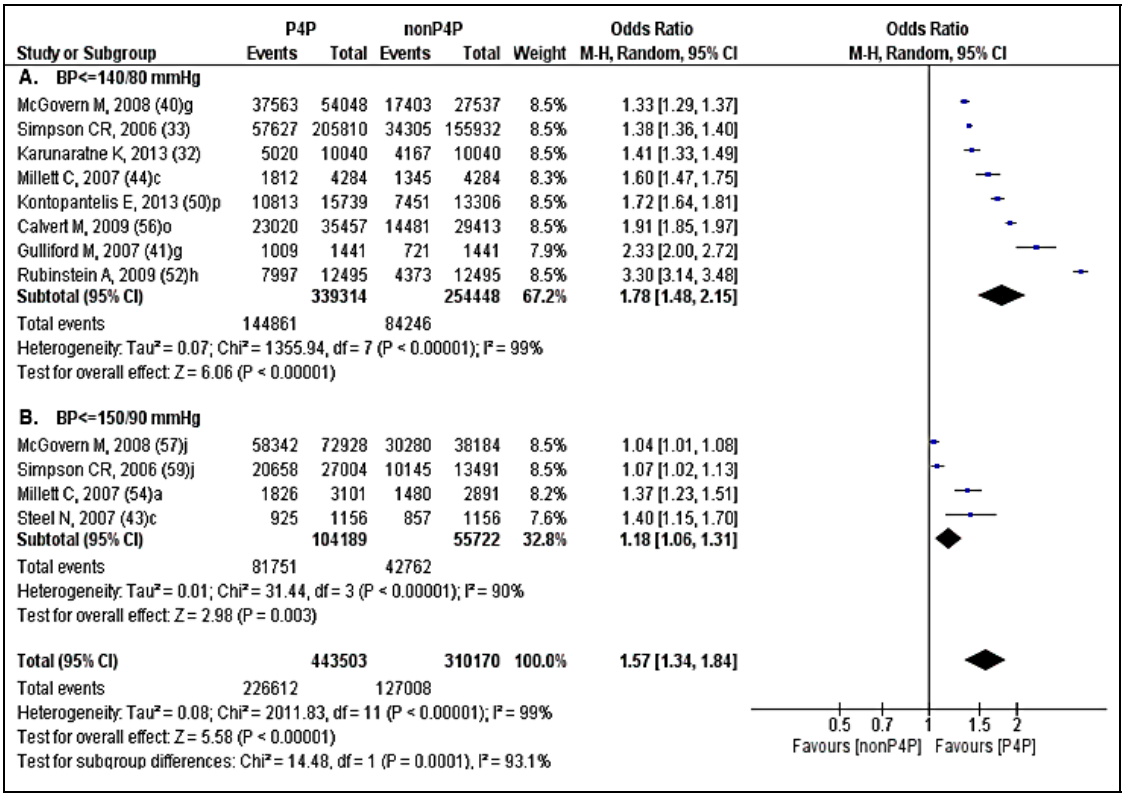


Figure 3.6 Example of forest plots

- **Publication bias analysis.** The basic causes of publication bias are; uninteresting/unfavorable studies less likely to be published, evidence suggests statistical significance most important factor [80].

Publication bias assessment revealed by a funnel plot that depends on a number of publications in the analysis. The evaluation can be performed among more than ten studies.

The Cochrane handbook [74] explained that a funnel plot is a scatter plot of the effect estimates from individual studies against some measure of each study's size or precision. The standard error of the effect estimate is often chosen as the measure of study size and plotted on the vertical axis with a reversed scale that places the larger, most powerful studies towards the top. The effect estimates from smaller studies should scatter more widely at the bottom, with the spread narrowing among larger studies. In the absence of bias and between study heterogeneity, the scatter will be due to sampling variation alone and the plot will resemble a symmetrical inverted funnel.

Hypothetical funnel plots; Figure 3.7 presented symmetrical plot in the absence of bias, Figure 3.8 depicted asymmetrical plot in the presence of reporting bias, Figure 3.9 demonstrated asymmetrical plot in the presence of bias because some smaller studies (open circles) are of lower methodological quality and therefore produce exaggerated intervention effect estimates.

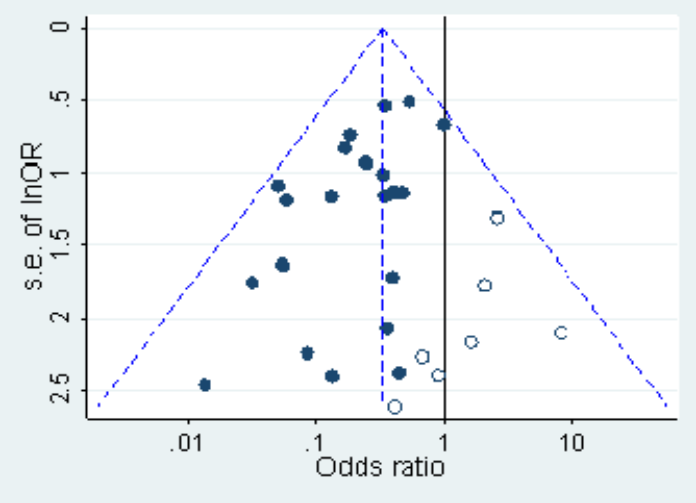


Figure 3.7 Symmetrical plot in the absence of bias

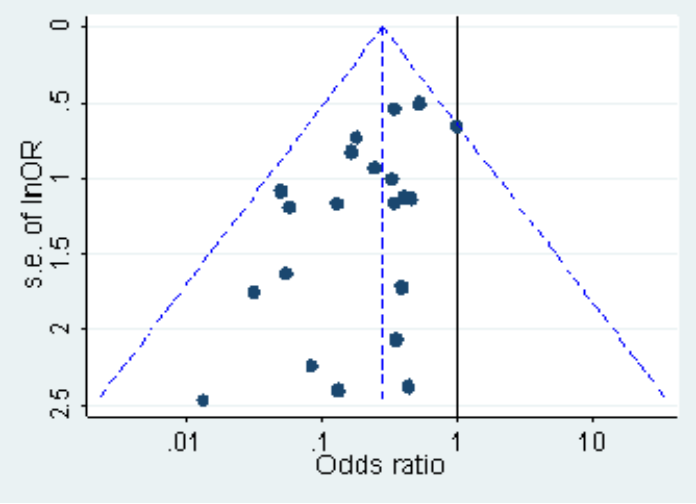


Figure 3.8 Asymmetrical plot in the presence of reporting bias

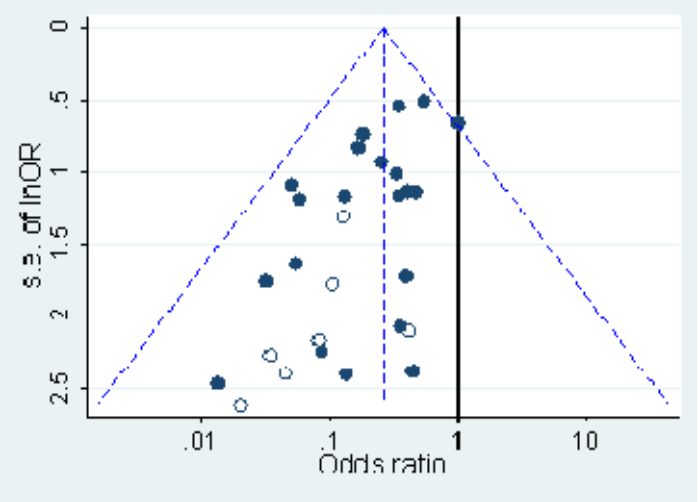


Figure 3.9 Asymmetrical plot in the presence of bias because some smaller studies (open circles) are of lower methodological quality and therefore produce exaggerated intervention effect estimates

3.11 Ethical consideration

This study documentary proof of exemption by Ethnical Review Committee for Human Research, Faculty of Public Health, Mahidol University as available in Appendix A.

CHAPTER IV
RESULTS

4.1 Result of the searches

This systematic review study performed to assess the effects of pay-for-performance (P4P) incentive scheme on health care systems with universal health coverage (UHC), in terms of the quality improvement and the accessibility to care services.

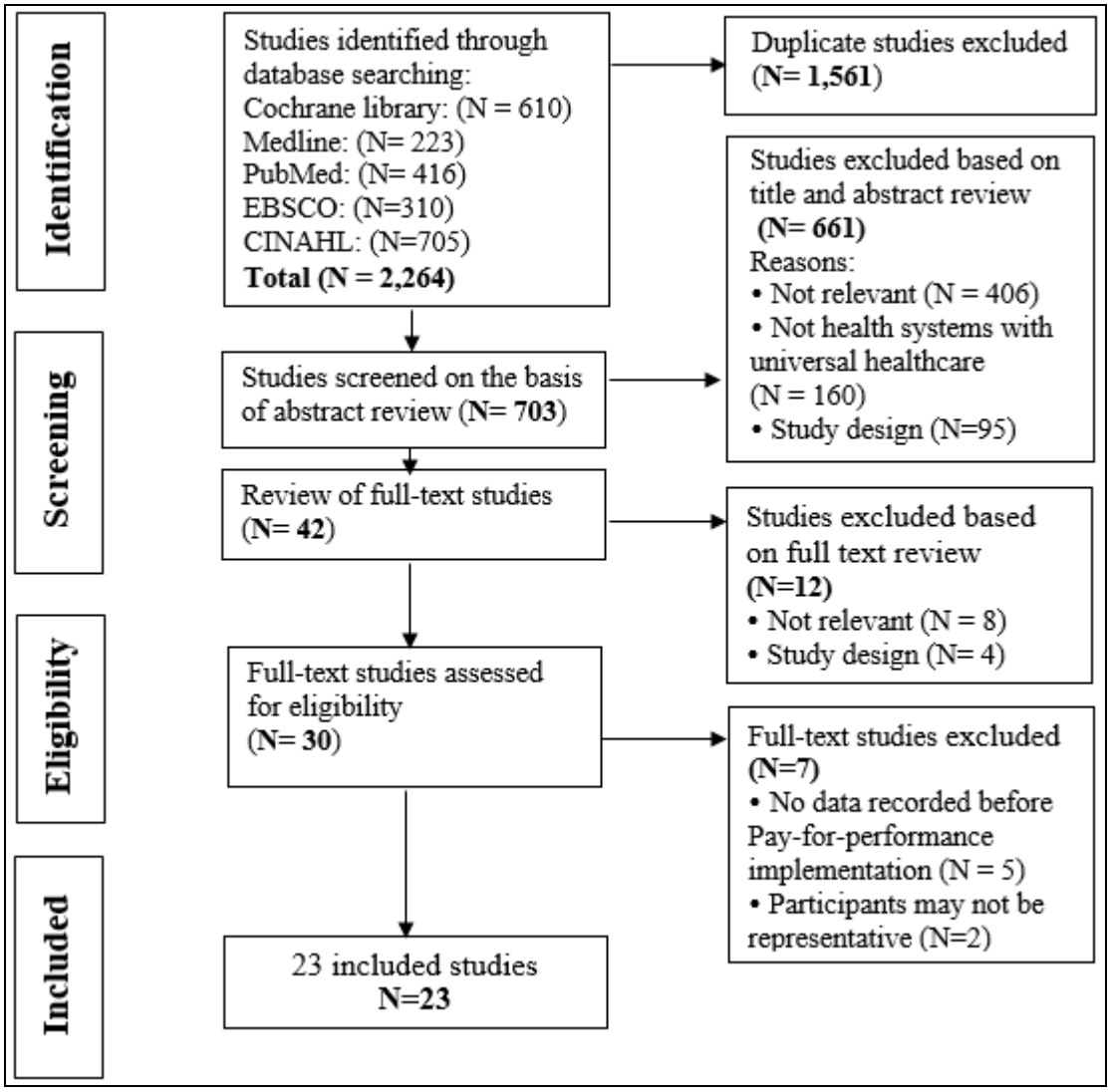


Figure 4.1 PRISMA flow chart of studies selection process

4.1.1 Identification of studies

Figure 4.1 demonstrated a PRISMA flow chart of the studies selection process. This study conducted its searching in five databases. The selected databases were the Cochrane Library, PUBMED, MEDLINE, EBSCO host service and CINAHL, focusing on English language studies published between January 2000 and March 2014. An inclusive search strategy was developed to identify the relevant studies according to MeSH (Medical Subject Heading) search terms; quality of health care, utilization, health services accessibility, equity for care, quality effectiveness, quality improvement, access to care. These search terms were combined with; pay for performance, incentive awards, incentive-based compensation, reimbursement, incentive, physician incentive plans, health provider incentive, reward and bonus. The results of the searches were 2,264 studies, identified through the five databases searching: 610 studies from Cochrane library, 223 studies from Medline, 416 studies from PubMed, 310 studies from EBSCO and 705 studies from CINAHL. All studies had been considered the pertinence by their titles and abstracts. The 1,561 duplicate studies were rejected.

4.1.2 Screening of studies

The 703 studies had been screened theirs abstracts. The 661 studies were excluded based on three reasons: (1) the 406 studies were irrelevant studies, (2) the 160 studies did not conduct from the universal health coverage context (UHC) and (3) the 95 studies did not meet the including criteria. There existed 42 relevant studies for full-text reading. After the full-text screening, thirty full-text papers were included and the twelve studies were excluded according to two reasons; eight studies were irrelevant pay-for-performance (P4P) scheme, and the four were unevaluated studies.

4.1.3 Eligibility of studies

The thirty included full-text papers had been evaluated for their eligibility and quality assessment. The seven excluded studies were presented in table 4.1. They were excluded because of two reasons: (1) five studies did not report the data before the implementation of P4P and (2) two studies did not be the representative of the population.

Table 4.1 Characteristics of excluded studies

Study/year	Reason for exclusion
Doran T., 2008 [39].	The study did not report the analysis of the data before P4P implementation.
Crawley D., 2009 [60].	Participants may not be the representative of the population because this study merged the six social classes (I-professional, II-managerial and technical, IIIN-skilled non-manual, IIIM-skilled manual, IV-partly skilled, V- unskilled) into two occupational groups (non-manual-I, II, IIIN and manual-IIIM, IV, V).
Kirschner K., 2013 [62].	The study did not report the analysis of the data before P4P implementation.
Chen P, 2012 [81]	The study did not report the analysis of the data before P4P implementation and evaluated only the first year of the P4P.
Duckett S., 2008 [82].	The study did not report the analysis of the data before P4P implementation and evaluated only the first year of the P4P.
Benavent J., 2009 [83].	The study did not report the data before Pay-for-performance (P4P) implementation.
Cheng A., 2013 [84].	Participants may not be the representative of the population because the data collected from hospitals in the greater Vancouver area only.

4.1.4 Included of studies

There remained twenty-three ultimate included studies. The characteristics of twenty-three included studies by study available at Appendix E. Table 4.2 presented the characteristics of the included studies.

Table 4.2 Characteristics of twenty-three included studies

1st Author, Year [Ref], Country	Study design	Performance indicators for pay-for-performance (P4P)		The compositions of pay-for-performance (P4P) implementation				The effectiveness of pay-for-performance (P4P)		Quality assessment
		Quality indicator targets	Disease/ Conditions	Data collection	Performance measures	Basis of payment	Types of incentive	Quality indicator targets achievement	Recording of quality indicators	
Karunaratne K, 2013 [32], The United Kingdom	Cohort	According to P4P renal indicator targets	Chronic Kidney disease. (CKD)	The electronic recording of data on computer systems	Process measure	Incentive for individual general practitioner.	Financial incentive	Positive effects: improvement of blood pressure control in patients with CKD	Positive effects: the recording of blood pressure increased.	8
Simpson CR, 2011 [33], Scotland	Cross sectional	Targeting quality care of P4P on the management of patients with HT	Hypertension (HT)	The electronic recording of data on computer systems	Process measure	Incentive for individual general practitioner	Financial incentive	Positive effects: improvement of blood pressure control in cardiovascular diseases patients	Partial effects: the recording of clinical care increased, but not in older and deprived patients.	8
Doran T, 2011 [34], The United Kingdom	Cohort	To enhance the quality and performance of clinical indicator targets	Chronic disease.	The electronic recording of data on computer systems	Process measure	Incentive for individual general practitioner	Financial incentive	Positive effects: significant improvement in quality for all indicators between 2001-2007	Positive effects: the recording of clinical care increased in chronic disease patents.	7

Table 4.2 Characteristics of twenty-three included studies (cont.)

1st Author, Year [Ref], Country	Study design	Performance indicators for pay-for-performance (P4P)		The compositions of pay-for-performance (P4P) implementation				The effectiveness of pay-for-performance (P4P)		Quality assessment
		Quality indicator targets	Disease/ Conditions	Data collection	Performance measures	Basis of payment	Types of incentive	Quality indicator targets achievement	Recording of quality indicators	
McGovern M, 2008 [40], Scotland	Cross sectional	To improve the ascertainment of quality measures	Diabetes (DM)	The electronic recording of data on computer systems	Process measure	Incentive for individual general practitioner	Financial incentive	Positive effects: the proportion of patients achieving the DM-related quality indicator targets increased	Partial effects: the recording of DM quality indicator increased in the old, but less good for women.	8
Gulliford M, 2007 [41], The United Kingdom	Cross sectional	To improve the metabolic targets achievement	Diabetes (DM)	The electronic recording of data on computer systems	Process measure	Incentive for individual general practitioner	Financial incentive	Mixed effects: the clinical outcomes are improving, but not in deprived areas with less organized service	Positive effects: the recording of intermediate outcomes increased.	7

Table 4.2 Characteristics of twenty-three included studies (cont.)

1st Author, Year [Ref], Country	Study design	Performance indicators for pay-for-performance (P4P)		The compositions of pay-for-performance (P4P) implementation				The effectiveness of pay-for-performance (P4P)		Quality assessment
		Quality indicator targets	Disease/ Conditions	Data collection	Performance measures	Basis of payment	Types of incentive	Quality indicator targets achievement	Recording of quality indicators	
Millett C, 2007 [42], The United Kingdom	Cohort	To improve support for smoking cessation in DM patients	Reduce smoking in the Diabetes (DM)	The electronic recording of data on computer systems	Process measure	Incentive For individual general practitioner	Financial incentive	Positive effects: increases the provision of support for smoking quit.	Positive effects: the recording of smoking status increased.	8
Steel N, 2007 [43] , The United Kingdom	Cohort	To improve the quality of clinical care achievement	Chronic disease	The electronic recording of data on computer systems	Process measure	Incentive For individual general practitioner	Financial incentive	Positive effects: substantial quality improvement for incentivized targets	Positive effects: the recording of blood pressure increased.	8
Millett C, 2007[44], The United Kingdom	Cohort	To improve the management of people with Diabetes	Diabetes (DM)	The electronic recording of data on computer systems	Process measure	Incentive for individual general practitioner	Financial incentive	Positive effects: the proportion of patients reaching treatment targets increased significantly	Positive effects: the recording of DM intermediate outcomes increased.	7

Table 4.2 Characteristics of twenty-three included studies (cont.)

1st Author, Year [Ref], Country	Study design	Performance indicators for pay-for-performance (P4P)		The compositions of pay-for-performance (P4P) implementation				The effectiveness of pay-for-performance (P4P)		Quality assessment
		Quality indicator targets	Disease/ Conditions	Data collection	Performance measures	Basis of payment	Types of incentive	Quality indicator targets achievement	Recording of quality indicators	
Li Y, 2010 [46], Taiwan	Cohort	To improve cure rate and length of treatment	Tuberculosis (TB)	Not applicable	Outcome measure	Incentive for group of providers in TB-P4P program.	Financial incentive	Positive effects: the cure rate and average length of treatment improved significantly.	Positive effects: the recording of TB clinical information increased.	8
Kuo R, 2011[47], Taiwan,	Cohort	To encourage guideline-adhering therapy and better patients survival	Breast cancer (BC)	Not applicable	Outcome measure	Incentive for group of providers in BC-P4P program.	Financial incentive	Positive effects: improving better quality care and better outcome of breast cancer care	Positive effects: the recording of BC clinical information increased.	8
Tsai W, 2010[48], Taiwan	Cohort	To improve the treatment and reduce the default rate of tuberculosis.	Tuberculosis (TB)	Not applicable	Outcome measure	Incentive for group of providers in TB-P4P program.	Financial incentive	Positive effects: improving quality of tuberculosis treatment and decreasing the default rate.	Positive effects: the recording of TB clinical information increased.	8

Table 4.2 Characteristics of twenty-three included studies (cont.)

1st Author, Year [Ref], Country	Study design	Performance indicators for pay-for-performance (P4P)		The compositions of pay-for-performance (P4P) implementation				The effectiveness of pay-for-performance (P4P)		Quality assessment
		Quality indicator targets	Disease/ Conditions	Data collection	Performance measures	Basis of payment	Types of incentive	Quality indicator targets achievement	Recording of quality indicators	
Lai C, 2013 [49], Taiwan,	Cross sectional	To improve the guideline-adherence and quality of diabetes care	Diabetes (DM)	Not applicable	Process measure	Incentive for group of providers	Financial incentive	Positive effects: more effectively manage diabetes in general population.	Positive effects: the recording of the guideline-recommended, tests/ examination increased.	8
Kontopantelis E, 2013 [50], The United Kingdom	Cohort	To improve the recorded quality of care for patients with diabetes.	Diabetes (DM)	The electronic recording of data on computer systems	Process measure	Incentive For individual family doctor	Financial incentive	Partial effects: improvement in the recorded quality of DM care in the first year, but in subsequent years were more modest	Positive effects: the recording of DM clinical information increased.	8
Colais P, 2013[51], Italy	Cohort	To improve the quality of clinical care	Hip fracture.	Not applicable	Process measure	Incentive for group of providers	Financial incentive	Positive effects: improving quality of hip fracture for elderly patients.	Positive effects: the reduction of hospital admissions increased.	8

Table 4.2 Characteristics of twenty-three included studies (cont.)

1st Author, Year [Ref], Country	Study design	Performance indicators for pay-for-performance (P4P)		The compositions of pay-for-performance (P4P) implementation				The effectiveness of pay-for-performance (P4P)		Quality assessment
		Quality indicator targets	Disease/ Conditions	Data collection	Performance measures	Basis of payment	Types of incentive	Quality indicator targets achievement	Recording of quality indicators	
Rubinstein A, 2009 [52], Argentina	Cross sectional	To improve health outcomes	Chronic disease.	Not applicable	Process measure	Incentive for group of providers	Financial incentive	Positive effects: improving quality of chronic disease treatment.	Positive effects: the recording of clinical information increased.	7
Millett C, 2009 [54], The United Kingdom	Cross sectional,	To improve the quality of chronic disease management	Coronary Heart Disease (CHD)	The electronic recording of data on computer systems	Process measure	Incentive for individual general practitioner	Financial incentive	Positive effects: improving quality of chronic disease treatment.	Positive effects: better and more equitable management of CHD	7
Calvert M, 2009 [56], The United Kingdom	Cohort	To improve the quality of chronic disease management	Diabetes (DM)	The electronic recording of data on computer systems	Process measure	Incentive for individual general practitioner	Financial incentive	Partial effects: the management of people with DM has improved, but the impact of the P4P on care is uncertain.	Positive effects: the recording of clinical information increased.	7

Table 4.2 Characteristics of twenty-three included studies (cont.)

1st Author, Year [Ref], Country	Study design	Performance indicators for pay-for-performance (P4P)		The compositions of pay-for-performance (P4P) implementation				The effectiveness of pay-for-performance (P4P)		Quality assessment
		Quality indicator targets	Disease/ Conditions	Data collection	Performance measures	Basis of payment	Types of incentive	Quality indicator targets achievement	Recording of quality indicators	
McGovern M, 2008 [57], Scotland	Cross sectional	To improve the quality of Coronary Heart Disease management	Coronary Heart Disease (CHD)	The electronic recording of data on computer systems	Process measure.	Incentive for individual general practitioner	Financial incentive	Partial effects: the management of CHD improved, but the impact of the P4P on care is uncertain.	Partial effects: the recording of CHD-related increased, but not all groups of population.	8
Simpson CR, 2006 [59], Scotland	Cross sectional	To improve the recording of quality indicators for patients with stroke	Stroke	The electronic recording of data on computer systems	Process measure.	Incentive for individual general practitioner	Financial incentive	Partial effects: the management of stroke care improved, but not in female, older and more deprived groups	Partial effects: the recording of quality indicators increased substantially, but inequitable care exists.	8

Table 4.2 Characteristics of twenty-three included studies (cont.)

1st Author, Year [Ref], Country	Study design	Performance indicators for pay-for-performance (P4P)		The compositions of pay-for-performance (P4P) implementation				The effectiveness of pay-for-performance (P4P)		Quality assessment
		Quality indicator targets	Disease/ Conditions	Data collection	Performance measures	Basis of payment	Types of incentive	Quality indicator targets achievement	Recording of quality indicators	
Alshamsan R, 2012 [64], The United Kingdom	Interrupted Time Series (ITS)	To improve the achievement of intermediate outcomes targets	Diabetes (DM)	The electronic recording of data on computer systems	Process measure	Incentive for individual general practitioner	Financial incentive	Partial effects: Improving quality of clinic treatment, but not in vulnerable population.	Partial effects: the recording of quality indicators increased substantially, not all groups of population.	8
Kiran T, 2012[66], Canada	Cohort	To improve the quality of diabetes care based on guidelines	Diabetes (DM)	Not applicable	Process measure	Incentive for individual physicians.	Financial incentive	Partial effects: minimal quality improvement in quality of diabetes care	Partial effects: the recording of quality indicators increased, but not in all population.	8
Inoue Y, 2011[70], Japan	Cohort	To improve return home rate based on time of discharge.	Functional Independence Measure (FIM)	Not applicable	Outcome measure	Incentive for group of providers	Financial incentive	No effect of P4P on return home rate.	Not applicable	8

Table 4.2 Characteristics of twenty-three included studies (cont.)

1st Author, Year [Ref], Country	Study design	Performance indicators for pay-for-performance (P4P)		The compositions of pay-for-performance (P4P) implementation				The effectiveness of pay-for-performance (P4P)		Quality assessment
		Quality indicator targets	Disease/ Conditions	Data collection	Performance measures	Basis of payment	Types of incentive	Quality indicator targets achievement	Recording of quality indicators	
Adaji A, 2013[85], Australia	Cohort	To improve the quality of diabetes care based on guidelines	Diabetes (DM)	Not applicable	Process measure	Incentive for group of providers	Financial incentive	Positive effects: Increased testing of process measure that met guideline requirement for diabetes	Positive effects: the recording of clinical information increased.	7

4.2 Characteristics of the included studies

According to Table 4.2, the characteristics of twenty-three included studies had been categorized in the following descriptions:

4.2.1 The general characteristics of the included studies.

The twenty-three included studies were recent publication (2006-2013). There was only one study published in 2006, four studies published in 2007, two studies published in 2008, three studies published in 2009, two studies published in 2010, four studies published in 2011, two studies from 2012, and five studies from 2013.

The twenty-three included studies were conducted in eight health care systems based on universal health coverage (UHC) context. Ten studies performed in the United Kingdom (UK), four studies carried out in Scotland, four studies conducted in Taiwan and the other studies were brought from Italy, Argentina, Canada, Japan, Australia, respectively.

The included studies can be categorized into three types of study designs: thirteen cohort studies, eight cross-sectional studies and only one interrupted time series study.

4.2.2 The performance indicator targets of P4P implementation.

The performance indicator targets were achieving of quality indicator targets and recording of quality indicators in the UHC context. The included studied provided the evaluation on P4P implementing in ten diseases as the following details: ten studies analyzed the P4P in diabetes mellitus (DM), three studies in chronic disease, two studies in coronary heart disease, two studies in tuberculosis, the other six studies analyzed the P4P in the functional rehabilitation, stroke, hip fracture, breast cancer, chronic kidney disease and hypertension patients as presented in Table 4.3.

Table 4.3 The areas of P4P implementation in the included studies

Diseases/conditions	Number of studies
Diabetes mellitus	10
Coronary heart disease	2
Tuberculosis	2
Functional Independence Measure (FIM)	1
Stroke (ST)	1
Hip Fracture	1
Breast cancer (BC)	1
Chronic Kidney disease (CKD)	1
Hypertension (HT)	1

4.2.3 The compositions of pay-for-performance (P4P) implementation

- The data collection systems.

There were fourteen included studies from the United Kingdom and Scotland used the electronic recording of data on computer systems such as Electronic Medical Record (EMR) for data collections.

- The performance measures.

The performance measures in the twenty-three included studies were classified into process measures and outcome measures without structure measure. There were nineteen studies with process measures and the four outcome measures.

- Basis of payment and types of incentive

All payments in the included studies were the financial incentive for health care provider. There were eight studies paid for groups of provider or hospitals and fifteen studies paid for individual provider.

4.2.4 The quality assessment

The quality assessment of the included studies were reported in Table 4.4. There are seven studies with score 7 and seventeen studies with score 8.

Table 4.4 The quality assessment of the included studies

No.	1st Author, Year, [Reference]	Modified Newcastle-Ottawa Scale (NOS)								Total score
		1. Selection			2. Comparability		3. Outcome			
		1.1 ★	1.2 ★	1.3 ★	a ★	b ★	3.1 ★	3.2 ★	3.3 ★	
1	Karunaratne K, 2013 [32]	⋆b	⋆a	⋆a	⋆a	⋆b	⋆a	⋆b	⋆a	8
2	Simpson C, 2011 [33]	⋆b	⋆a	⋆a	⋆a	⋆b	⋆a	⋆b	⋆a	8
3	Doran T, 2011 [34]	⋆b	⋆a	⋆a	⋆a	-	⋆a	⋆b	⋆a	7
4	McGovern M, 2008 [40]	⋆b	⋆a	⋆a	⋆a	⋆b	⋆a	⋆b	⋆a	8
5	Gulliford M, 2007 [41]	⋆b	⋆a	⋆a	⋆a	-	⋆a	⋆b	⋆a	7
6	Millett C, 2007 [42]	⋆b	⋆a	⋆a	⋆a	⋆b	⋆a	⋆b	⋆a	8
7	Steel N, 2007 [43]	⋆b	⋆a	⋆a	⋆a	⋆b	⋆a	⋆b	⋆a	8
8	Millett C, 2007 [44]	⋆b	⋆a	⋆a	-	⋆b	⋆a	⋆b	⋆a	7
9	Li Y, 2010 [46]	⋆b	⋆a	⋆a	⋆a	⋆b	⋆a	⋆b	⋆a	8
10	Kuo R, 2011 [47]	⋆b	⋆a	⋆a	⋆a	⋆b	⋆a	⋆b	⋆a	8
11	Tsai W, 2010 [48]	⋆b	⋆a	⋆a	⋆a	⋆b	⋆a	⋆b	⋆a	8
12	Lai C, 2013 [49]	⋆b	⋆a	⋆a	⋆a	⋆b	⋆a	⋆b	⋆a	8
13	Kontopantelis E, 2013 [50]	⋆b	⋆a	⋆a	⋆a	⋆b	⋆a	⋆b	⋆a	8
14	Colais P, 2013 [51]	⋆b	⋆a	⋆a	⋆a	⋆b	⋆a	⋆b	⋆a	8
15	Rubinstein A, 2009 [52]	⋆b	⋆a	⋆a	⋆a	-	⋆a	⋆b	⋆a	7
16	Millett C, 2009 [54]	⋆b	⋆a	⋆a	⋆a	-	⋆a	⋆b	⋆a	7
17	Calvert M, 2009 [56]	⋆b	⋆a	⋆a	⋆a	-	⋆a	⋆b	⋆a	7
18	McGovern M, 2008 [57]	⋆b	⋆a	⋆a	⋆a	⋆b	⋆a	⋆b	⋆a	8
19	Simpson C, 2006, [59]	⋆b	⋆a	⋆a	⋆a	⋆b	⋆a	⋆b	⋆a	8

Table 4.4 The quality assessment of the included studies (cont.)

No.	1st Author, Year, [Reference]	Modified Newcastle-Ottawa Scale (NOS)								Total score
		1. Selection			2. Comparability		3. Outcome			
		1.1	1.2	1.3	a	b	3.1	3.2	3.3	
		★	★	★	★	★	★	★	★	
20	Alshamsan R, 2012, [64]	⋆b	⋆a	⋆a	⋆a	⋆b	⋆a	⋆b	⋆a	8
21	Kiran T, 2012 [66]	⋆b	⋆a	⋆a	⋆a	⋆b	⋆a	⋆b	⋆a	8
22	Inoue Y, 2011 [70]	⋆b	⋆a	⋆a	⋆a	⋆b	⋆a	⋆b	⋆a	8
23	Adaji A, 2013 [85]	⋆b	⋆a	⋆a	⋆a	-	⋆a	⋆b	⋆a	7

Modified Newcastle-Ottawa Scale (NOS) as the following details

A study can be awarded a maximum of three stars for Selection and Outcome categories and maximum of two stars can be given for Comparability.

Total score = 8 High quality: = 8

Middle quality: = 5-7 Low quality: <= 4

1. Assessment of “Selection” categories comprised 3 items (3 Scores)

(1) Representativeness

(a) Truly representative of the general population ⋆

(b) Somewhat representative of the general population ⋆

(c) Selected group of users e.g. nurses, volunteers

(d) No description of the derivation of the cohort

(2) Selection of the control or baseline groups.

(a) Conducted from the same community as the intervention group ⋆

(b) Conducted from a different source

(c) No description.

(3) Ascertainment of the data

(a) Validated data resources ✯

(b) No validation is mentioned

(c) No description

2. Assessment of "Comparability" categories comprised 2 items. (2 Scores)

Comparability on the basis of the design or analysis

(a) Study controls for conditions ✯

(b) Study controls for gender, age, comorbidities, deprivation status ✯

3. Assessment of "Outcome" categories comprised 3 items. (3 Scores)

(1) Assessment of outcome

(a) Objective measurements (the recorded quality indicators, the achievement of quality targets.) ✯

(b) Prescribing of medications ✯

(c) Self report

(d) No description

(2) Sufficient follow-up time to allow outcome to occur.

(a) Yes (majority of population at least 1 year) ✯

(b) No

(3) Adequacy of follow-up.

(a) Yes (majority of population at least 1 year) ✯

(b) No

4.2.5 The effectiveness of pay-for-performance (P4P)

The P4P effectiveness had been extracted from twenty-three included studies. Each of them contained data before P4P implementation periods (non-P4P period) and after P4P implementation periods (P4P period). Comparison statistics calculated by Review Manager (RevMan) versions 5.3 (The Cochrane Collaboration, Oxford, London, UK) [78]. Overall results were available in Appendix E. Further subgroup analysis was performed to obtain a summary effect size of P4P in order to understand factors that affect the effect size [86].

4.3 Subgroup analysis

In a subgroup analysis, the effect estimates reported from at least three studies, and data were statistical analyzed using the RevMan as the following description;

The P4P effectiveness finding with 95% Confidence Interval (CI) were calculated by using a random effect. P-values less than 0.05 were considered statistically significant. The different variables presented in forest plots.

An ascertainment for heterogeneity was carried out to detect inconsistency in the study results. I^2 statistics was used to assess heterogeneity, where $P \leq 0.05$ or $I^2 > 50\%$. The heterogeneity < 50% referred to the effect is assumed to be due to change. $I^2 > 50\%$ was interpreted the effect is random and the effectiveness of P4P on health care system seem to be unclear ($I^2 > 50\%$), it is necessary to explore the heterogeneity, in order to elucidate the possible variables that more influence to the P4P intervention and detect the magnitudes of the effects.

Publication bias among more than ten studies was evaluated using funnel plots.

All similar effect estimates had been selected into the process of subgroup comparison as the following sequences;

4.3.1 The comparison of clinical targets estimates.

1. Comparison of the blood pressure (BP) target estimates.

The twelve BP target estimates had been extracted from twelve included studies. There were different variables that affect the achieving of blood pressure target in the patients. DeCoster J. [87] suggested that the appropriateness of a moderating variable can be evaluated by the process of moderator analysis. In order to approve a distinguished variable that influence to the P4P effectiveness.

In this moderator analysis process, there were four groups of selected variables as the following details;

(1) The blood pressure (BP) target levels;

A. The BP target lesser than or equal to140/80 mmHg

B. The BP target lesser than or equal to150/90 mmHg.

(2) The methodological difference between the studies.

A. Cohort study design.

B. Cross-sectional study design.

(3) The database setting / country.

A. The United Kingdom (UK)

B. Scotland

C. Argentina.

(4) The condition / disease.

A. Diabetes Mellitus (DM)

B. Coronary Heart Disease (CHD)

C. Stroke (ST)

D. Hypertension. (HT)

All variables had been tested heterogeneity across their subgroup for moderator analysis. Statistics calculated by the Review Manager (RevMan). The comparison was demonstrated in Table 4.5.

There was very high heterogeneity (I^2) of each comparison groups. However, the decreasing of the I^2 in all subgroups were detected (*). Thus, the

selected moderating variables were suitable for further subgroup analysis of the blood pressure target estimates.

Table 4.5 Subgroup analysis of the blood pressure targets estimates

Subgroups	Estimates	Participants	Effect Estimate Odds Ratio (M-H, Random, 95% CI)	Hetero-geneity (I^2)
1) BP (target levels)	**12**	**753673**	**1.57 (1.34, 1.84)**	**99%**
A. BP <=140/80	8	593762	1.78 (1.48, 2.15)	99%
B. BP <=150/90	4	159911	1.18 (1.06, 1.31)	90% *
2) BP (study designs)	**12**	**683056**	**1.57 (1.35, 1.83)**	**99%**
A. Cohort	5	124875	1.61 (1.42, 1.84)	96% *
B. Cross-sectional	7	558181	1.55 (1.23, 1.94)	100%
3) BP (settings)	**12**	**753673**	**1.57 (1.34, 1.84)**	**99%**
A. The UK	7	133749	1.65 (1.46, 1.86)	95% *
B. Scotland	4	594934	1.20 (1.04, 1.38)	99%
C. Argentina	1	24990	3.30 (3.14, 3.48)	1 study
4) BP (conditions)	**12**	**753673**	**1.57 (1.34, 1.84)**	**99%**
A. Diabetes	5	186950	1.74 (1.45, 2.09)	99%
B. Hypertension	3	389044	1.86 (0.94, 3.66)	100%
C. CHD	2	117104	1.19 (0.91, 1.55)	96% *
D. CKD	1	20080	1.41 (1.33, 1.49)	1 study
E. Stroke	1	40495	1.07 (1.02, 1.13)	1 study

1) Subgroup analysis of blood pressure target estimates based on target levels

Figure 4.2, there were twelve studies with eight estimates that referred to the achieving blood pressure (BP) target of <= 140/80 mmHg and four estimates that referred to the achieving BP target of <= 150/90 mmHg.

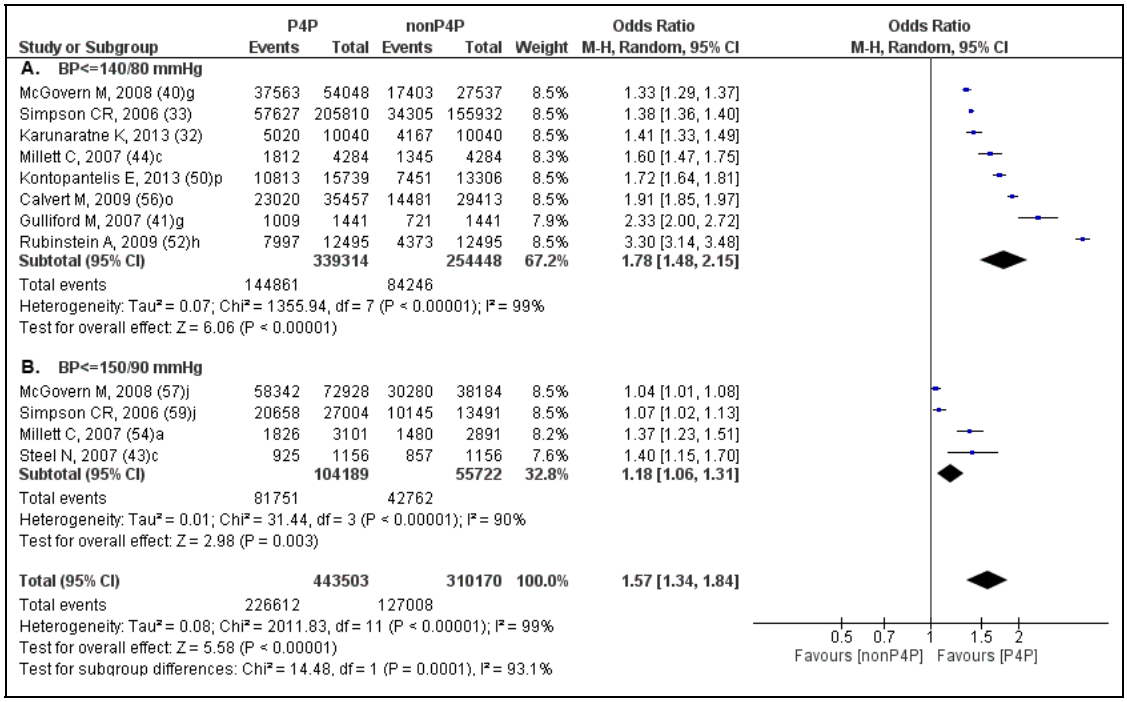


Figure 4.2 Subgroup analysis blood pressure target estimates based on target levels

P4P period had substantially greater odds of target achievement than non-P4P period. The overall Mantal-Haenszel odds ratio (OR) was 1.57; 95% CI 1.34 to 1.84; P < 0.00001, in a random effect model. Significant heterogeneity existed among the estimates; I^2 = 99% (P < 0.00001).

The association of the BP target levels and the achievement of the estimates showed in Figure 4.2.

Figure 4.2.A. reported that the achieving BP target of <=140/80 mmHg, in P4P period had significantly greater odds of target achievement than non-P4P period. The odds ratio was 1.78; 95% CI 1.48 to 2.15; P < 0.00001, random effect model. Heterogeneity existed; I^2 = 99% (P < 0.00001).

Figure 4.2.B. showed the achieving BP target of <=150/90 mmHg, in P4P period had significantly greater odds of target achievement than non-P4P period. The odds ratio was 1.18; 95% CI 1.06 to 1.31; P = 0.003, random effect model. There was reducing heterogeneity but still remained; I^2 = 90% (P < 0.00001).

2) Subgroup analysis of the blood pressure target estimates based on study designs.

Figure 4.3, there were twelve studies with five achieving estimates, which extracted from cohort study and seven achieving estimates that extracted from cross-sectional study.

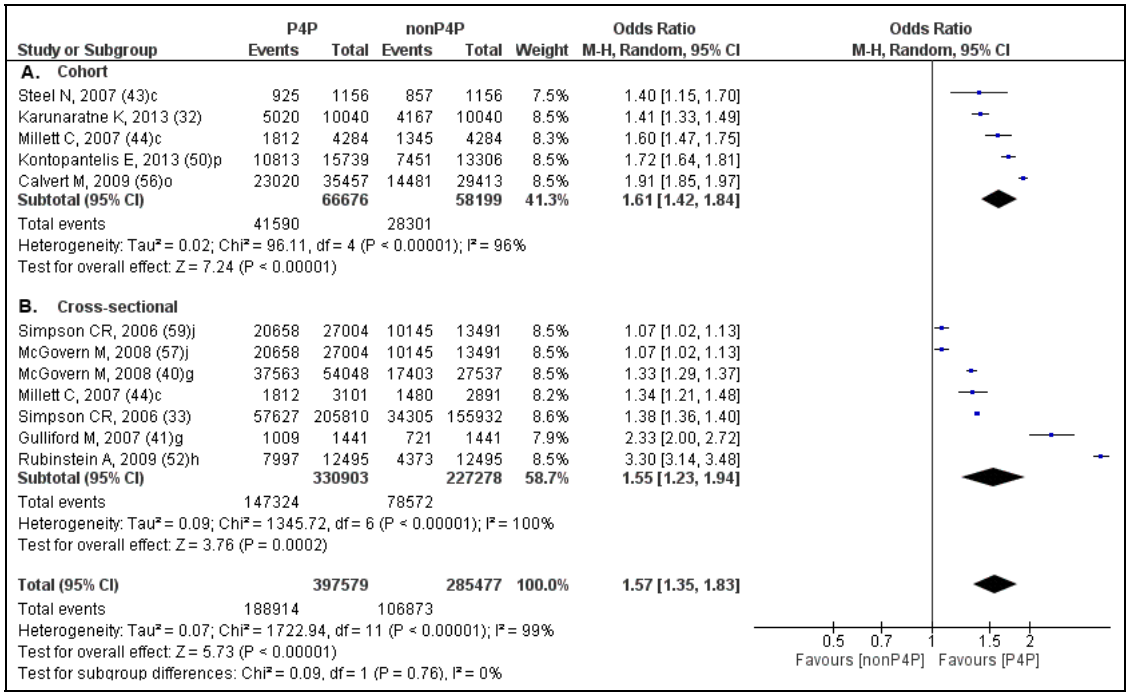


Figure 4.3 Subgroup analysis blood pressure target estimates based on study designs

P4P period had meaningfully greater odds of target achievement than non-P4P period. The overall Mantal-Haenszel odds ratio (OR) was 1.57; 95% CI 1.35 to 1.83; P = 0.002, in a random effect model. Considerably heterogeneity existed among the estimates; I^2 = 99% (P < 0.00001).

Subgroup analysis demonstrated the association of each the study designs and the target achievement of separate estimates as the following details;

Figure 4.3.A reported, the estimates from cohort studies that performed in P4P period had significantly greater odds of target achievement than non-P4P period. The odds ratio (OR) was 1.61; 95% CI 1.48 to 1.84; P < 0.00001, random effect model. There was reducing heterogeneity but still remained; I^2 = 96% (P < 0.00001).

Figure 4.3.B showed, the estimates from cross-sectional that conducted in P4P period had significantly greater odds of target achievement than non-P4P period. The odds ratio (OR) was 1.55; 95% CI 1.23 to 1.94; P=0.0002, random effect model. Significant heterogeneity existed among the estimates; I^2 = 100% (P < 0.00001).

3) Subgroup analysis of the blood pressure target achievement based on database setting or country.

Figure 4.4, there were twelve studies with seven achieving blood pressure (BP) target estimates that conducted from the United Kingdom (the UK) and four estimates that conducted from Scotland and only one estimates from Argentina.

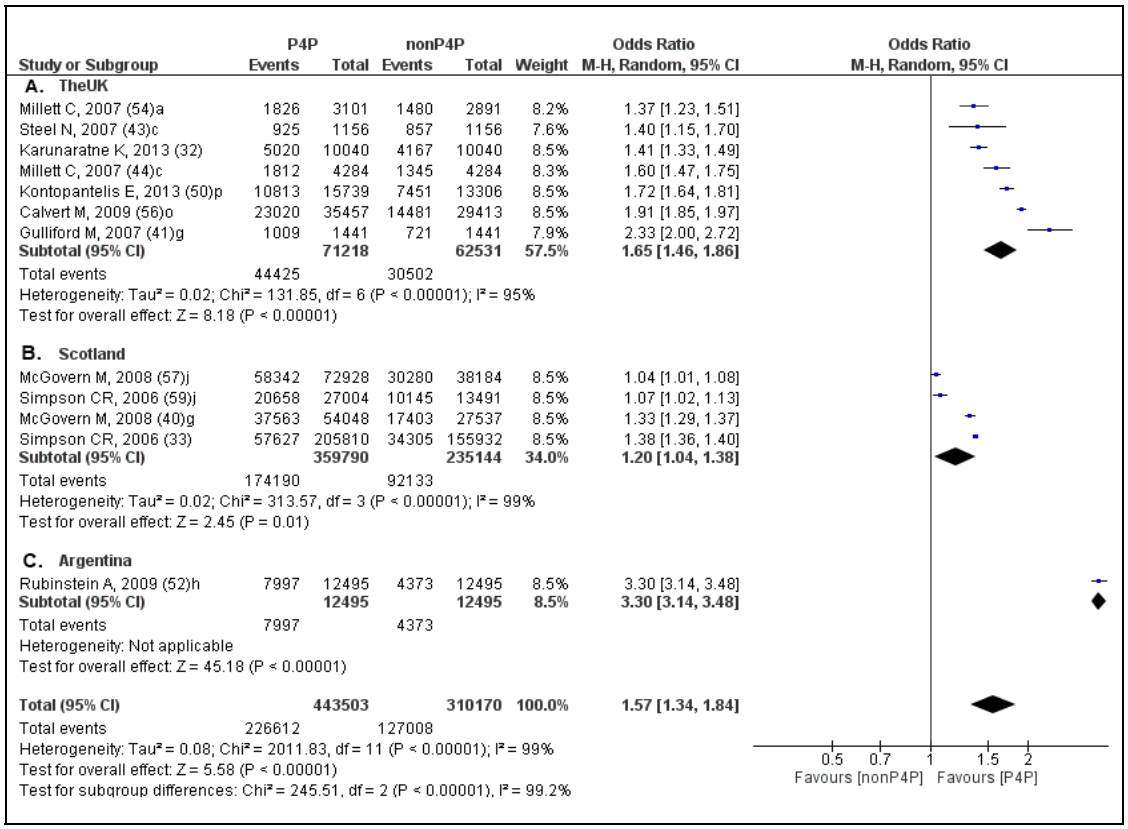


Figure 4.4 Subgroup analysis blood pressure target estimates based on database settings

P4P period had significantly greater odds of target achievement than non-P4P period. The overall Mantal-Haenszel odds ratio (OR) was

1.57; 95% CI 1.34 to 1.84; P < 0.00001, in a random effect model. Significantly heterogeneity existed among the estimates; I^2 = 99% (P < 0.00001).

Subgroup analysis demonstrated the association of each the database setting and the achievement of each separate estimates as the following details;

Figure 4.4.A reported that the BP target estimates from the UK database conducted in P4P period had significantly greater odds of target achievement than non-P4P period. The odds ratio (OR) was 1.65; 95% CI 1.46 to 1.86; P < 0.00001, random effect model. The heterogeneity was decreased but still remained; I^2 = 95% (P < 0.00001).

Figure 4.4.B demonstrated the BP target estimates from Scotland database conducted in P4P period had significantly greater odds of target achievement than non-P4P period. The odds ratio (OR) was 1.20; 95% CI 1.04 to 1.38; P=0.01, random effect model. Significant heterogeneity existed among the estimates; I^2 = 99% (P < 0.00001)

Figure 4.4.C showed the BP target estimates from Argentina database conducted in the P4P period had significantly greater odds of target achievement than the non-P4P period. The odds ratio (OR) was 3.30; 95% CI 3.14 to 3.48; P< 0.00001, random effect model. Heterogeneity was not applicable.

4) Subgroup analysis of the achieving blood pressure targets based on diseases

Figure 4.5, there were twelve studies with five achieving blood pressure (BP) target estimates that collected from the patient with diabetes mellitus (DM), three estimates collected from hypertensive patents (HT), two estimates from patients with coronary heart disease (CHD) and other two estimates from different two diseases, chronic kidney disease (CKD) and stroke (ST).

P4P period had significantly greater odds of target achievement than non-P4P period. The overall Mantal-Haenszel odds ratio (OR) was 1.57; 95% CI 1.34 to 1.84; P < 0.00001, in a random effect model. Significantly high heterogeneity existed among the estimates; I^2 = 100% (P < 0.00001).

Subgroup analysis demonstrated the association of each the diseases and their BP target estimates the following details;

Figure 4.5.A reported that the DM patients estimates collected in P4P period had significantly greater odds of target achievement than non-P4P period. The odds ratio (OR) was 1.74; 95% CI 1.45 to 2.09; P < 0.00001, random effect model. The heterogeneity was decreased but still remained; I^2 = 99% (P < 0.00001).

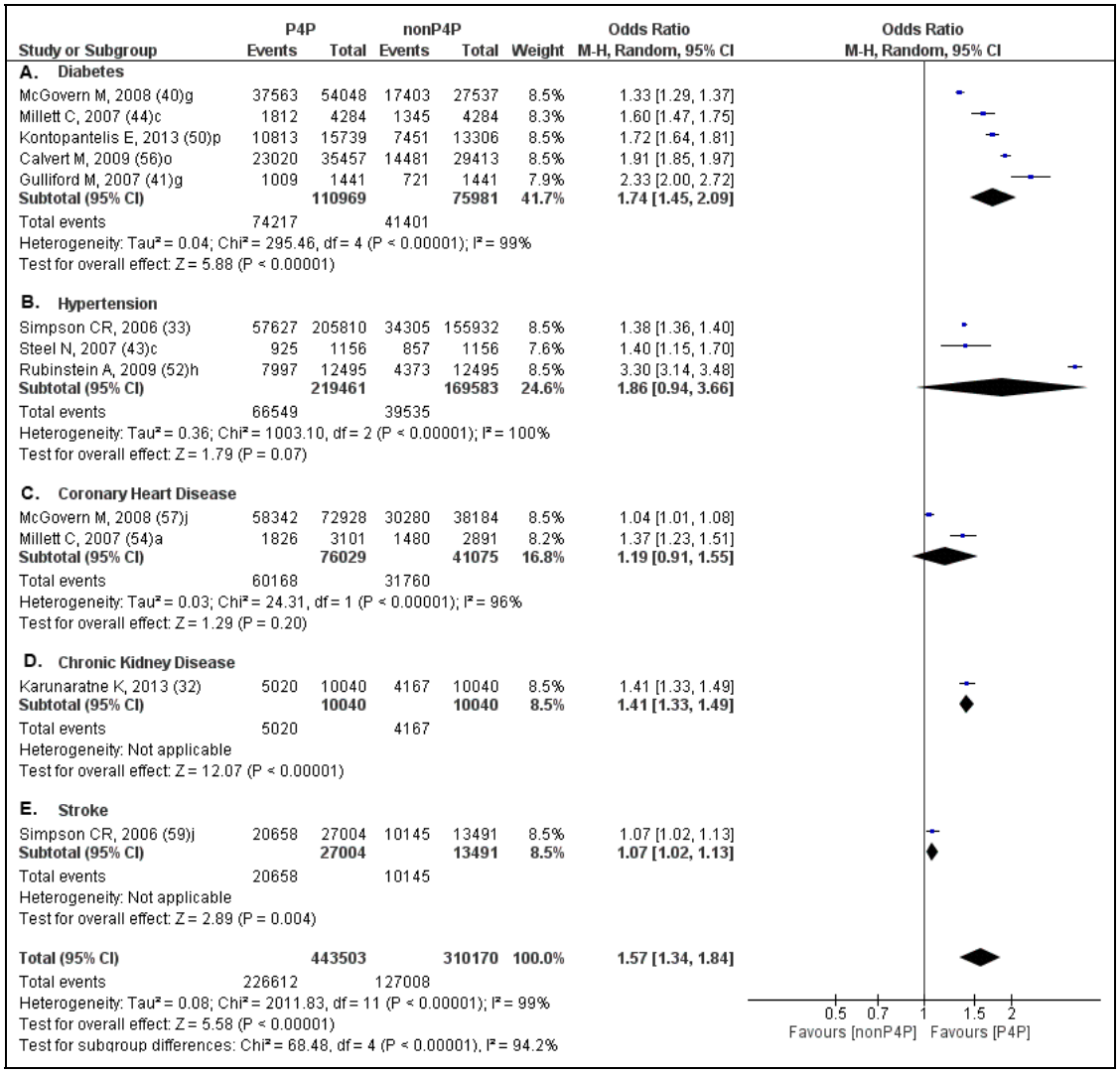


Figure 4.5 Subgroup analysis blood pressure target estimates based on diseases

Figure 4.5.B presented the HT patients estimates were not significantly different between collected in P4P period and non-P4P period.

The odds ratio (OR) was 1.86; 95% CI 0.94 to 3.66; P= 0.07, random effect model. Significant heterogeneity existed; I^2 = 100% (P < 0.00001).

Figure 4.5.C presented the CHD patients estimates were not significantly different between collected in P4P period and non-P4P period. The odds ratio (OR) was 1.19; 95% CI 0.91 to 1.55; P= 0.20, random effect model. Significant heterogeneity existed; I^2 = 96% (P < 0.00001).

Figure 4.5.D showed the CKD blood pressure estimates collected in P4P period had significantly greater odds of BP achievement than non-P4P period. The odds ratio (OR) was 1.41; 95% CI 1.33 to 1.49; P< 0.00001, random effect model. Heterogeneity was not applicable.

Figure 4.5.E showed the ST blood pressure estimates collected in P4P period had significantly greater odds of BP achievement than non-P4P period. The odds ratio (OR) was 1.07; 95% CI 1.02 to 1.13; P< 0.00001, random effect model. Heterogeneity was not applicable.

There were various studies and different separate estimates. The publication bias of each variables comparison was evaluated with funnel plots in Figure 4.6

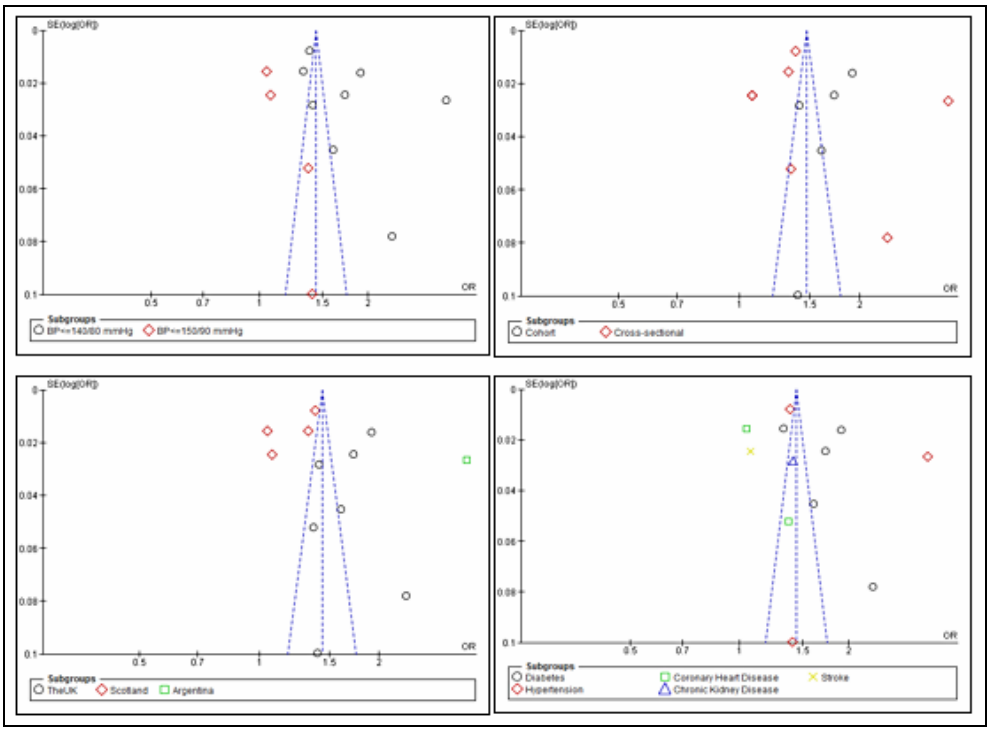


Figure 4.6 Funnel plot for blood pressure target achievements

The three blood pressure target estimates were distributed symmetrically within the 95% CIs while the others showed asymmetrical, reporting significant publication bias.

2. Comparison of the cholesterol target estimates.

There were three groups of selected variables as the following details;

1) The methodological difference between the studies.
    A. Cohort study design.
    B. Cross-sectional study design.
2) The database setting/country.
    A. The United Kingdom (UK)
    B. Scotland
    C. Argentina.
3) The disease/condition
    D. Diabetes Mellitus (DM)
    E. Coronary Heart Disease (CHD)
    F. Stroke (ST)
    G. Hypertension. (HT)

All mentioned variables had been tested heterogeneity across their subgroup for moderator analysis. Statistics calculated by the Review Manager (RevMan) [78]. The comparison was demonstrated in Table 4.6.

Table 4.6 Subgroup analysis of the cholesterol targets estimates

Subgroups	Estimates	Participants	Effect Estimate, Odds Ratio (M-H, Random, 95% CI)	Hetero-geniety (I^2)
1) Cholesterol (study design)	**8**	**299794**	**1.68 (1.00, 2.82)**	**100%**
A. Cohort	3	101450	2.22 (1.74, 2.83)	98% *
B. Cross-sectional	5	198344	1.42 (0.81, 2.49)	100%
2) Cholesterol (setting)	**8**	**299794**	**1.68 (1.00, 2.82)**	**100%**
A. The UK	5	110324	2.62 (2.12, 3.23)	98% *
B. Scotland	3	189470	0.80 (0.49, 1.31)	100%
3) Cholesterol (diseases)	**8**	**299794**	**1.37 (1.34, 1.39)**	**100%**
A. Diabetes (DM)	5	180752	1.75 (1.72, 1.79)	100%
B. CHD	2	91644	0.64 (0.62, 0.67)	100%
C. Stroke (ST)	1	27398	1.11 (1.03, 1.19)	1study

1) Subgroup analysis of total cholesterol target estimates based on study designs

Figure 4.7, there were eight studies with three achieving cholesterol target estimates, which extracted from cohort studies and five achieving estimates that extracted from cross-sectional studies.

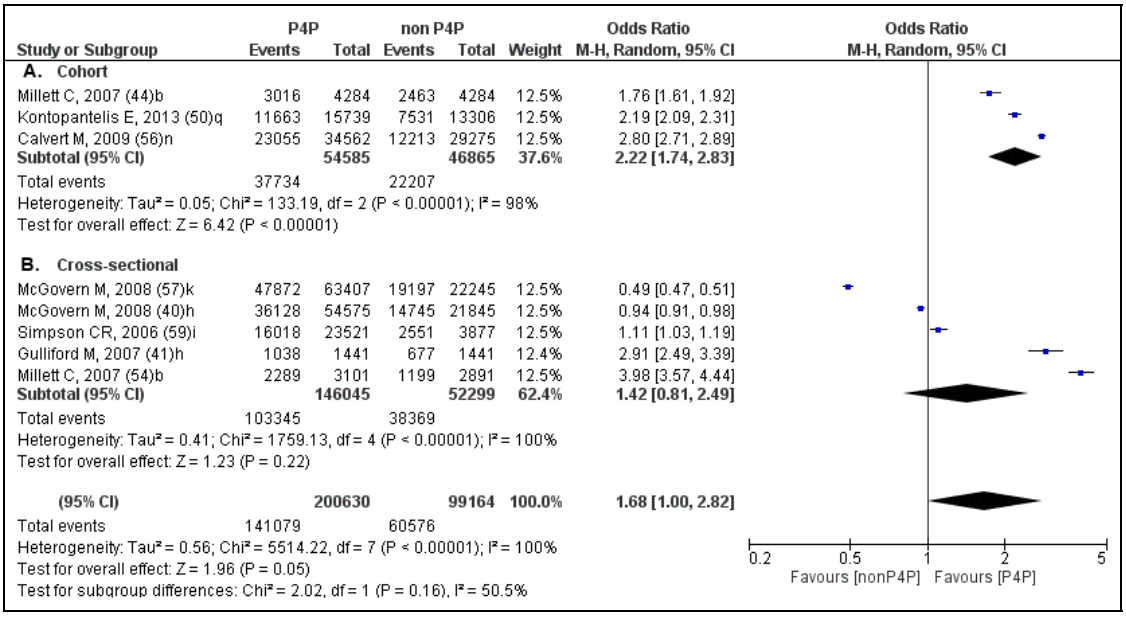


Figure 4.7 Cholesterol target estimates subgroup analysis based on study designs

There was positive association but not significantly different between pay-for-performance (P4P) intervention and the achieving total cholesterol target estimates. The overall Mantal-Haenszel odds ratio (OR) was 1.68; 95% CI 1.00 to 2.82; P=0.05, in a random effect model. Significantly heterogeneity existed among the estimates; I^2 = 100% (P < 0.00001).

Subgroup analysis demonstrated the association of each study designs and their cholesterol target estimates as the following details;

Figure 4.7.A reported that the estimates from cohort studies which performed in P4P period had significantly greater odds of cholesterol target achievement than non-P4P period. The odds ratio (OR) was 2.22; 95% CI 1.74 to 2.83; P < 0.00001, random effect model. There was reducing heterogeneity but still remained; I^2 = 98% (P < 0.00001).

Figure 4.7.B showed that the estimates from cross-sectional studies were not significantly different between P4P and non-P4P period. The odds ratio (OR) was 1.42; 95% CI 0.81 to 2.49; P=0.22, random effect model. Significant heterogeneity existed among the estimates; I^2 = 100% (P < 0.00001).

No funnel plot analysis because the relevant studies were less than ten.

2) Subgroup analysis of total cholesterol target estimates based on database setting or country.

Figure 4.8, there were eight studies with five achieving cholesterol target estimates that conducted from the United Kingdom and three achieving estimates that conducted from Scotland.

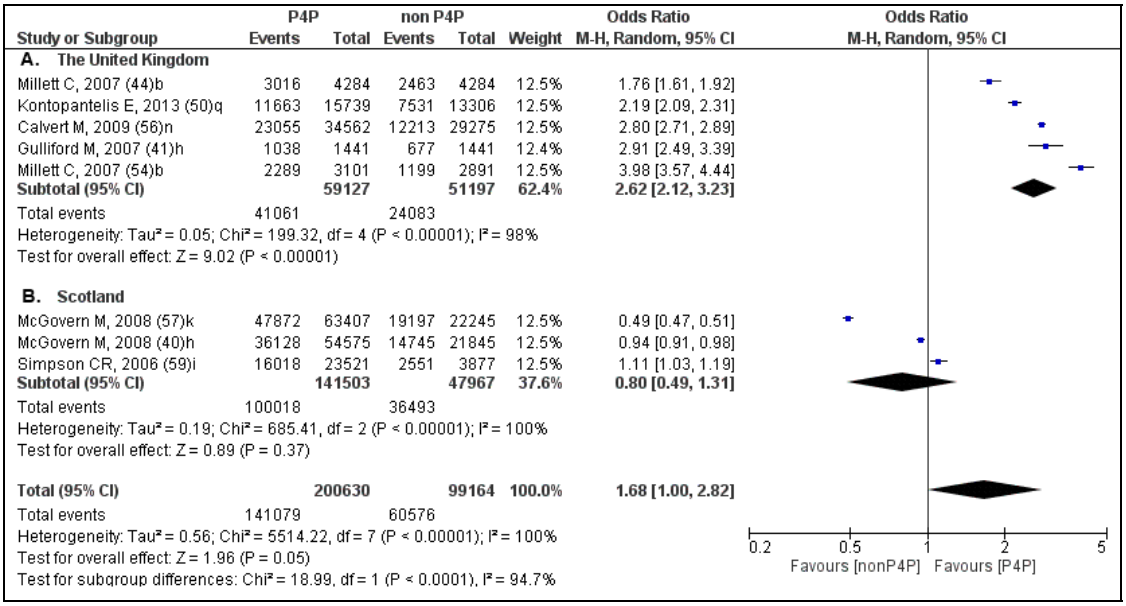


Figure 4.8 Cholesterol target estimates subgroup analysis based on database settings

There was positive association but not significantly different between pay-for-performance (P4P) intervention and the achieving total cholesterol target estimates based on database setting or country. The overall Mantal-Haenszel odds ratio (OR) was 1.68; 95% CI 1.00 to 2.82; P=0.05, in a random effect model. Significantly heterogeneity existed among the estimates; I^2 = 100% (P < 0.00001).

Subgroup analysis demonstrated the association of each database settings and their achieving cholesterol target estimates as the following details;

Figure 4.8.A reported that the estimates that from UK database in P4P period had significantly greater odds of achievement than non-P4P period. The odds ratio (OR) was 2.62; 95% CI 2.12 to 3.23; P < 0.00001, random effect model). The heterogeneity was decreased but still remained; I^2 = 98% (P < 0.00001).

Figure 4.8.B showed the estimates that from Scotland database were not significantly different between P4P and non-P4P period. The odds ratio (OR) was 0.80; 95% CI 0.49 to 1.31; P=0.22, random effect model). Significant heterogeneity existed among the estimates; I^2 = 100% (P < 0.00001).

No funnel plot analysis because the relevant studies were less than ten.

4.3.2 The comparison of clinical care recorded estimates;

1. Comparison of the blood pressure levels recorded estimates

There were three groups of selected variables as the following details;

1) The methodological difference between the studies.
    A. Cohort study design.
    B. Cross-sectional study design.
2) The database setting/country.
    A. The United Kingdom (UK)
    B. Scotland
    C. Argentina.
3) The disease/condition
    A. Diabetes Mellitus (DM)
    B. Coronary Heart Disease (CHD)
    C. Stroke (ST)
    D. Hypertension. (HT)

All above variables had been tested heterogeneity across their subgroup for moderator analysis. Statistics calculated by the Review Manager (RevMan). The comparison were demonstrated in Table 4.7.

There was very high heterogeneity (I^2) of each comparison groups. However, the decreasing of the I^2 in all subgroups were detected (*). Thus, the selected moderating variables were suitable for further subgroup analysis of the blood pressure recorded estimates.

Table 4.7 Subgroup analysis of pressure recorded estimates

Subgroups	Estimates	Participants	Effect Estimate, Odds Ratio (M-H, Random, 95% CI)	Hetero-geniety (I^2)
1) BP recorded (study design)	**12**	**611821**	**4.02 (2.71, 5.97)**	**100%**
A. Cohort	7	315418	3.01 (2.53, 3.58)	97% *
B. Cross-sectional	5	296403	6.27 (4.15, 9.48)	100%
2) BP Recorded (setting)	**12**	**611821**	**4.02 (2.71, 5.97)**	**100%**
A. The United Kingdom	8	318300	2.77 (2.33, 3.30)	97%
B. Scotland	3	268531	11.14 (9.69, 12.82)	95%*
C. Argentina	1	24990	4.43 (4.14, 4.75)	-
3) BP Recorded (conditions)	**12**	**611821**	**4.02 (2.71, 5.97)**	**100%**
A. Diabetes	5	231334	3.10 (1.52, 6.35)	100%
B. CHD	2	163411	6.32 (2.08, 19.23)	100%
C. Stroke	2	68842	7.74 [2.92, 20.51]	100%
D. Hypertension	3	148234	3.07 [2.25, 4.18]	98%*

1) Subgroup analysis of blood pressure recorded estimates based on study designs

Figure 4.9, there were four studies with seven blood pressure (BP) recorded estimates that extracted from cohort study and five estimates that extracted from cross-sectional study. The total BP recorded had significantly higher clinical recorded in P4P than non-P4P period. The overall Mantal-Haenszel odds ratio (OR) was 4.02; 95% CI 2.71 to 5.97; $P < 0.00001$, in a random effect model. Significantly heterogeneity existed among the estimates; $I^2 = 100\%$ ($P < 0.00001$).

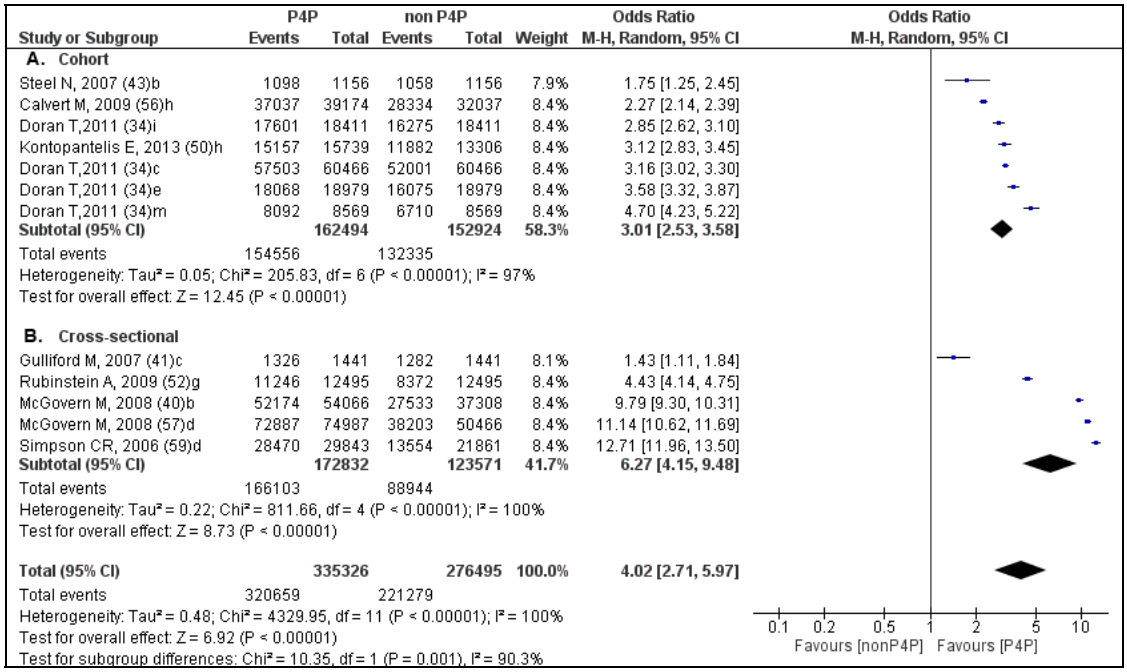


Figure 4.9 Subgroup analysis of blood press recorded estimates based on study designs

Subgroup analysis demonstrated the association of each study designs and their BP recorded estimates as the following details;

Figure 4.9.A reported that the cohort studies performed in P4P period had significantly greater odds of clinical recorded than non-P4P period. The odds ratio (OR) was 3.01; 95% CI 2.53 to 3.58; P < 0.00001, random effect model. There was reducing heterogeneity but still remained; I^2 = 97% (P < 0.00001).

Figure 4.9.B showed the cross-sectional studies performed in P4P period had significantly greater odds of clinical recorded than non-P4P period. The odds ratio (OR) was 6.27; 95% CI 4.15 to 9.48; P < 0.00001, random effect model. Significant heterogeneity existed; I^2 = 100% (P < 0.00001).

2) Subgroup analysis of blood pressure record estimates based on database settings

Figure 4.10, there were five studies with eight blood pressure recorded estimates conducted from the United Kingdom and three estimates conducted from Scotland and only one estimates from Argentina.

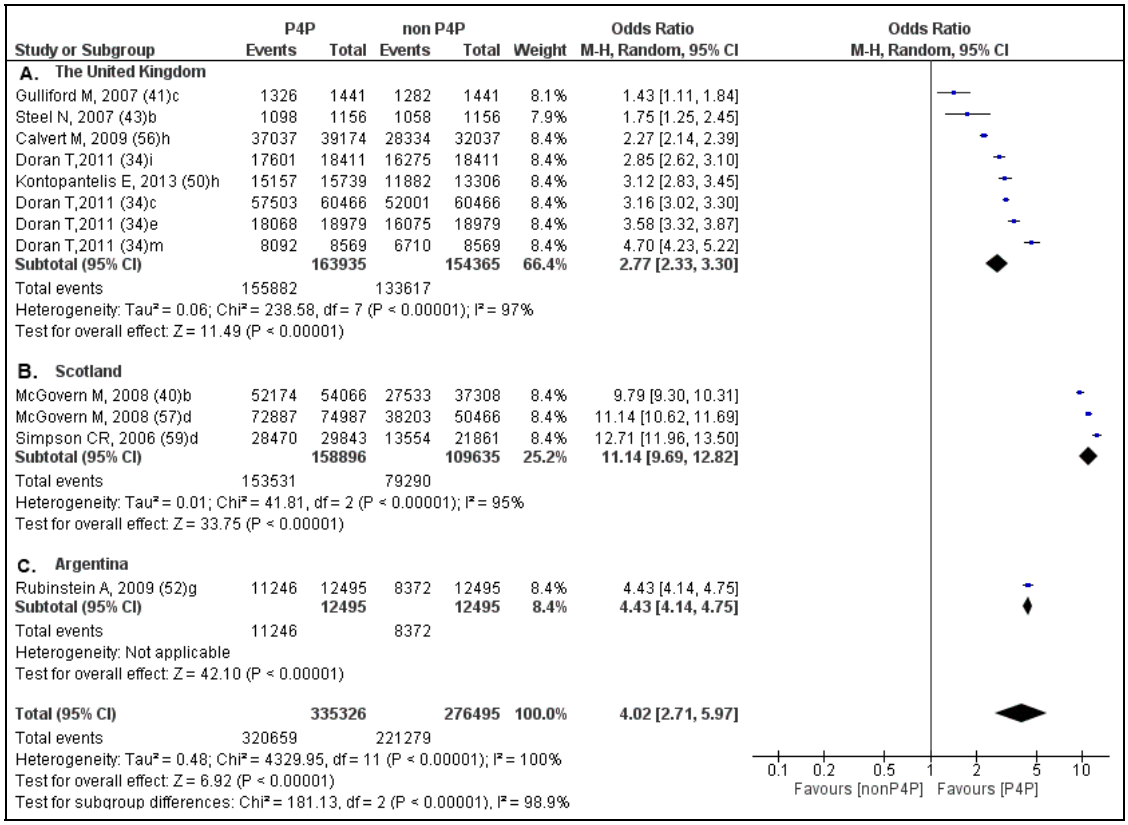


Figure 4.10 Subgroup analysis of BP recorded estimates based on database setting

P4P period had significantly greater odds of clinical recorded than non-P4P period. The overall Mantal-Haenszel Odds ratio (OR) was 4.02; 95% CI 2.71 to 5.97; P < 0.00001, in a random effect model. Significantly heterogeneity existed among the estimates; I^2 = 100% (P < 0.00001).

Subgroup analysis demonstrated the association of each database settings and their BP recorded estimates as the following details;

Figure 4.10.A reported that the BP estimates from the UK database conducted in P4P period had significantly greater odds of clinical recorded than non-P4P period. The odds ratio (OR) was 2.77; 95% CI 2.33 to 3.30; P < 0.00001, random effect model. The heterogeneity was decreased but still remained; I^2 = 97% (P < 0.00001).

Figure 4.10.B demonstrated the BP recorded estimates from Scotland database conducted in P4P period had significantly greater odds of clinical recorded than non-P4P period. The odds ratio (OR) was 11.14; 95% CI 9.69 to 12.82; P < 0.00001, random effect model. The heterogeneity was decreased but still remained; I^2 = 95% (P < 0.00001).

Figure 4.10.C showed the BP recorded estimates from Argentina database conducted in the P4P period had significantly greater odds of clinical recorded than the non-P4P period. The odds ratio (OR) was 4.43; 95% CI 4.14 to 4.75; P < 0.00001, random effect model. Heterogeneity was not applicable.

3) Subgroup analysis of blood pressure recorded estimates based on diseases

Figure 4.11, there were twelve studies with five blood pressure recorded estimates, that collected from the patient with diabetes mellitus (DM), two estimates collected from patients with coronary heart disease (CHD), two estimates from patients with stroke (ST) and three estimates collected from patients with hypertension (HT).

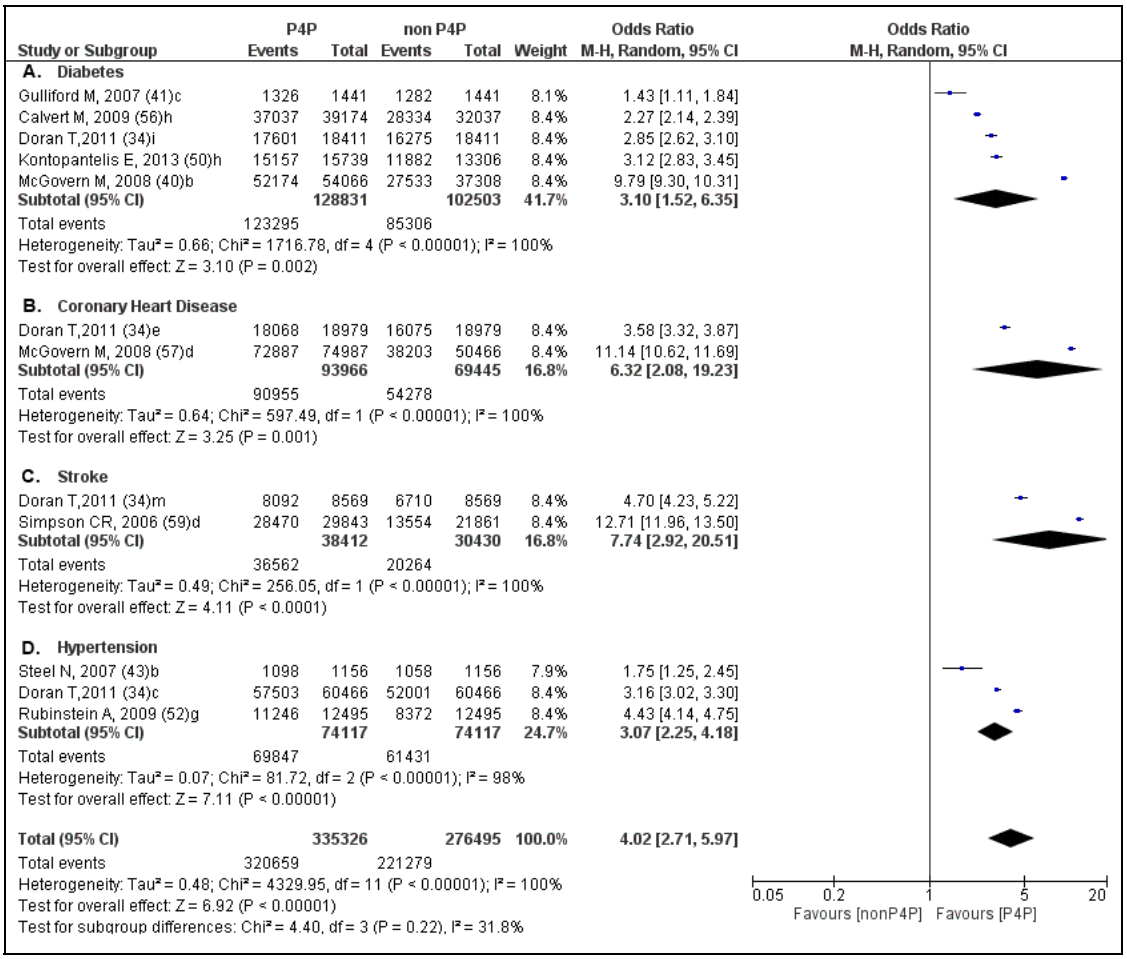


Figure 4.11 Subgroup analysis of blood pressure recorded estimates based on diseases

P4P period had significantly greater odds of clinical recorded than non-P4P period. The overall Mantal-Haenszel odds ratio (OR) was 4.02; 95% CI 2.71 to 5.97; P < 0.00001, in a random effect model. Significantly high heterogeneity existed among the estimates; I^2 = 100% (P < 0.00001).

Subgroup analysis demonstrated the association of each diseases and their BP recorded estimates as the following details;

Figure 4.11.A reported that the DM blood pressure recorded estimates collected in P4P period had significantly greater odds of clinical recorded than non-P4P period. The odds ratio (OR) was 3.10; 95% CI 1.52 to 6.35; P =0.002, random effect model. Significant heterogeneity existed; I^2 = 100% (P < 0.00001).

Figure 4.11.B presented the CHD blood pressure recorded estimates collected in P4P period had significantly greater odds of clinical recorded than non-P4P period. The odds ratio (OR) was 6.32; 95% CI 2.08 to 19.23; P= 0.001, random effect model. Significant heterogeneity existed; I^2 = 100% (P < 0.00001).

Figure 4.11.C presented the ST blood pressure recorded estimates collected in P4P period had significantly greater odds of BP recorded estimates than non-P4P period. The odds ratio (OR) was 7.74; 95% CI 2.92 to 20.51; P<0.001, random effect model). Significant heterogeneity existed; I^2 = 100% (P < 0.00001).

Figure 4.11.D presented the HT blood pressure estimates collected in P4P period had significantly greater odds of clinical recorded than non-P4P period. The odds ratio (OR) was 3.07; 95% CI 2.25 to 4.18; (P < 0.00001), random effect model). Significant heterogeneity existed; I^2 = 100% (P < 0.00001).

2. Comparison of the total cholesterol levels recorded

There were three groups of selected variables as the following details;

1) The methodological difference between the studies

A. Cohort study design.

B. Cross-sectional study design.

2) The database setting/country

A. The United Kingdom (UK)

B. Scotland

3) The disease/condition

A. Diabetes Mellitus (DM)

B. Coronary Heart Disease (CHD)

C. Stroke (ST)

All above variables had been tested heterogeneity across their subgroup for moderator analysis. Statistics calculated by the Review Manager (RevMan). The comparison were demonstrated in Table 4.8.

There was very high heterogeneity (I^2) of each comparison groups. However, the decreasing of the I^2 in some subgroup variables were detected (*); study design and database setting variables. But no difference in the disease variables, therefore, the disease variables were not suitable for further subgroup analysis.

Table 4.8 Subgroup analysis of the cholesterol recorded estimates

Subgroups	Outcomes	Participants	Effect Estimate, Odds Ratio (M-H, Random, 95% CI)	Hetero-geniety (I^2)
1) Cholesterol recorded (Study design)	**9**	**466291**	**5.27 (3.71, 7.51)**	**100%**
A. Cohort	5	192175	3.86 (3.16, 4.70)	98% *
B. Cross-sectional	4	274116	8.02 (6.26, 10.28)	99%
2) Cholesterol recorded (Setting)	**9**	**466291**	**5.27 (3.71, 7.51)**	**100%**
A. The United Kingdom	6	195057	3.87 (3.23, 4.64)	98% *
B. Scotland	3	271234	9.71 (.46, 12.64)	99%

Table 4.8 Subgroup analysis of the cholesterol recorded estimates (cont.)

Subgroups	Outcomes	Participants	Effect Estimate, Odds Ratio (M-H, Random, 95% CI)	Hetero-geniety (I^2)
3) Cholesterol recorded conditions)	**9**	**466291**	**5.27 (3.71, 7.51)**	**100%**
A. Diabetes	5	235184	4.50 (2.37, 8.55)	100%
B. CHD	2	162303	5.49 (2.85, 10.56)	100%
C. Stroke	2	68804	7.50 (3.92, 14.35)	100%

1) Subgroup analysis of cholesterol recorded estimates based on study designs

Figure 4.12, there were seven studies with five cholesterol recorded estimates that collected from cohort study design and four estimates that collected from cross-sectional study design.

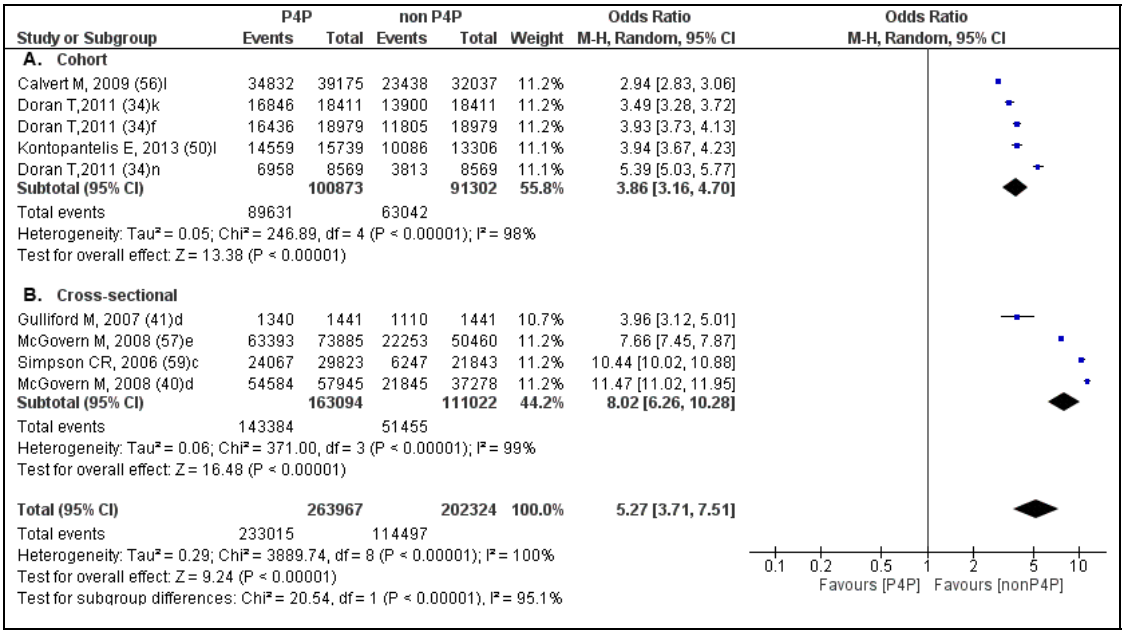


Figure 4.12 Subgroup analysis of cholesterol recorded estimates accessibility based on study designs

The total cholesterol recorded had significantly higher clinical recorded in P4P than non-P4P period. The overall Mantal-Haenszel

Odds ratio (OR) was 5.27; 95% CI 3.71 to 7.51; P < 0.00001, in a random effect model. Significantly heterogeneity existed among the estimates; I^2 = 100% (P < 0.00001).

Subgroup analysis demonstrated the association of each study designs and their cholesterol recorded as the following details;

Figure 4.12.A reported that the estimates from the cohort studies which retrieved in P4P period had significantly greater odds of clinical recorded than non-P4P period. The odds ratio (OR) was 3.86; 95% CI 3.16 to 4.70; P < 0.00001, random effect model. There was reducing heterogeneity but still remained; I^2 = 98% (P < 0.00001).

Figure 4.12.B showed the estimates from the cross-sectional studies which retrieved in P4P period had significantly greater odds of clinical recorded than non-P4P period. The odds ratio (OR) was 8.02; 95% CI 6.26 to 10.28; P < 0.00001, random effect model. There was significant heterogeneity; I^2 = 99% (P < 0.00001).

No funnel plot analysis because the relevant studies were less than ten.

2) Subgroup analysis of cholesterol record estimates based on database setting or country

Figure 4.13, There were seven studies with six cholesterol recorded estimates conducted from the United Kingdom and three estimates conducted from Scotland.

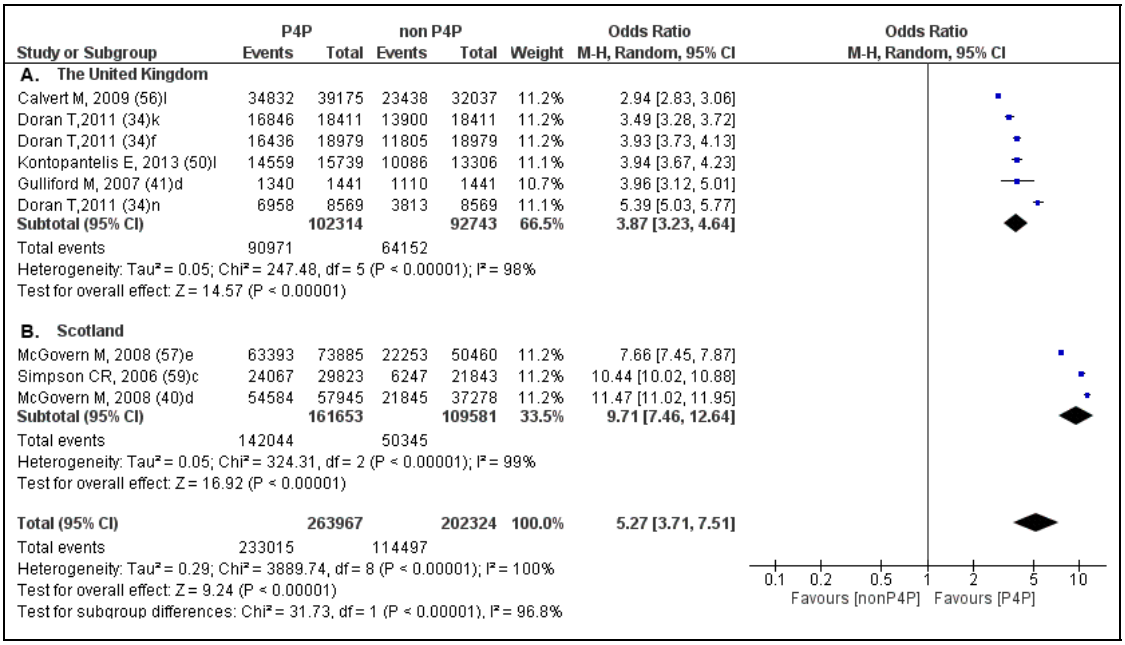


Figure 4.13 Subgroup analysis of cholesterol record estimates based on database setting

P4P period had significantly greater odds of clinical recorded estimates than non-P4P period. The overall Mantal-Haenszel odds ratio (OR) was 5.27; 95% CI 3.71 to 7.51; P < 0.00001, in a random effect model. Significantly heterogeneity existed among the estimates; I^2 = 100% (P < 0.00001).

Subgroup analysis demonstrated the association of each database settings and their cholesterol recorded estimates as the following details; Figure 4.13.A reported that the cholesterol recorded estimates from the UK database conducted in P4P period had significantly greater odds of clinical recorded than non-P4P period. The odds ratio (OR) was 3.87; 95% CI 3.23 to 4.64; P < 0.00001, random effect model. The heterogeneity was decreased but still remained; I^2 = 98% (P < 0.00001).

Figure 4.13.B reported that the cholesterol recorded estimates from Scotland database conducted in P4P period had significantly greater odds of clinical recorded than non-P4P period. The odds ratio (OR) was 9.71; 95% CI 7.46 to 12.64; P < 0.00001, random effect model. The heterogeneity was decreased but still remained; I^2 = 99% (P < 0.00001).

No funnel plot analysis because the relevant studies were less than ten.

4.3.3 The supplementary comparison of the effect estimates.

The effect estimates reported dichotomous outcomes. Most of them demonstrated that the achievement in the post-P4P period had a better outcome when compared to the pre-P4P period. When P4P scheme implemented in context of the lower baseline of performance shown more improvements than in the higher baseline. There were the associations of percentage achievement in pre-P4P period and the percentage change after P4P implementation as depicted in Figure 4.14

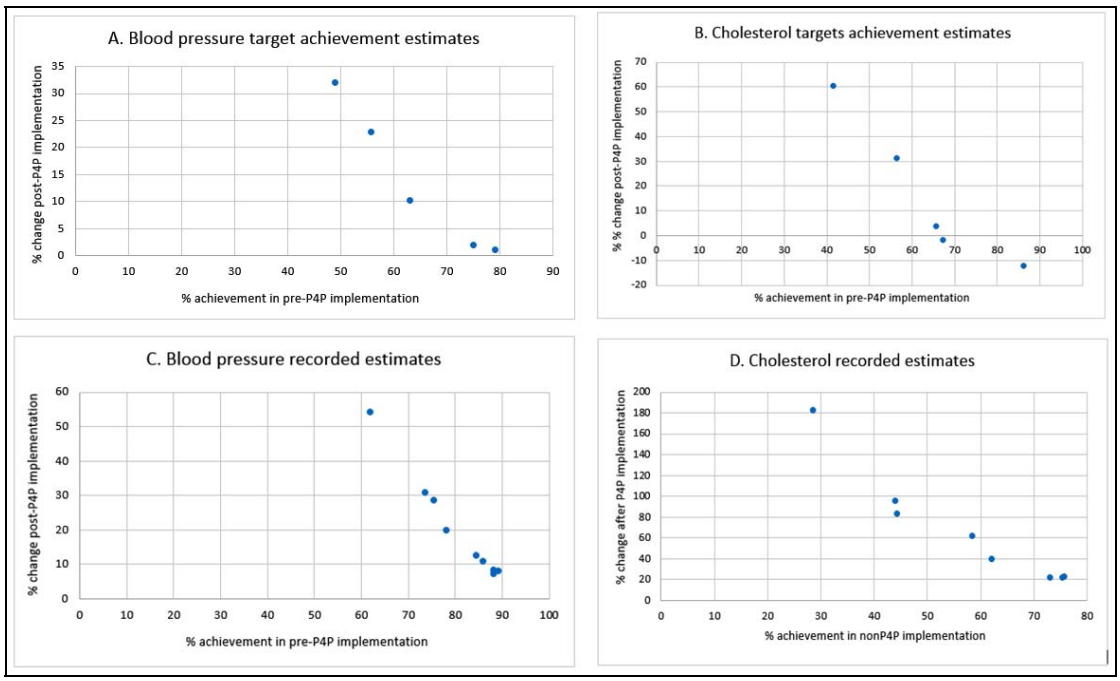


Figure 4.14 The percentage change post-P4P implementation (y-axis) when pre-P4P implementation were baselines (x-axis)

Moreover, there are three included studies; Kontopantelis E. [50], Calvert M. [56], Doran T. [34] that contained the data from more than one year before and after P4P implementation in diabetes patients. In Figure 4.15, the year 2001-2003 period was before P4P implementation, the year 2004 was the initial year of implementation and the year 2005-2007 period was after P4P implementation.

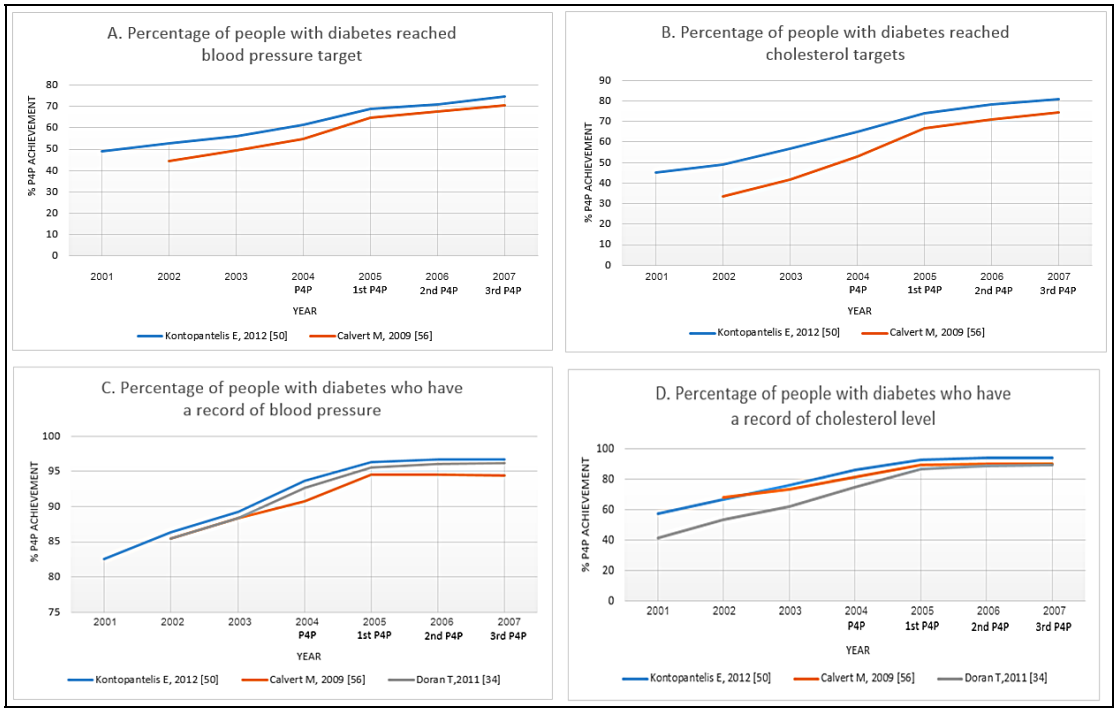


Figure 4.15 Percentage of diabetes patients achieving quality targets

There was the highest slope in the first year of P4P implementation (2004-2005), which showed the increase of blood pressure recorded in the diabetes patients. In the second year (2005-2006) and the third year (2006-2007) of P4P implementation, the slope was declined but remained significantly improvement. The P4P implementation enhances significant increasing in both quality improvements and the accessibility of clinic care in the first three year of P4P implementation.

CHAPTER V
DISCUSSION

5.1 Research findings

This review provides the effects of P4P implementation and the conditions associated with the achievement of pay-for-performance (P4P) that summarized from the whole twenty-three included studies.

5.1.1 The supportive compositions of pay-for-performance (P4P) scheme.

To our findings, the three considerable supportive compositions of P4P implementation were summarized as the following details:

(1) Appropriate information technology (IT) infrastructure

According to Table 4.2, there were fourteen studies conducted from the UK and Scotland presented that they collected and submitted their pay-for-performance (P4P) data electronically. In the United kingdom (UK) and Scotland, a P4P system has been called the Quality and Outcomes Framework (QOF). The QOF is a financial incentive with one national agreed set and a single payer; it is covering primary care for ten chronic diseases, organization management, patient satisfaction and the other specific priorities [88]. The program establishes groups of quality indicators that the general practitioners (GP) have to reach for their score points which according to the guideline recommendations. Most of the measurements were process measures. In addition, performance measurement is based on data recorded from the electronic medical record (EMR), they use a computer program for showing the target indicators to the GP. The information technology (IT) system accelerated data collection, facilitated the GPs for call and recall their patients, provided an update of score points achieved and prompted the GPs with computer pop-up boxes for added any requirement activities [89].

Therefore, the substantial supporting system for P4P implementation is developing the information technology (IT) infrastructure for payment purposes.

(2) Incentive payments should base on teamwork and explicit measurements related to health outcomes of the patients

According to Table 4.2, the study from Italy [51], Argentina [52] and four studies conducted from Taiwan [46,47,48,49] showed significant improvement of health outcomes.

The study which conducted in Italy [51], demonstrated that the P4P hospital tended to increased quality of care for aging patients with hip fracture. Particularly, when added P4P to the Diagnosis Related Group payment system.

The study from Argentina [52], reported that the multimodal intervention based on P4P implementation had significant improvement in most of the explicit quality indicator targets.

The four studies conducted from Taiwan [46,47,48,49] with two studies about P4P in tuberculosis, and other two studies reported P4P in breast cancer and diabetes. All four studies from Taiwan showed significant improvement of health outcomes and accessibility of care services. Taiwan introduced a nationwide, single pay financial incentive covered 99% of Taiwan's population in 2001. The Bureau of National Health Insurance (BNHI) implemented the pay-for-performance (P4P) in the initiative five particular diseases; diabetes, breast cancer, cervical cancer, tuberculosis and asthma. The P4P quality measures of diabetes and asthma based on process measures according to the explicit measures in the guideline recommendations, and the breast cancer based on outcome measures such as recurrent and survival rate. The participants were voluntary care providers in primary health care facilities such as hospital, clinic and other health service settings. For participation, a volunteer has to meet formal qualification or certification medical requirements, treatment by follow clinical guideline recommendations and set up appropriate data collection [90].

Hence, the important supporting systems for P4P implementation are team-base treatment and the explicit measurements in the guideline recommendations for improving quality and health outcomes of the patients.

(3) Variety of payment incentive plans for different target population was provided

According to Table 4.2, the study from Canada [66] and Australia [85], reported the variety of payment incentive plans for different target population.

The study which performed in Canada [66] reported that Canada implemented P4P only at Ontario's health care system since 2004. The P4P scheme has been called "Family Health Network" targeted in the following areas; preventive health care, severe mental illness care, chronic disease management, after-hour care, smoking cessation and leadership-group management. In 2008, Ontario began a pay-for-result for decreasing of waiting times. Currently, there is not active P4P implementation in other provinces of Canada [88].

The study which conducted in Australia [85], presented that the general practices under the General Practice Management Plan (GPMP), Medicare Australia significantly increased clinical recording followed the process measure recommendations for diabetes patients.

In 1997, Australia launched the General Practice Immunisation Incentive scheme for primary health care. After that in 1998, the Medicare Australia paid bonus payments to the GPs under the Practice Incentives Program (PIP-1). In 2006, Veteran's Affairs Department of the Commonwealth paid P4P financial incentive for the hospital that served veterans according to the established quality targets. Recently, in 2007 Medicare Australia drove the Practice Incentives Program (PIP-2) focused on GP practices [91]. Furthermore, Australia has a pilot P4P indicators called the "Clinical Practice Improvement Payment" (CPIP) for the gap areas in care [82]. There was the only one included study which conducted in Australia [85], reported that the general practices under the General Practice Management Plan (GPMP), Medicare Australia significantly increased testing according to the process measure recommendations for diabetes patients.

Thus, the one of prudential supporting systems for P4P implementation is the available of various payment incentive plans for different target population.

5.1.2 The effectiveness of pay-for-performance (P4P) implementation.

P4P scheme implemented in context of the lower baseline of performance shown more improvements than in the higher baseline. There is some accordant evidence from the previous studies, for example, Peterson et al. [11] suggested that the lower baseline performance may better achieve to quality targets. The policy maker has to consider the aim of P4P implementation, Rosenthal et al. [92] reported that after P4P implementation, the former lower performance groups of physician achieved the quality targets the most while the previous higher performance improved the least.

P4P is probably efficient incentive scheme for the low productive health care service areas.

5.1.3 The effective period of pay-for-performance (P4P) implementation

P4P scheme could enhance a steady increase in both quality of services and accessibility in the first three years of implementation. The increasing rates were declined after that, but still shown improvements. However, there were some conceivable explanations for the high growth rate in the first year, but not sustained. In the first year of P4P implementation, the health care providers might be enthusiastic in the new incentive scheme and also tended to follow the clinical recommendations. After that when the set of targets were reached above the established thresholds, there were no more financial incentives for further improvement. Thus, P4P scheme may be considered to evaluate the outcomes.

There were several studies supported that the P4P encouraged the improving performance in a short period; MacBride-Stewart S [37] suggested that the prescribing of relevant drugs increased before the P4P implementation in 2004; the growth existed in the first two years but in the third year of P4P, and the significant improvement was at a lower level. McGovern M. [40] presented that in the first year of P4P period, the patients with diabetes who have an electronic recorded increase 54.2%. Moreover, the recorded of blood pressure, glycated hemoglobin (HbA1c), cholesterol and serum creatinine improve substantially. Cheng S. [49] reported that the diabetes intermediate outcome examinations and tests increased meaningfully in the P4P program, and the improvement dropped in the second and third year but remained more achievements than in non-P4P program, Campbell S. [53] suggested that in the

first two year of the P4P implementation, the quality of care in the patients with asthma and diabetes increased but not for the heart disease patients

5.2 The limitation of the study

There were a lot of limitations of this study as below.

First, P4P implementation in the universal health coverage context (UHC) was introduced nationally. Therefore, all of our included studies are non-randomized studies and most of them reported positive effects to the P4P implementation. Therefore, they did not have control over the implementation of P4P, and they could not use a randomized controlled trial to assess their P4P effectiveness. There was a limited of methodological design using in the studies that relevant to P4P scheme in the UHC context and may lead to an over-estimate of the effectiveness of P4P. Schatz [12] discussed that the non-randomized studies generally reported positive effects of the intervention.

Second, the included study limited to English language and retrieved from only five databases; the Cochrane Library, PUBMED, MEDLINE, EBSCO host service and CINAHL within the same timeframe. So, they did not provide any various publications.

Third, this study did not add unpublished documents and gray literatures in the review. Thus, there were substantial publication bias.

Finally, because of the large variety in the variables across the included studies. The data reported high heterogeneity in this systematic review, it was not ordinary to perform any statistical meta-analysis on the non-homogenous estimate data. Therefore, this study performed the supplementary comparison of the extracted effect estimates of P4P implementation. There were the associations of percentage achievement in non-P4P period and the percentage change after P4P implementation that the lower baseline of performance shown more improvements, within a short period. This study make it clear that P4P probably appropriate for activate the unproductive area of health care systems with continuous evaluations.

CHAPTER VI
CONCLUSION

6.1 Conclusion

Pay-for-performance (P4P) incentive scheme is one of the effective strategies that tend to increase quality improvement of healthcare system with universal health coverage (UHC) in some conditions, as the following;

1. P4P implementation most likely encourages the achievement of quality indicator targets and the recording of quality indicators in the lower baseline better than, the higher baseline.

2. P4P implementation enhances the improvement of quality care in first three years of the implementation, and the growth may decline but remain significant improvement.

3. There are three supporting systems for the effectiveness of P4P scheme;

(1) Appropriate information technology (IT) infrastructure for the payment purposes.

(2) Incentive payments should base on teamwork and explicit measurements related to health outcomes of the patients.

(3) Variety of payment incentive plans for different target population should be provided.

P4P is probably appropriate to implement in the low productive health care service areas such as health services for the underprivileged population. Furthermore, the effectiveness of P4P last for a reliable period; after that a new incentive scheme may be considered to boost the outcomes.

6.2 Recommendations

The recognitions that associate with the effectiveness of Pay-for-performance (P4P) implementation in the health care system based on universal health coverage context (UHC) are the following;

1) The ultimate goals of P4P implementation.

All policy makers have to evaluate the baseline performance of their healthcare systems in order to scope the scope the unproductive sectors that need the improvement.

2) The accurate measurements.

P4P implementation most likely improved in the explicit quality indicators and clear process measurements.

3) The payment mechanism.

The principal mechanism for potential P4P effectiveness is the information technology (IT) infrastructure on payment purposes.

4) Monitoring and evaluating the P4P incentive scheme.

P4P needs continuous monitoring and evaluating in order to reduce unintended consequences and maintain productivity in long term of the implementation.

REFERENCES

1. Carrin G, Hanvoravongchai P. Provider payments and patient charges as policy tools for cost-containment: How successful are they in high-income countries? Human Resources for Health. (2003, Jan); 11-10. Available from: Academic Search Complete.
2. OECD (2013), Health at a Glance 2013: OECD Indicators, OECD Publishing. [Retrieved May 23, 2014]. Available from: http://www.oecd.org/els/health-systems/Health-at-a-Glance-2013.pdf
3. Scheffler RM, Pay for Performance (P4P) Programs in Health Services: What is the Evidence? World Health Report (2010) Background Paper, 31; 2010. [Retrieved March 3, 2014]. Available from: http://www.who.int/healthsystems/topics/financing/healthreport/P4PWHR2010SheffIerFINAL.pdf
4. Cromwell J, Trisolini MG, Pope GC, Mitchell JB and Greenwald LM. Pay for Performance in Health Care: Methods and Approaches. RTI Press publication No. BK-0002-1103. Research Triangle Park, NC: RTI Press. (2011). [Retrieved March 3, 2014]. Available from: http://www.rti.org/rtipress.
5. The Organisation for Economic Co-operation and Development. OECD Health Policy Studies: Value for Money in Health Spending. Organisation for Economic Co-operation and Development; 2010 [Retrieved March 3, 2014]. Available from: http://www.consorci.org/accessos-directes/patronal/documents-i-publicacions/articles-i-publicacions-d2019interes/value%20for%20money%20in%20health.pdf
6. Jankhotkaew J, Schröder-Bäck P, Brand H, Pongsupap Y, Pratoomnun S. An evaluation of universal health care coverage among low income groups in Thailand by using four selected characteristics of good governance. Journal of Health Systems Research 2011; 5(2): 195-204.

7. Wibulpolprasert S. Inequitable Distribution of Doctors: Can it be Solved? Human Resources for Health Development Journal (HRDJ) (1999 Jan-Apr); 3(1): 2-39.
8. Ministry of Public Health of Thailand. Pay for Performance: P4P, (2556); [cited March 3, 2014]. Available from: http://www.moph.go.th/ops/oic/data/20131016151652_1_.pdf
9. Kutzin J. Health financing for universal coverage and health system performance: concepts and implications for policy. Bulletin of The World Health Organization. (2013, Aug); 91(8): 602-611. Available from: CINAHL Plus with Full Text.
10. Houle S, McAlister F, Jackevicius C, Chuck A, Tsuyuki R. Does Performance-Based Remuneration for Individual Health Care Practitioners Affect Patient Care? Annals of Internal Medicine. (2012, Dec 18); 157(12): 889-899. Available from: Academic Search Complete.
11. Petersen L, Woodard L, Urech T, Daw C, Sookanan S. Does Pay-for-Performance Improve the Quality of Health Care? Annals of Internal Medicine. 2006, Aug 15); 145(4): 265-W71. Available from: Academic Search Complete.
12. Schatz M. Does pay-for-performance influence the quality of care? Current Opinion in Allergy & Clinical Immunology. (2008, June); 8(3): 213-221. Available from: CINAHL Plus with Full Text.
13. Bruin S, Baan C, Struijs. Pay-for-performance in disease management: a systematic review of the literature. BMC Health Services Research. (2011, Jan 2); 11(Suppl 1): 272-285. Available from: Academic Search Complete.
14. Eijkenaar F, Emmert M, Scheppach M, Schöffski O. Effects of pay for performance in health care: A systematic review of systematic reviews. Health Policy. (2013, Jan 1); Available from: Scopus®.
15. McDonald R, Roland M. Pay for performance in primary care in England and california: comparison of unintended consequences. Annals of Family Medicine. (2009, Mar); 7(2): 121-127. Available from: CINAHL Plus with Full Text.

16. Van Herck P, De Smedt D, Annemans L, Remmen R, Rosenthal M, Sermeus W. Systematic review: Effects, design choices, and context of pay-for-performance in health care. BMC Health Services Research. (2010, Jan); 10:247-259. Available from: Academic Search Complete.
17. Stuckler D, Feigl AB, Basu S, Mckee M. The political economy of universal health coverage: Background paper for the global symposium on the health systems research. World Health Organization; 2010. [Retrieved March 3, 2014]. Available from: http://www.pacifichealthsummit.org/downloads/UHC/the%20political%20economy%20of%20uhc.PDF
18. The World Bank. Universal Health Coverage Study Series. World Bank; 2014. [Retrieved March 3, 2014]. Available from: http://www.worldbank.org/en/topic/health/publication/universal-health-coverage-study-series.print
19. The World Health Report 2010: Health Systems Financing: The path to universal coverage. Geneva: World Health Organization; 2010. [Retrieved March 3, 2014]. Available from: http://www.who.int/healthsystems/topics/financing/healthreport/en/
20. Gulliford M, Figueroa-Munoz J, Morgan M, Hughes D, Gibson B, Hudson M, et al. What does' access to health care' mean? Journal of Health Services Research & Policy. (2002, July); 7(3): 186-188. Available from: MEDLINE.
21. The World Health Organization. Quality of care: a process for making strategic choices in health systems. Geneva: World Health Organization; 2006 [Retrieved March 3, 2014]. Available from: http://www.who.int/management/quality/assurance/QualityCare_B.Def.pdf
22. True Cost – Analyzing our economy, government policy, and society through the lens of cost-benefit: List of Countries with Universal Healthcare. [Retrieved March 3, 2014]. Available from: http://truecostblog.com/2009/08/09/countries-with-universal-healthcare-by-date/#link1
23. O'Dougherty S, Cashin C, Langenbrunner J, World B. Designing and Implementing Health Care Provider Payment Systems: How-to Manuals [monograph on the Internet]. Washington, D.C.: World Bank; 2009. Available from: Discovery eBooks.

24. Batalden P, Davidoff F. What is "quality improvement" and how can it transform healthcare? Quality and Safety In Health Care. (2007, Feb 1); 16(1): 2-3. Available from: Scopus®.
25. Eichler R and Susna D. United States Agency for International Development (USAID). May 2008. Paying for Performance in Health Guide to Developing the Blueprint. Bethesda, MD: Health Systems 20/20 project, Abt Associates Inc.
26. Miller G, Babiarz KZ. Pay-for-performance incentives in low- and middle-income country health programs. National Boueau of Economic Research Working Paper No. 18932. (2013, April), [Retrieved March 3, 2014]. Available from: http://www.stanford.edu/~ngmiller/w18932.pdf
27. Donabedian A. Evaluating the Quality of Medical Care. Milbank Quarterly [serial on the Internet]. (2005, Dec); 83(4): 691-729. Available from: Business Source Complete.
28. Agency for Healthcare Research and Quality (AHRQ). Pay for Performance: A Decision Guide for Purchasers. [Retrieved March 8, 2014]. Available from: http://www.ahrq.gov/professionals/quality-patient-safety/quality-resources/tools/p4p/p4pguide.pdf
29. Fenter T, Lewis S. Pay-for-Performance Initiatives. J Manag Care Pharm. (2008); 14(6); (suppl S-c): S12-S15
30. The World Bank. How are providers (hospitals, clinics, managers, physicians, nurses, etc.) paid? World Bank; [Retrieved March 9, 2014]. Available from: http://web.worldbank.org/WBSITE/EXTERNAL/TOPICS/EXTHEALTHNUTRITIONANDPOPULATION/EXTHSD/0,,contentMDK:22516696~menuPK:6485077~pagePK:148956~piPK:216618~theSitePK:376793~isCURL:Y,00.html
31. Eichler R, Levine R. Performance incentives for global health: potential and pitfalls / Rena Eichler, Ruth Levine and the Performance-Based Incentives Working Group [monograph on the Internet]: Washington, D.C.: Center for Global Development, c2009; 2009. Available from: Harvard Library Bibliographic Dataset.

32. Karunaratne K, Stevens P, Irving J, Hobbs H, Kilbride H, Farmer C, et al. The impact of pay for performance on the control of blood pressure in people with chronic kidney disease stage 3–5. Nephrology Dialysis Transplantation. (2013, Aug); 28(8): 2107-2116. Available from: Academic Search Complete.
33. Simpson C, Hannaford P, Ritchie L, Sheikh A, Williams D. Impact of the pay-for-performance contract and the management of hypertension in Scottish primary care: a 6-year population-based repeated cross-sectional study. The British Journal of General Practice: The Journal of the Royal College Of General Practitioners. (2011, July); 61(588): e443-e451. Available from: MEDLINE.
34. Doran T, Kontopantelis E, Campbell S, Reeves D, Valderas J, Salisbury C, et al. Effect of financial incentives on incentivised and non-incentivised clinical activities: Longitudinal analysis of data from the UK Quality and Outcomes Framework. Bmj [serial on the Internet]. (2011, July 9), [cited June 15, 2014]; 343(7814): Available from: Scopus®.
35. Vamos E, Pape U, Bottle A, Hamilton F, Ng A, Molokhia M, et al. Association of practice size and pay-for-performance incentives with the quality of diabetes management in primary care. Cmaj. (2011, Sep 6); 183(12): E809-E816. Available from: Scopus®.
36. Cheng S, Lee T, Chen C. A Longitudinal Examination of a Pay-for-Performance Program for Diabetes Care: Evidence from a Natural Experiment. Medical Care. (2012, Feb); 50(2): 109-116. Available from: CINAHL Plus with Full Text.
37. MacBride-Stewart S, Elton R, Walley T. Do quality incentives change prescribing patterns in primary care? An observational study in Scotland. Family Practice (2008, Feb 1); 25(1): 27-32. Available from: Scopus®.
38. Doran T, Fullwood C, Reeves D, Roland M, Gravelle H. Exclusion of patients from pay-for-performance targets by English physicians. New England Journal of Medicine. (2008, July 17); 359(3): 274-284. Available from: Scopus®.

39. Doran T, Fullwood C, Kontopantelis E, Reeves D. Effect of financial incentives on inequalities in the delivery of primary clinical care in England: analysis of clinical activity indicators for the quality and outcomes framework. Lancet (2008, Aug 30); 372(9640): 728-736. Available from: CINAHL Plus with Full Text.
40. McGovern M, Williams D, Hannaford P, Taylor M, Lefevre K, Simpson C, et al. Introduction of a new incentive and target-based contract for family physicians in the UK: good for older patients with diabetes but less good for women?. Diabetic Medicine [serial on the Internet]. (2008, Sep); 25(9): 1083-1089. Available from: Academic Search Complete.
41. Gulliford M, Ashworth M, Robotham D, Mohiddin A. Achievement of metabolic targets for diabetes by English primary care practices under a new system of incentives. Diabetic Medicine (2007, May); 24(5): 505-511. Available from: Academic Search Complete.
42. Millett C, Gray J, Saxena S, Netuveli G, Majeed A. Impact of a pay-for-performance incentive on support for smoking cessation and on smoking prevalence among people with diabetes. CMAJ: Canadian Medical Association Journal Supplement (2007, June 5); 176(12): 1705-1710. Available from: Academic Search Complete.
43. Steel N, Maisey S, Clark A, Fleetcroft R, Howe A. Quality of clinical primary care and targeted incentive payments: An observational study. British Journal of General Practice (2007, June 1); 57(539): 449-454. Available from: Scopus®.
44. Millett C, Gray J, Saxena S, Netuveli G, Khunti K, Majeed A. Ethnic Disparities in Diabetes Management and Pay-for-Performance in the UK: The Wandsworth Prospective Diabetes Study. Plos Medicine [serial on the Internet] (2007, June); 4(6): e191-1093. Available from: Academic Search Complete.
45. Scott A, Schurer S, Jensen P, Sivey P. The effects of an incentive program on quality of care in diabetes management. Health Economics (2009, Sep); 18(9): 1091-1108. Available from: CINAHL Plus with Full Text.

46. Li Y, Tsai W, Khan M, Yang W, Lee T, Kung P, et al. The effects of pay-for-performance on tuberculosis treatment in Taiwan. Health Policy & Planning (2010, July); 25(4): 334-341. Available from: CINAHL Plus with Full Text.
47. Kuo R, Kuo-Piao C, Mei-Shu L. Effect of the Pay-for-Performance Program for Breast Cancer Care in Taiwan. Journal of Oncology Practice (2011, May 2); e8s-e15s. Available from: Academic Search Complete.
48. Tsai W, Kung P, Khan M, Campbell C, Yang W, Li Y, et al. Effects of pay-for-performance system on tuberculosis default cases control and treatment in Taiwan. Journal of Infection (2010, Sep 1); 61(3): 235-243. Available from: Scopus®.
49. Lai C, Hou Y. Original Article: The association of clinical guideline adherence and pay-for-performance among patients with diabetes. Journal of the Chinese Medical Association (2013, Feb 1); 76102-107. Available from: ScienceDirect.
50. Kontopantelis E, Reeves D, Valderas J, Campbell S, Doran T. Recorded quality of primary care for patients with diabetes in England before and after the introduction of a financial incentive scheme: a longitudinal observational study. BMJ Quality & Safety (2013, Jan); 22(1): 53-64. Available from: MEDLINE.
51. Colais P, Pinnarelli L, Fusco D, Davoli M, Braga M, Perucci C. The impact of a pay-for-performance system on timing to hip fracture surgery: experience from the Lazio Region (Italy). BMC Health Services Research (2013, Oct 7); 13393. Available from: MEDLINE.
52. Rubinstein A, Rubinstein F, Botargues M, Barani M, Kopitowski K. A multimodal strategy based on pay-per-performance to improve quality of care of family practitioners in Argentina. Journal of Ambulatory Care Management (2009, Apr 1); 32(2): 103-114. Available from: Scopus®.
53. Campbell S, Reeves D, Kontopantelis E, Sibbald BS, Roland M. Effects of Pay for Performance on the Quality of Primary Care in England. The New England Journals of Medicine (2009, July 23); 361(4): 368-378. Available from: CINAHL Plus with Full Text.

54. Millett C, Netuveli G, Saxena S, Majeed A. Impact of pay for performance on ethnic disparities in intermediate outcomes for diabetes: a longitudinal study. Diabetes Care (2009, Mar); 32(3): 404-409. Available from: CINAHL Plus with Full Text.
55. Lee J, Netuveli G, Majeed A, Millett C. The effects of pay for performance on disparities in stroke, hypertension, and coronary heart disease management: interrupted time series study. Plos One (2011); 6(12): e27236. Available from: MEDLINE.
56. Calvert M, Shankar A, McManus R, Lester H, Freemantle N. Effect of the quality and outcomes framework on diabetes care in the United Kingdom: retrospective cohort study. BMJ (Clinical Research Ed.) (2009, May 26); 338b1870. Available from: MEDLINE.
57. McGovern M, Boroujerdi M, Taylor M, Williams D, Hannaford P, Simpson C, et al. The effect of the UK incentive-based contract on the management of patients with coronary heart disease in primary care. Family Practice (2008, Feb); 25(1): 33-39. Available from: CINAHL Plus with Full Text.
58. Hamilton F, Bottle A, Vamos E, Anthea (1), Majeed A, Molokhia M, et al. Impact of a pay-for-performance incentive scheme on age, sex, and socioeconomic disparities in diabetes management in UK primary care. Journal of Ambulatory Care Management (2010, Oct 1); 33(4): 336-349. Available from: Scopus®.
59. Simpson CR, Hannaford PC, Lefevre K, Williams D: Effect of the UK incentive-based contract on the management of patients with stroke in primary care. Stroke 2006, 37:2354-2360.
60. Crawley D, Ng A, Mainous A, Majeed A, Millett C. Impact of pay for performance on quality of chronic disease management by social class group in England. Journal of The Royal Society Of Medicine (2009, Mar); 102(3): 103-107. Available from: CINAHL Plus with Full Text.
61. Sutton M, Nikolova S, Boaden R, Lester H, McDonald R, Roland M. Reduced mortality with hospital pay for performance in England. New England Journal of Medicine (2012, Nov 8); 367(19): 1821-1828. Available from: Scopus®.

62. Kirschner K, Braspenning J, Jacobs J, Grol R. Design choices made by target users for a pay-for-performance program in primary care: An action research approach. BMC Family Practice (2012, Jan 1); 13Available from: Scopus®.
63. Campbell S, McDonald R, Lester H. The experience of pay for performance in english family practice: A qualitative study. Annals of Family Medicine (2008, May 1); 6(3): 228-234. Available from: Scopus®.
64. Alshamsan R, Lee J, Majeed A, Netuveli G, Millett C. Effect of a UK pay-for-performance program on ethnic disparities in diabetes outcomes: Interrupted time series analysis. Annals of Family Medicine. (2012, May 1); 10(3): 228-234. Available from: Scopus®.
65. Lee T, Cheng S, Chen C, Lai M. A pay-for-performance program for diabetes care in Taiwan: a preliminary assessment. American Journal of Managed Care. (2010); 16(1): 65-69. Available from: CINAHL Plus with Full Text.
66. Kiran T, Victor J, Kopp A, Shah B, Glazier R. The relationship between financial incentives and quality of diabetes care in ontario, Canada. Diabetes Care (2012, May); 35(5): 1038-1046. Available from: CINAHL Plus with Full Text.
67. Tsung-Tai C, Kuo-Piao C, I-Chin L, Mei-Shu L. The Unintended Consequence of Diabetes Mellitus Pay-for-Performance (P4P) Program in Taiwan: Are Patients with More Comorbidities or More Severe Conditions Likely to Be Excluded from the P4P Program? Health Services Research (2011, Feb); 46(1p1): 47-60. Available from: Academic Search Complete.
68. Dalton A, Alshamsan R, Majeed A, Millett C. Exclusion of patients from quality measurement of diabetes care in the UK pay-for-performance programme. Diabetic Medicine (2011); (5): 525. Available from: Academic OneFile.
69. Doran T, Fullwood C, Gravelle H, Reeves D, Kontopantelis E, Roland M, et al. Pay-for-performance programs in family practices in the United Kingdom. New England Journal of Medicine (2006, July 27); 355(4): 375-384. Available from: CINAHL Plus with Full Text.

70. Inoue Y, Jeong S, Kondo K. Changes in hospital performance after implementation of pay-for-performance in Japan. Clinical Audit (2011); (default): 17. Available from: Directory of Open Access Journals.
71. Serumaga B, Ross-Degnan D, Avery A, Elliott R, Majumdar S, Soumerai S, et al. Effect of pay for performance on the management and outcomes of hypertension in the United Kingdom: interrupted time series study. BMJ (Clinical Research Ed.). (2011, Jan 25); 342d108. Available from: MEDLINE.
72. Systematic Reviews CRD's guidance for undertaking reviews in health care: Centre for Reviews and Dissemination, University of York, 2008. Published by CRD, University of York (2009, Jan), [Retrieved March 5, 2014]. Available from: http://www.york.ac.uk/inst/crd/pdf/Systematic_Reviews.pdf
73. Jackson N. Systematic Reviews of Health Promotion and Public Health Interventions: train-the-trainer handbook. Cochrane Health Promotion and Public Health Field, Victorian Health Promotion foundation (2005, May) [Retrieved March 3, 2014].
74. Higgins JPT, Green S (editors). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from www.cochrane-handbook.org. Downs, S.H. and Black, N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomized and non-randomized studies of health care interventions. Journal of Epidemiology Community Health (1998); 52: 377-384.
75. National Collaborating Centre for Methods and Tools (2008). Quality Checklist for Health Care Intervention Studies. Hamilton, ON: McMaster University. (Updated 01 June, 2010). Retrieved from http://www.nccmt.ca/registry/view/eng/9.html
76. GA Wells, B Shea, D O'Connell, J Peterson, V Welch, M Losos, P Tugwell. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. Retrieved from http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp

77. Review Manager (RevMan) [Computer program]. Version 5.0. Copenhagen: The Nordic Cochrane Centre, The Cochrane Collaboration, 2008.
78. MacPherson P, Houben R, Glynn J, Corbett E, Kranzer K. Pre-treatment loss to follow-up in tuberculosis patients in low-and lower-middle-income countries and high-burden countries: a systematic review and meta-analysis. Bulletin of the World Health Organization (2014, Feb); 92(2): 126-138. Available from: CINAHL Plus with Full Text.
79. Rust R, Lehmann D, Farley J. Estimating Publication Bias in Meta-Analysis. Journal of Marketing Research (JMR) [serial on the Internet]. (1990, May), [cited October 8, 2014]; 27(2): 220-226. Available from: Business Source Complete.
80. Chen P, Lee Y, Kuo R. Differences in patient reports on the quality of care in a diabetes pay-for-performance program between 1 year enrolled and newly enrolled patients. International Journal for Quality in Health Care. (2012, Apr 1); 24(2): 189-196. Available from: Scopus®.
81. Duckett S, Daniels S, Kamp M, Stockwell A, Walker G, Ward M. Pay for performance in Australia: Queensland's new Clinical Practice Improvement Payment. Journal of Health Services Research & Policy (2008, July); 13(3): 174-177. Available from: CINAHL Plus with Full Text.
82. Benavent J, Juan C, Clos J, Sequeira E, Gimferrer N, Vilaseca J. Using pay-for-performance to introduce changes in primary healthcare centres in Spain: first year results. Quality in Primary Care (2009, Apr); 17(2): 123-131. Available from: CINAHL Plus with Full Text.
83. Cheng A, Sutherland J. British Columbia's pay-for-performance experiment: Part of the solution to reduce emergency department crowding? Health Policy (2013, Nov 1); 113(1-2): 86-92. Available from: Scopus®.
84. Adaji A, Schattner P, Jones K, Beovich B, Piterman L. Care planning and adherence to diabetes process guidelines: Medicare data analysis. Australian Health Review (2013, Feb); 37(1): 83-87. Available from: CINAHL Plus with Full Text.

85. Walker E, Hernandez A, Kattan M. Meta-analysis: Its strengths and limitations. Cleveland clinic journal of medicine (2008, June); 25(6): 431-439.
86. DeCoster J. Meta-analysis notes. Department of Psychology University of Alabama, USA. (2004, Sep) Available from http://stat-help.com
87. Wilson K. Pay-for-performance in health care: What can we learn from international experience. Quality Management in Health Care [serial on the Internet]. (2013, Jan 1); 22(1): 2-15. Available from: Scopus®.
88. McDonald R, White J, Marmor T. Paying for Performance in Primary Medical Care: Learning about and Learning from "Success" and "Failure" in England and California. Journal of Health Politics, Policy & Law [serial on the Internet]. (2009, Oct); 34(5): 747-776. Available from: Business Source Complete.
89. Cheng M. Pay-for-performance in Taiwan, Academy Health Annual Research Meeting 2006. Seattle, Washington (2006, June). Available from: www.academyhealth.org
90. Greene J. An examination of pay-for-performance in general practice in Australia. Health Services Research (2013); (4): 1415. Available from: Business Insights: Essentials.
91. Rosenthal M, Frank R, Li Z. Early Experience with Pay-for-performance: From Cincept to practice. JAMA (2005); 294(14): 1788-1793.

APPENDICES

APPENDIX A
ETHICAL CONSIDERATION

Documentary Proof of Exemption
Ethical Review Committee for Human Research
Faculty of Public Health, Mahidol University

Protocol Title :	THE EFFECTS OF PAY-FOR-PERFORMANCE IN HEALTH CARE SYSTEMS WITH UNIVERSAL HEALTH COVERAGE: A SYSTEMATIC REVIEW
Protocol No. :	63/2557
Principal Investigator :	Miss Nithimar Sermsuti-Anuwat
Affiliation :	Master of Science (Public Health) Program in Health Administration Faculty of Public Health, Mahidol University

This protocol complies with a "Research with Exemption" category

Date of Issue : 28 March 2014

The aforementioned project have been reviewed and approved according to the Standard Operating Procedures of Ethical Review Committee for Human Research, Faculty of Public Health, Mahidol University.

S. Natthamongkolchai

.....

(Assoc. Prof. Dr. Sutham Nanthamongkolchai)

Chairman of Ethical Review Committee for Human Research

420/1 Rajvithi Road, Bangkok, Thailand 10400
Tel. (662) 3548543-9 ext. 1127, 7404 Fax. (662) 6409854

APPENDIX B
HEALTH SYSTEMS WITH UNIVERSAL HEALTH COVERAGE

Health systems with universal health coverage (UHC) context:

The healthcare systems that exist a legal mandate for provide the same health protection more than 90% of citizens with equal accessibility to the quality health services regardless of individual status and ability to pay according to the political economy of universal health coverage: background paper for the global symposium on the health systems research by the World Health Organization [17].

(1) Andorra, (2) Antigua, (3) Argentina, (4) Armenia, (5) Australia, (6) Austria, (7) Azerbaijan, (8) Bahrain, (9) Belarus, (10) Belgium, (11) Bosnia and Herzegovina, (12) Botswana, (13) Brunei Darussalam, (14) Bulgaria, (15) Canada, (16) Chile, (17) Costa Rica, (18) Croatia, (19) Cuba, (20) Cyprus, (21) Czech Republic (22) Denmark, (23) Estonia, (24) Finland, (25) France, (26) Germany, (27) Greece, (28) Hungary, (29) Iceland, (30) Ireland, (31) Israel, (32) Italy, (33) Japan, (34) Kuwait, (35) Luxembourg, (36) Moldova, (37) Mongolia, (38) Netherlands, (39) New Zealand (40) Norway, (41) Oman, (42) Panama, (43) Portugal, (44) Romania, (45) Singapore, (46) Slovakia, (47) Slovenia, (48) South Korea, (49) Spain, (50) Sweden, (51) Switzerland, (52) Taiwan, (53) Thailand, (54) Tunisia, (55) United Arab Emirates, (56) Ukraine, (57) The United Kingdom, (58) Venezuela.

APPENDIX C
DATA EXTRACTION FORM

Study ID:	**Report ID:**	Date form Completed:
First author:	Year of study:	Data extractor:
Citation:		

1. General Information

Publication type:	Journal Article ☐ / Abstract ☐ / Other_____
Country of study:	

2. Study Eligibility

Study Characteristics			**Page/ Para/ Figure #**
Type of study	☐ Randomized controlled Trial(RCT) ☐ Cluster Randomized Controlled Trial (cluster RCT)	☐ Controlled Before and After (CBA) study • Contemporaneous data collection • Comparable control site • At least 2x intervention and 2x control clusters	
	☐ Interrupted Time Series (ITS) • At least 1 time point before and 1 after the intervention with clearly defined intervention point	☐ Other design (specify)	
	☐ A process evaluation of an included study design	Does the study design meet the criteria for inclusion? yes☐ / No☐ ➡Exclude / Unclear ☐	

Study Characteristics			**Page/ Para/ Figure #**
Participants	Describe the participants included:		
	Are participants defined as a group having specific social or cultural characteristic?	Yes☐ / No☐ / Unclear ☐ Details:	
	How is the geographic boundary defined?	Details: Specific location (e.g. state / country):	
	Do the participants meet the criteria for inclusion?	Yes☐ / No☐➡Exclude Unclear ☐	
Types of intervention	Strategies included in the intervention**:**		
	Focus of the intervention:		
	Does the intervention meet the criteria for inclusion?	Yes☐ / No☐➡Exclude Unclear ☐	
Duration of intervention	Start date: Stop date:	Intervention duration:	
	Is the duration of intervention adequate for inclusion?	Yes☐ / No☐➡Exclude Unclear ☐	
Types of outcome measures	List outcomes: Do the outcome measures meet the criteria for inclusion?	Details: Yes☐ / No☐➡Exclude Unclear ☐	

Summary of Assessment for Inclusion

Include in review ☐ **Exclude from review ☐**

DO NOT PROCEED IF PAPER EXCLUDED FROM REVIEW

APPENDIX D
CHARACTERISTICS OF INCLUDED STUDIES BY STUDY

	Karunaratne K, 2013 [32] / The United Kingdom
Type of study	A prospective cohort study.
Aim of study	To assess the effectiveness of renal quality indicators in P4P on hypertension management in primary care.
Participants	This cohort study was conducted from a clinical decision support system of the CKD in the United Kingdom primary care database. The study examined a total population of 90,250 individuals registered with a valid serum creatinine testing in the 6-year study period. All 10,040 patients were stage 3-5 CKD in the two years pre-P4P implementation.
Method	Patients were studied over three time periods, pre-P4P (1 April 2004 to 31 March 2006), 2 years post-P4P (1 April 2006 to 31 March 2008) and finally the two subsequent years (1 April 2008 to 31 March 2010).
Interventions	To rewarded health care provider with Pay for performance (P4P) incentive which has been called the Quality and Outcomes Framework (QOF).
Outcomes	Process measures: P4P incentives were based on the achievement of the mean systolic and diastolic blood pressures (BP).
Results	• The percentage of patients with CKD 3-5 attaining the BP target of 145/80 increased from 41.5% in the pre-P4P period to 50.0% in the post-P4Pperiod. • The mean blood pressure in hypertensive patients decrease from 146/79 mmHg to 140/76 in the first 2 years post-P4P ($P < 0.01$) • This blood pressure reduction was sustained in the last 2 years of the study, 139/75 ($P < 0.01$, ANOVA).

Simpson C, 2011 [33] / Scotland	
Type of study	A population-based repeated cross-sectional
Aim of study	To investigate the impact of Pay-for-performance (P4P) on the management of patients with hypertension in the primary care.
Participants	There were 315 from 1030 physicians in Scotland contributed their computerized patient data to the Primary Care Clinical Informatics Unit (PCCIU). The 826,973 patients registered with these physicians have been found to be representative of the Scottish population. All patients aged ≥ 40 years with a computer record of hypertension were identified.
Method	The study had been assessed at six time points (from1 April 2001 until 1 April 2006). To ascertained the evidence of achieving target blood pressures that followed the British Hypertension Society and new General Medical Services (nGMS): a recorded blood pressure measurement of systolic/diastolic ≤ 140/90 mmHg or ≤ 150/90 mmHg.
Interventions	The nGMS contract (pay-for-performance)
Outcomes	The achievement of target blood pressure levels and provision of hypertension related prescribing for each year.
Results	• Increasing treatment for hypertension with absolute difference (AD) 9.2%; 95% CI = 9.0 to 9.5 • The majority of increases found in blood pressure measurement (AD) 46.8%; 95% CI = 46.5 to 47.1 • Recorded hypertension (AD) 5.9%; 95% CI = 5.7 to 6.0 • Blood pressure controlled ≤ 140/90 mmHg absolutely increased 18.9%; 95% CI = 18.5 to 19.4

Doran T, 2011 [34] / The United Kingdom	
Type of study	Longitudinal analysis of achievement rates for 42 activities.
Aim of study	To investigate whether the incentive scheme for UK general practitioners lead to neglect the non-incentive activities.
Participants	Patient level data were extracted from the General Practice Research Database (GPRD), which contains anonymized, patient based data on morbidity, prescribing, treatment, and referral collected from over 500 general practices, covering about 7% of the UK population (4.4 million patients).
Method	The study selected a sample of 148 practices that provided data to the GPRD continuously between January 2000 and December 2007, structured to include a range of list (patient panel) sizes. Selected practices were nationally representative in terms of patient sex and age distribution and area socioeconomic deprivation but had a relatively large average list size, reflecting a bias towards larger practices in the GPRD. The final sample consisted of 653,500 patients. Patients with relevant conditions were identified from their diagnostic Read codes.
Interventions	Pay-for-performance incentive
Outcomes	Achievement rates projected from trends in the pre-incentive period (2000-1 to 2002-3) and actual rates in the first three years of the scheme (2004-5 to 2006-7).
Results	Achievement rates improved for most indicators in the pre-incentive period. There were significant increases in the rate of improvement in the first year of the incentive scheme (2004-5) for 22 of the 23 incentivized indicators. Achievement for these indicators reached a plateau after 2004-5, but quality of care in 2006-7 remained higher than that predicted by pre-incentive trends for 14 incentivized indicators.

McGovern M, 2008 [40] / Scotland	
Type of study	A serial cross-sectional study
Aim of study	To determine whether the recording of diabetes-related health indicators has increased and differences diminished between age, gender and deprivation groups, following the introduction of the new General Medical Services contract (nGMS),
Participants	Total of 310 (out of 1030) practices in Scotland contributed their IT data to PCCIU. From the accumulated data, They identified all patients aged 17 years and over who had a computer record of diabetes (read code C10 and below—categorized as diabetes mellitus which includes Type 1 and Type 2) on 31 March 2004 was designated the 'pre-contract' dataset. The total population at risk pre-contract was 1,578,902 registered patients. All patients with diabetes on the 31 March 2005 (1 year after the introduction of the new contract) formed the 'post-contract' dataset. The total population at risk post-contract was 1,533,802 patients.
Method	They excluded patients from our analysis if such an exception code existed.
Interventions	Pay-for-performance incentive
Outcomes	Main outcome measures were the recording of diabetes health indicators and prescribing of medicines at pre- and post-contract time points.
Results	One year after the introduction of the nGMS contract, there was a 54.2% relative increase in the number of patients electronically recorded as having diabetes. In addition, measurement of the quality indicators glycated haemoglobin (HbA1c), blood pressure, serum creatinine and cholesterol significantly increased ($P< 0.05$). Women were less likely than men to have HbA1c [odds ratio (OR) 0.85, 95% confidence intervals (CI) 0.80–0.91].

	Gulliford M, 2007 [41] / The United Kingdom
Type of study	A cross-sectional design.
Aim of study	To analyze achievement of metabolic targets by English general practices following the introduction of a new system of incentives.
Participants	Clinical data were abstracted for 2099/2442 (86%) eligible diabetic subjects at 26 participating practices. There were 1441/2099 (69%) of subjects who had a diagnosis of diabetes.
Method	The study used three data sources. First, we used data from a local diabetes care assessment in 26 general practices in South London to determine trends over time from 2000 to 2005. Use of this data was approved by the Guy's Hospital Research Ethics Committee. Secondly, we employed data from the QOF to describe the national situation in 8484 practices in 2005. Thirdly, we included information concerning the socio-demographic characteristics of registered patient populations as well as practice organizational characteristics to understand factors that are associated with achieving metabolic targets.
Interventions	The Quality and Outcomes Framework (QOF).
Outcomes	Quality targets from the national GP contract were achieved in each calendar year for glycated haemoglobin (≤7.4%), cholesterol (≤5.0 mmol/l) or blood pressure (≤145/85 mmHg).
Results	Among 26 practices in South London, the median practice-specific proportion of patients achieving HbA1c ≤ 7.4% increased from 38% before P4P in 2003 to 57% after P4P implementation in 2005.

Millett C, 2007 [42] / The United Kingdom	
Type of study	A population-based longitudinal study.
Aim of study	To examined the impact of a pay-for-performance incentive on improving support for smoking cessation and to reduce the prevalence of smoking among people with chronic diseases such as diabetes.
Participants	Wandsworth Primary Care Trust, located in southwest London, England, has established comprehensive primary care based diabetes registers. The study area contains 36 primary care practices and has a total registered population of 243,519.
Method	The study identified all patients with type 1 or type 2 diabetes by searching computerized general practice records for “read codes” for diagnoses of diabetes (C10) or diabetes care (66A).[Read codes are the clinical classification system used in primary care in the United Kingdom.] Patients who received repeat prescriptions for diabetic medications or whose glycosylated hemoglobin level was greater than 7.5% were also included in our study. Patients under 18 years of age and women with gestational diabetes or who received treatment for polycystic ovarian syndrome rather than diabetes were excluded. The data were collected before (June–October 2003) and after (November 2005–January 2006).
Interventions	Pay-for-performance incentive
Outcomes	The study examined smoking status and cessation advice based on information recorded on practice computers during the 2003 and 2005 study periods.
Results	Significantly more patients with diabetes had their smoking status ever recorded in 2005 than in 2003 (98.8% v. 90.0%, $p < 0.001$). The proportion of patients with documented smoking cessation advice also increased significantly over this period, from 48.0% to 83.5% ($p < 0.001$).

Steel N, 2007 [43] / The United Kingdom	
Type of study	A retrospective observational study.
Aim of study	To examine the relationship between changes in recorded quality of care for four common chronic conditions from, 2003 to 2005, and the payment of incentives.
Participants	Medical records were examined for 1156 patients.
Method	Data were extracted from electronic and paper patient records to assess quality of care provided for asthma (seven indicators, two in the QOF, five not); hypertension (14 indicators, four in QOF, 10 not); osteoarthritis (nine indicators, none in QOF); and depression (six indicators, none in QOF).
Interventions	The Quality and Outcomes Framework (QOF) provided practices with substantial financial rewards (Pay-for-performance) for achievement of quality indicators in 10 chronic conditions.
Outcomes	To compared incentivized and non-incentivized indicators of quality of care.
Results	A significant increase occurred for the six indicators linked to incentive payments: from75% achieved in 2003 to 91% in 2005 (change=16%, 95%confidence interval [CI] =10to22%, P<0.01). A significant increase also occurred for 15 other indicators linked to 'incentivized conditions'; 53to64% (change=11%, 95% CI = 6to15%, P<0.01). The 'non-incentivized conditions' started at a lower achievement level, and did not increase significantly: 35to36% (change = 2%, 95%CI = –1 to 4%, P=0.19).

	Millett C, 2007 [44] / The United Kingdom
Type of study	A population-based longitudinal survey.
Aim of study	To examined disparities in management of people with diabetes and intermediate clinical outcomes within a multiethnic population in primary care before and after the introduction of the Pay-for-performance (P4P) in April 2004.
Participants	The study area contained 36 general practices with a registered population of 243,519 patients. The median list size of practices was 6,349 patients and there was an even distribution of large, medium, and small practices in the study area.
Method	The study used recorded data from electronic general practice records, in an ethnically diverse part of southwest London in 2003 and 2005. All patients with type 1 and type 2 diabetes mellitus were identified from computerized general practice records in participating practices by searching for diagnoses of diabetes (C10) or diabetes care (66A) Read codes. Patients with repeat prescribing for diabetic medications, or with an HbA1cgreater than 7.4%, were also included in our sample. Patients under 18 y of age, or women with gestational diabetes or receiving treatment for polycystic ovarian syndrome rather than diabetes, were excluded.
Interventions	Pay-for-performance
Outcomes	Outcome measures were prescribing levels and achievement of national treatment targets (HbA1c $\leq$7.0%; blood pressure [BP] <140/80 mm Hg; total cholesterol $\leq$5 mmol/l or 193 mg/dl).
Results	The proportion of patients reaching treatment targets for HbA1c, BP, and total cholesterol increased significantly after the implementation of the P4P. The extents of these increases were broadly uniform across ethnic groups, with the exception of the black Caribbean patient group, which had a significantly lower improvement in HbA1c.

Li Y, 2010 [46] / Taiwan	
Type of study	The retrospective study.
Aim of study	This study investigates the effectiveness of the P4P system in terms of cure rate and length of treatment.
Participants	The number of cases in pre-P4P years (2002 and 2003) was 25,754 compared with 33,536 in the post-P4P implementation years (2004 and 2005).
Method	The effectiveness of the program was evaluated by comparing the TB cure rate and length of treatment before and after the implementation of the P4P program, and between participating and non-participating hospitals. Logistic regression analysis was conducted to explore the factors affecting TB patients' cure rate within a 12-month treatment period.
Interventions	'Pay-for-Performance on Tuberculosis' program (P4P on TB)
Outcomes	The study compared the outcome of TB treatment before and after the implementation of P4P on TB and the outcomes between hospitals by participation status in the P4P program.
Results	The cure rate and the average length of treatment before the implementation of P4P were 46.9% and 256.24 days, respectively, compared with 63.0% and 249.74 days after implementation of P4P. The cure rate and length of treatment in P4P hospitals were 68.1% and 249.13 days, respectively, compared with 42.4% and days in non-P4P hospitals.

Kuo R, 2011 [47] / Taiwan	
Type of study	A population-based observational study with cross-sectional design.
Aim of study	To evaluate the impact of the nationwide pay-for-performance (P4P) program for breast cancer care (BC-P4P) in Taiwan on care quality, patient survival, and recurrence.
Participants	A total of 4,528 patients with stage I or II breast cancer diagnosed in 2002 or 2003 who received curative surgery were observed until the end of 2008.
Method	Retrospective analysis of population-based cancer registration and claims data was used in this study. This study applied multivariate linear regression to explore the association between BC-P4P enrollment and quality of care. Cox regression was applied to examine the effect of BC-P4P enrollment on 5-year recurrence and overall survival among patients with breast cancer.
Interventions	A nationwide breast cancer P4P (BC-P4P) initiative program
Outcomes	The comparison if quality of care provided by enrolled and non-enrolled hospitals and evaluate the effects of the BC-P4P program on patient survival and recurrence.
Results	After controlling for age, stage, type of surgery, and other factors, BC-P4P enrollees were found to have received better quality care than non-enrollees (P=0.001). Cox regression models also indicated that after controlling for patient characteristics, quality of care was related to better 5-year overall survival (odds ratio [OR], 0.212; P=0.001) and recurrence (OR, 0.289; P<0.001). Even when controlled by quality of care provided to patients and its interaction with status of BC-P4P enrollment, BC-P4P enrollment remained statistically significant regarding 5-year overall survival (OR, 0.167; P<0.001) and recurrence (OR, 0.370; P=0.002).

Tsai W, 2010 [48] / Taiwan	
Type of study	A retrospective study.
Aim of study	To investigate the effectiveness of the P4P system in terms of default rate.
Participants	National Health Insurance Research Dataset in Taiwan from 2002 to 2005 has been used for the study. Total study samples are 13,191 patients in the years 2002 and 2003, and 27,142 patients in the years 2004 and 2005.
Method	The duration of default is calculated by taking the difference between a medical visit and the next visit when the gap is 2 months or more. The date of medical visits and duration of medication days provide information to understand the patients' compliance behavior. Patient with as long as 4 months of treatment from start of treatment was included in the study since diagnosis errors are common in the beginning of TB treatment
Interventions	The "pay-for-performance on Tuberculosis" program (P4P on TB)
Outcomes	The study compared the differences of TB default rate before and after the implementation of P4P program, between participating and non-participating hospitals, and between P4P hospitals with and without case managers.
Results	The treatment default rate after "P4P on TB" was 11.37% compared with the 15.56% before "P4P on TB" implementation. The treatment default rate in P4P hospitals was 10.67% compared to 12.7% in non-P4P hospitals. In addition, the default rate was 10.4% in hospitals with case managers compared with 12.68% in hospitals without case managers.

Lai C, 2013 [49] / Taiwan	
Type of study	A cross-sectional study.
Aim of study	The study examined the effects of the DM-P4P program on guideline adherence among patients with diabetes.
Participants	Participation in the DM-P4P program is voluntary for physicians; in addition, the physician has the ability to select patients to be enrolled in the P4P program.
Method	The data for this study came from an NHI claim dataset that was obtained from the National Health Research Institute. The dataset included the registry for medical personnel, board certified specialists, contracted medical facilities, and ambulatory and inpatient care records. Physicians who had treated at least 50 patients with DM in 2008 were included. Physicians who were generalists, or who did not specialize in internal medicine, endocrinology, family medicine, or pediatrics were excluded.
Interventions	The diabetes mellitus pay-for-performance (DM-P4P) program.
Outcomes	To improve guideline adherence and the quality of care.
Results	A total of 520,804 patients were included in the analysis. Patients enrolled in the DM-P4P program were more likely to receive all of the guideline-recommended tests/examinations than patients treated by non-P4P physicians. Patients who were not enrolled in the program but who were treated by DM-P4P-participating physicians were more likely to receive three of the seven recommended tests/examinations than were those treated by non-P4P physicians.

	Kontopantelis E, 2013 [50] / The United Kingdom
Type of study	A longitudinal observational study.
Aim of study	To assessed the effect of the incentives on recorded quality of care for diabetes patients and its variation by patient and practice characteristics.
Participants	The study used the General Practice Research Database (GPRD) and selected a stratified sample of 148 English general practices in England. The GPs contributing data from 2000/1 to 2006/7, and obtained a random sample of 653,500 patients in which 23,920 diabetes patients identified.
Method	The study selected a structured sample of 148 practices. Selected practices were nationally representative in terms of patient sex and age distribution and area socioeconomic deprivation, but had a relatively high average list size, reflecting a bias towards larger practices in the GPRD. All patients who registered with a selected practice for at least 1 day during the study period were selected from practices with list sizes of 4500 or below. For larger practices, a random selection of 4500 patients was drawn from each practice. By including a larger percentage of patients from smaller practices we were able better control for the nested structure of the data in the analyses.
Interventions	The UK's Quality and Outcomes Framework (QOF)
Outcomes	The study quantified annually recorded quality of care at the patient-level, as measured by the 17 QOF diabetes indicators.
Results	Recorded quality of care improved for all subgroups in the pre-incentive period. In the first year of the incentives, composite quality improved over-and-above this pre-incentive trend by 14.2% (13.7–14.6%). By the third year the improvement above trend was smaller, but still statistically significant.

Colais P, 2013 [51] / Italy	
Type of study	A retrospective cohort study.
Aim of study	To compare the proportion of surgery for hip fracture performed within 48 hours of admission among Lazio hospitals according to different payment systems, before and after the implementation of the regional act.
Participants	This study is based on information from the Hospital Information System (HIS) and the Healthcare Emergency Information System (HEIS). The data of patients aged 65 years and over, residing in the Lazio region and admitted to an acute care hospital for hip fracture before (1 July 2008−30 June 2009) and after (1 July 2010−30 June 2011) the pay-for-performance act.
Method	The proportion of surgeries performed within 48 hours of hospital arrival was calculated. An adjusted multivariate regression analysis was applied to assess the effect of hospital payment type on the likelihood of surgery within 48 hours of hospital arrival.
Interventions	Pay-for-performance
Outcomes	The proportion of surgery for hip fracture performed within 48 hours.
Results	The share of patients with hip fracture that had surgery within 48 hours was 11.7% before the introduction of the pay-for-performance act and 22.2% after. The proportion of early hip fracture operations increased after the pay-for-performance act, regardless of hospital payment type. The largest increase of surgery within 48 h occurred in private hospitals (adjusted Relative Risk = 2.80, $p < 0.001$)

Rubinstein A, 2009 [52] / Argentina	
Type of study	A cross-sectional study
Aim of study	To assess the effectiveness of a multimodal intervention based on pay-for-performance, teamwork, continuous education, and audit and feedback to improve quality.
Participants	18 primary care centers distributed across the metropolitan area of Buenos Aires.
Method	The Division of Family and Community Medicine takes care of approximately 80,000 individuals. Each physician, who is in charge of a defined panel of approximately 1200 patients, belongs to 1 of 5 different primary care groups (UDAs), which are composed of 10 to 15 physicians responsible for the care of a population of 10,000 to 15,000 patients. The intended goal of analyzing quality by UDA was to limit the individual variability, to increase the pool of patients with less prevalent conditions, and to encourage group rather than individual commitment.
Interventions	Pay-for-performance
Outcomes	To analyzing quality by UDA was to limit the individual variability, to increase the pool of patients with less prevalent conditions, and to encourage group rather than individual commitment.
Results	After 2 years, a significant improvement was observed in most of the indicators measuring clinical effectiveness and some improvements were observed in other domains

Millett C, 2009 [54] / The United Kingdom	
Type of study	A cross sectional surveys.
Aim of study	To examine disparities in coronary heart disease management and intermediate clinical outcomes within a multiethnic population before and after the introduction of a major pay for performance initiative in April 2004.
Participants	2,891 individuals with coronary heart disease registered with participating practices in 2003 and 3,101 in 2005.
Method	Comparison of two cross-sectional surveys using electronic general practice records. Thirty-two family practices in south London, United Kingdom (UK). All practices in the study area were asked to participate and all patients with CHD were identified by searching for diagnoses of CHD or a repeat prescription for nitrates. Patients with management codes for a positive angiography test or for cardiac bypass surgery and coronary angioplasty were also identified. Medical records were then checked to confirm the diagnosis of CHD.
Interventions	Pay-for-performance
Outcomes	Percentage achievement by ethnic group of quality indicators in the management of coronary heart disease
Results	The proportion of patients reaching national treatment targets increased significantly for blood pressure (51.2% to 58.9%) and total cholesterol (65.7% to 73.8%) after the implementation of a major pay for performance initiative in April 2004. Improvements in blood pressure control were greater in the black group compared to whites, with disparities evident at baseline being attenuated (black 54.8% vs. white 58.3% reaching target in 2005).

	Calvert M, 2009 [56] / The United Kingdom
Type of study	A retrospective cohort study.
Aim of study	To examine the management of diabetes between 2001 and 2007 in the United Kingdom and to assess whether changes in the quality of care reflect existing temporal trends or are a direct result of the implementation of the quality and outcomes framework.
Participants	The study obtained data from the doctors' independent network (DIN)-LINK database. The study conducted patients with type 1 or type 2 diabetes recorded data from147 general practices.
Method	The study identified people with diabetes if they had a Read code for diabetes or one or more prescriptions for oral anti-diabetic drugs, insulin, or glucose testing kits. Read codes included those in the C10 hierarchy and other diabetes related Read codes including diabetes monitoring, referrals, and diabetes related eye and foot complications.
Interventions	Pay-for-performance programs
Outcomes	Annual prevalence of diabetes and attainment of process and clinical outcomes over the three years before and the three years after the introduction of the quality and outcomes framework.
Results	Significant improvements in process and intermediate outcome measures were observed during the six year period, with consecutive annual improvements observed before the introduction of incentives. The introduction of the quality and outcomes framework did not lead to improvement in the management of patients with type 1 diabetes, nor to a reduction in the number of patients with type 2 diabetes who had HbA1c levels greater than 10%.

McGovern, 2008 [57] / Scotland	
Type of study	A serial cross-sectional study
Aim of study	To determine whether the recording of CHD-related health indicators and prescribing of medicines have increased following the introduction of the nGMS contract (Pay-for-Performance) and whether differences in the treatment of patients of differing age, gender and deprivation have been affected.
Participants	The study carried out with 310 general practices in Scotland. They identified everyone who had a computer record of CHD (G3 to G3401, G342 to G366 and G38 to G3z) on March 31, 2004 (designated 'pre-contract' as the nGMS was introduced on April 1, 2004; total population at risk 1,806,266 registered patients) and March 31, 2005 (1 year after the introduction of the new contract; designated 'post-contract' 1,775,397 patients) from the Primary Care Clinical Informatics Unit (PCCIU)database,
Method	The key characteristics of each identified person at each time point were included: gender, age (<64, 65–75 or >75 years), number of CHD-related comorbidities (0, 1, 2 or 3+), deprivation status (deprivation quintile 1 (most affluent), to 5 (most deprived), based on Carstair's DEPCAT postcode categorization.
Interventions	Pay-for-performance programs
Outcomes	Main outcome measures were the recording of CHD-related health indicators and prescribing of medicines at pre- and post-contract time points.
Results	The recording of CHD-related quality indicators and prescribing increased dramatically (mean absolute increase of 17.1%) after the introduction of the nGMS contract. Post-contract, disparities between patient subgroups, continued for certain components of care. Women were less likely to be recorded than men in 9 of 11 components of care, with older patients.

Simpson CR, 2006 [59] / Scotland	
Type of study	A serial cross-sectional study
Aim of study	To ascertain whether a new contract based on financial incentives for general practitioners has been associated with improved recording of quality indicators for patients with stroke and whether there was evidence of any difference in change between sex, age, and deprivation groups
Participants	Anonymous retrospective data from all 310 of the 850 Scottish practices that use the General Practice Administrative Software System were conducted on March 31, 2004 (1 year before introduction of the new contract in April 2004, designated as the "pre-contract" period; total population at risk 1,806,266 patients) and March 31, 2005 (1 year after introduction of the new contract in April 2004; designated as the "post-contract" period; 1,775,397 patients).
Method	All registered patients with a recording of stroke before the 2 time points were included in the analyses and excluded anyone with a record of "exception codes from the analysis.
Interventions	Pay-for-performance programs
Outcomes	The recording of CHD-related health indicators and prescribing of medicines at pre- and post-contract time points.
Results	Documentation of quality indicators increased over time, with absolute increases for individual indicators ranging from 32.3% to 52.1%. There was a large increase in the documentation of quality indicators among the oldest patients (>75 years) and the most affluent patients. This tended to attenuate age groups differences and to exacerbate differences between deprivation groups. Women tended to have larger increases in documentation than men; however, sex differences persisted, with women less likely than men to have smoking habits recorded.

	Alshamsan R, 2012 [64] / The United Kingdom
Type of study	An interrupted time series analysis.
Aim of study	To examine the long-term effects of the Quality and Outcomes Framework (QOF), a major pay-for-performance program in the United Kingdom, on ethnic disparities in diabetes outcomes.
Participants	Electronic medical record data of diabetes patients registered with 29 family practices in Wandsworth, South West London, where the population is younger than that of England as a whole.
Method	Patients were identified by searching diagnostic and management Read codes in the patient electronic record using an established methodology. Read codes are the clinical classification system used in primary care in the United Kingdom. Historical clinical data were extracted on each patient for the years 2000 to 2007 from his or her electronic record.
Interventions	Pay-for-performance programs
Outcomes	The outcome measures were mean blood pressure, total cholesterol, and HbA1c based on each diabetes patient's last recorded measurement in each year.
Results	• Initial accelerated improvements in systolic blood pressure in white and black patients, but these improvements were sustained only in black patients (annual decrease: –1.68 mmHg; 95% CI, –2.41 to –0.95 mm Hg). • Initial improvements in diastolic blood pressure in white patients (–1.01 mm Hg; 95% CI, –1.79 to –0.24 mm Hg) and in cholesterol in white (–0.13 mmol/L; 95% CI, –0.21 to –0.05 mmol/L) and black (–0.10 mmol/L; 95% CI, –0.20 to –0.01 mmol/L) • There was no beneficial impact of QOF on HbA1cin any ethnic group. Existing disparities in risk factor control remained largely intact (for example; mean HbA1c: white 7.5%, black 7.8%, south Asian 7.8%; P<.05) at the end of the study period.

Kiran T, 2012 [66] / Canada	
Type of study	A retrospective cohort study
Aim of study	To assessed the impact of a diabetes incentive code introduced for primary care physicians in Ontario, Canada, in 2002 on quality of diabetes care at the population and patient level.
Participants	The study analyzed administrative data for 757,928 Ontarians with diabetes to examine the use of the code and receipt of three evidence-based monitoring tests from 2006 to 2008.
Method	They used available administrative claims data to examine the use of the diabetes incentive code and assess receipt of evidence-based monitoring tests among individuals with diabetes in Ontario. Data were accessed through a comprehensive research agreement with the Ontario Ministry of Health and Long Term Care. Prior to data analysis, all patient and provider identifiers were removed and replaced with unique encrypted numbers.
Interventions	Pay-for-performance programs
Outcomes	They assessed use of the Diabetes Management Assessment fee code using physician service claims to the Ontario Health Insurance Plan (OHIP).
Results	The proportion receiving the optimal number of all three monitoring tests (HbA1c, cholesterol, and eye tests) rose gradually from 16% in 2000 to 27% in 2008. Individuals who were younger, lived in rural areas, were not enrolled in a primary care model, or had a mental illness were less likely to receive all three recommended tests.

Inoue Y, 2011 [70] / Japan	
Type of study	The retrospective study.
Aim of study	The purpose of this study was to evaluate the effect of pay-for-performance (P4P) in hospitals in Japan, and to determine if any improvement occurred in the Functional Independence Measure (FIM™) score at the time of discharge from hospital, the return home rate.
Participants	The Rehabilitation Patients Databank of Japan was developed to facilitate this evaluation, with financial support from the Ministry of Health, Labor, and Welfare of Japan. By May 2011, 37 hospitals had contributed structured data for a total of 9,031 patients to the databank.
Method	The study divided the data into two groups, i.e. before P4P (April 2006 to March 2008) and after P4P (April 2008 to March 2010). Data were included for patients aged 65–100 years who had a hospital stay of 77–180 days. Finally, the total numbers of patients and hospitals included were 903 and 26, respectively (before P4P [14 hospitals, n = 530] and after P4P [12 hospitals, n = 373]).
Interventions	Pay-for-performance programs
Outcomes	Improvement occurred in the Functional Independence Measure (FIM™) score at the time of discharge from hospital, the return home rate.
Results	There was improvement in the process by which health care delivery was provided, but neither the FIM gain nor the return home rate was significantly higher after P4P was introduced.

Adaji A, 2013 [85] / Australia	
Type of study	A retrospective cohort study
Aim of study	To test the association, in patients with a diagnosis of diabetes I and II, between having or not having a care plan, (i.e. General Practice Management Plans (GPMPs), Team Care Arrangements (TCAs)). The checks comprised HbA1c, HDL cholesterol and urinary micro-albumin.
Participants	Data obtained from the Medicare Australia database
Method	Chi-square analysis of retrospective group data obtained from the Medicare database (from 'billing' patterns only).
Interventions	In Australia, GPs were originally paid for care planning by Medicare in 1999. This incentive scheme was replaced in 2005 by Medicare's Chronic Disease Management item numbers.
Outcomes	To test the association, in patients with a diagnosis of diabetes, between having or not having a GPMP or TCA.
Results	The creation of GPMPs was associated with general practitioners (GPs) requesting checks for HbA1c (59.7%), HDL cholesterol (36.9%) and micro albumin (50.8%) for diabetes patients in accordance with guideline recommendations. Although the TCA was associated with an increase in the frequency of HbA1c checks (61.3%), there was a reduction in the number of HDL cholesterol (23.7%) and micro albumin (36.8%) checks. The group with no care plans had the lowest association with HbA1c (47.8%), HDL cholesterol (19.7%) and micro albumin (29.3%) checks that met guideline requirements for diabetes.

APPENDIX E

SUMMARY OF THE SEPARATE PERFORMANCE ESTIMATES

No.	1st Author Year[Ref] Setting	Study design	Condition	Outcome measured		Results				Odds Ratio M-H, (95%CI) % Change between P4P and nonP4P
				Clinical information recorded	Quality indicators reached	P4P Number Achievement (%)	P4P Total	Non-P4P Number achievement (%)	Non-P4P Total	
1	Karunaratne K, 2013[32], The United Kingdom	Cohort	Chronic Kidney disease	-	• Blood pressure < 140/85 mm Hg (%)	5,020 (50)	10,040	4,167 (41.5)	10,040	1.41(1.33, 1.49) P < 0.00001 20.48% increase
2	Simpson CR, 2011[33], Scotland,	Cross sectional	Hyper-tension	-	•Blood pressure <140/90 mm Hg (%)	57,627 (28)	205,810	34,305 (22)	155,932	1.38 (1.36, 1.40) P < 0.00001 27.28% increase
3	Doran T, 2011[34]a, The United Kingdom	Cohort	Asthma	• Record of smoking status (%)	-	20,082 (84.6)	23,737	8,878 (37.4)	23,737	9.20 (8.80, 9.61) P < 0.00001 126.20% increase
4	Doran T, 2011[34]b, The United Kingdom	Cohort	Hyper-tension	• Record of smoking status (%)	-	48,857 (80.8)	60,466	22,251 (36.8)	60,466	7.23 (7.04, 7.42) P < 0.00001 119.57% increase
5	Doran T, 2011[34]c The UK	Cohort	Hyper-tension	• Record of blood pressure (%)	-	57,503 (95.1)	60,466	52,001 (86.0)	60,466	3.16 (3.02, 3.30) P < 0.00001 10.58% increase
6	Doran T, 2011[34]d The United Kingdom	Cohort	Coronary Heart Disease	• Record of smoking status (%)	-	16,607 (87.5)	18,979	9,110 (48.0)	18,979	7.58 (7.20, 7.99) P < 0.00001 82.29% increase

No.	1st Author Year[Ref] Setting	Study design	Condition	Outcome measured		Results				Odds Ratio M-H, (95%CI)
				Clinical information recorded	Quality indicators reached	P4P		Non-P4P		% Change between P4P and nonP4P
						Number Achievement (%)	Total	Number achievement (%)	Total	
7	Doran T, 2011[34]e The United Kingdom	Cohort	Coronary Heart Disease	• Record of blood pressure (%)	-	18,068 (95.2)	18,979	16,075 (84.7)	18,979	3.58 (3.32, 3.87) P < 0.00001 12.40% increase
8	Doran T, 2011[34]f, The United Kingdom	Cohort	Coronary Heart Disease	• Record of total cholesterol (%)	-	16,436 (86.6)	18,979	11,805 (62.2)	18,979	3.93 (3.73, 4.13) P < 0.00001 39.23% increase
9	Doran T, 2011[34]g, The United Kingdom	Cohort	Chronic Obstruc-tive Pulmona ry Disease	• Record of smoking status (%)	-	6,266 (93.2)	6,723	2,978 (44.3)	6,723	17.24 (15.50,19.18) P < 0.00001 110.38% increase
10	Doran T, 2011[34]h, The United Kingdom	Cohort	Diabetes	• Record of HbA1c or equivalent level (%)	-	17,049 (92.6)	18,411	15,134 (82.2)	18,411	2.71 (2.54, 2.90) P < 0.00001 12.65% increase
11	Doran T, 2011[34]i, The United Kingdom	Cohort	Diabetes	• Record of blood pressure (%)	-	17,601 (95.6)	18,411	16,275 (88.4)	18,411	2.85 (2.62, 3.10) P < 0.00001 8.14% increase
12	Doran T, 2011[34]j, The United Kingdom	Cohort	Diabetes	• Record of serum creatinine testing (%)	-	16,938 (92.0)	18,411	13,035 (70.8)	18,411	4.74 (4.46, 5.05) P < 0.00001 29.94% increase
13	Doran T, 2011[34]k, The United Kingdom	Cohort	Diabetes	• Record of total cholesterol (%)	-	16,846 (91.5)	18,411	13,900 (75.5)	18,411	3.49 (3.28, 3.72) P < 0.00001 21.19% increase

No.	1st Author Year[Ref] Setting	Study design	Condition	Outcome measured		Results				Odds Ratio M-H, (95%CI)
				Clinical information recorded	Quality indicators reached	P4P		Non-P4P		% Change between P4P and nonP4P
						Number Achievement (%)	Total	Number achievement (%)	Total	
14	Doran T, 2011[34]l, The United Kingdom	Cohort	Transient Ischaemic Attack (TIA) or Stroke	• Record of smoking status (%)	-	7,369 (86.0)	8,569	2,973 (34.7)	8,569	11.56 (10.72,12.47) P < 0.00001 147.8386% increase
15	Doran T, 2011[34]m, The United Kingdom	Cohort	TIA or Stroke	• Record of blood pressure (%)	-	8,092 (93.7)	8,569	6,710 (78.3)	8,569	4.70 (4.23, 5.22) P < 0.00001 19.67% increase
16	Doran T, 2011 [34]n, The United Kingdom	Cohort	TIA or Stroke	• Record of total cholesterol (%)	-	6,958 (81.2)	8,569	3,813 (44.5)	8,569	5.39 (5.03, 5.77) P < 0.00001 82.47% increase
17	Doran T, 2011[34]o, The United Kingdom	Cohort	Hypo-thyroidism	• Record of thyroid function tests (%)	-	13,183 (93.6)	14,084	10,831 (76.9)	14,084	4.39 (4.06, 4.75) P < 0.00001 21.72% increase
18	McGovern M, 2008 [40]a, Scotland	Cohort	Diabetes	•Record of HbA1c (%)	-	54,148 (96.4)	56,170	23,230 (62.3)	37,287	16.20 (15.43,17.02) P < 0.00001 54.74% increase
19	McGovern M, 2008 [40]b, Scotland	Cross sectional	Diabetes	• Record of blood pressure (%)	-	52,174 (96.5)	54,066	27,533 (73.8)	37,308	9.79 (9.30, 10.31) P < 0.00001 30.76% increase
20	McGovern M, 2008 [40]c, Scotland	Cross sectional	Diabetes	• Record of serum creatinine (%)	-	52,838 (94.2)	56,091	20,336 (54.5)	37,314	13.56 (13.02,14.13) P < 0.00001 72.84% increase

No.	1st Author Year[Ref] Setting	Study design	Condition	Outcome measured		Results				Odds Ratio M-H, (95%CI) % Change between P4P and nonP4P
				Clinical information recorded	Quality indicators reached	P4P Number Achievement (%)	P4P Total	Non-P4P Number achievement (%)	Non-P4P Total	
21	McGovern M, 2008 [40]d, Scotland	Cross sectional	Diabetes	• Record of total cholesterol (%)	-	54,584 (94.2)	57,945	21,845 (58.6)	37,278	11.47 (11.02,11.95) P < 0.00001 60.75% increase
22	McGovern M, 2008 [40]e, Scotland	Cross sectional	Diabetes	-	• HbA1c ≤ 7.4 % (%)	28,535 (52.7)	54,146	10,449 (45)	23,220	1.36 (1.32, 1.40) P < 0.00001 17.11% increase
23	McGovern M, 2008 [40]f, Scotland	Cross sectional	Diabetes	-	• HbA1c ≤ 10 % (%)	48,877 (90.3)	54,127	18,577 (80)	23,221	2.33 (2.23, 2.43) P < 0.00001 12.88% increase
24	McGovern M, 2008 [40]g, Scotland	Cross sectional	Diabetes	-	• Blood pressure ≤145/85 mm Hg (%)	37,563 (69.5)	54,048	17,403 (63.2)	27,537	1.33 (1.29,1.37) P < 0.00001 9.97% increase
25	McGovern M, 2008 [40]h, Scotland	Cross sectional	Diabetes	-	•Total cholesterol ≤ 5 mmol/L (%)	36,128 (66.2)	54,575	14,745 (67.5)	21,845	0.94 (0.91,0.98) P=0.0006 **1.93% decrease**
26	Gulliford M, 2007 [41]a, The United Kingdom	Cross sectional	Diabetes	• Record of Body mass index (%)	-	1,355 (94)	1,441	1,182 (82)	1,441	3.45 (2.67, 4.46) P < 0.00001 14.63% increase
27	Gulliford M, 2007, [41]b, The United Kingdom	Cross sectional	Diabetes	• Record of HbA1c (%)	-	1,369 (95)	1,441	1,124 (78)	1,441	5.36 (4.10, 7.01) P < 0.00001 21.79% increase
28	Gulliford M, 2007 [41]c, The United Kingdom	Cross sectional	Diabetes	• Record of blood pressure (%)	-	1,326 (92)	1,441	1,282 (89)	1,441	1.43 (1.11, 1.84) P = 0.0054 3.37% increase

No.	1st Author Year[Ref] Setting	Study design	Condition	Outcome measured		Results				Odds Ratio M-H, (95%CI)
				Clinical information recorded	Quality indicators reached	P4P		Non-P4P		% Change between P4P and nonP4P
						Number Achievement (%)	Total	Number achievement (%)	Total	
29	Gulliford M, 2007 [41]d, The United Kingdom	Cross sectional	Diabetes	• Record of total cholesterol (%)	-	1,340 (93)	1,441	1,110 (77)	1,441	3.96 (3.12, 5.01) P < 0.00001 20.78% increase
30	Gulliford M, 2007 [41]e, The United Kingdom	Cross sectional	Diabetes	-	• HbA1c ≤ 7.4% (%)	821 (57)	1,441	548 (38)	1,441	2.16 (1.86, 2.50) P < 0.00001 50% increase
31	Gulliford M, 2007 [41]f, The United Kingdom	Cross sectional	Diabetes	-	• HbA1c ≤ 10% (%)	1,282 (89)	1,441	1,038 (72)	1,441	3.13 (2.56, 3.83) P < 0.00001 23.61% increase
32	Gulliford M, 2007[41]g, The United Kingdom	Cross sectional	Diabetes	-	• Blood pressure ≤ 145/85 mm Hg (%)	1,009 (70)	1,441	721 (50)	1,441	2.33 (2.00, 2.72) P < 0.00001 40% increase
33	Gulliford M, 2007 [41]h, The United Kingdom	Cross sectional	Diabetes	-	• Total cholesterol ≤ 5 mmol/L (%)	1038 (72)	1441	677 (47)	1,441	2.91 (2.49, 3.39) P < 0.00001 53.19% increase
34	Millett C, 2007[42]a, The United Kingdom	Cohort	Diabetes	• Record of smoking status (%)	-	3,581 (83.6)	4,284	2,896 (67.6)	4,284	2.44 (2.20, 2.71) P < 0.00001 23.67% increase
35	Millett C, 2007[42]b, The United Kingdom	Cohort	Diabetes	• Record of smoking cessation advice (%)	-	3,577 (83.5)	4,284	2,056 (48)	4,284	5.48 (4.96, 6.06) P < 0.00001 73.96% increase

No.	1st Author Year[Ref] Setting	Study design	Condition	Outcome measured		Results				Odds Ratio M-H, (95%CI) % Change between P4P and nonP4P
				Clinical information recorded	Quality indicators reached	P4P Number Achievement (%)	P4P Total	Non-P4P Number achievement (%)	Non-P4P Total	
36	Steel N, 2007[43]a, The United Kingdom	Cohort	Hypertension	• Record of smoking cessation advice (%)	-	1,110 (96)	1,156	1,080 (93.4)	1,156	1.70 (1.17, 2.47) P = 0.0057 2.78% increase
37	Steel N, 2007[43]b, The United Kingdom	Cohort	Hypertension	• Record of Blood pressure (%)	-	1,098 (95)	1,156	1,058 (91.5)	1,156	1.75 (1.25, 2.45) P = 0.0010 3.83% increase
38	Steel N, 2007[43]c, The United Kingdom	Cohort	Hypertension	-	• Blood pressure≤150/90 mm Hg (%)	925 (80)	1,156	857 (74.1)	1,156	1.40 (1.15, 1.70) P = 0.0008 7.96% increase
39	Millett C, 2007[44]a, The United Kingdom	Cohort	Diabetes	-	• HbA1c ≤ 7.0% (%)	1,602 (37.4)	4,284	1,504 (35.1)	4,284	1.10 (1.01, 1.21) P = 0.0277 6.55% increase
40	Millett C, 2007[44]b, The United Kingdom	Cohort	Diabetes	-	• Cholesterol ≤5 mmol/l (%)	3,016 (70.4)	4,284	2,463 (57.5)	4,284	1.76 (1.61, 1.92) P < 0.00001 22.43% increase
41	Millett C, 2007[44]c, The United Kingdom	Cohort	Diabetes	-	• Blood pressure < 140/80 mm Hg (%)	1,812 (42.3)	4,284	1,345 (31.4)	4,284	1.60 (1.47, 1.75) P < 0.00001 34.71% increase
42	Li Y, 2010[46], Taiwan	Cohort	Tuberculosis	-	• Cases cured within 9 months (%)	21,137 (63)	33,551	12,060 (46.9)	25,714	1.93 (1.87, 1.99) P < 0.00001 34.33% increase
43	Kuo R, 2011[47], Taiwan,	Cohort	Breast Cancer	-	• No cancer recurrence within 5 years (%)	1,204 (86.4)	1393	2,593 (82.7)	3,135	1.33 (1.11, 1.59) P = 0.0017 4.47% increase

No.	1st Author Year[Ref] Setting	Study design	Condition	Outcome measured		Results				Odds Ratio M-H, (95%CI)
				Clinical information recorded	Quality indicators reached	P4P		Non-P4P		% Change between P4P and nonP4P
						Number Achievement (%)	Total	Number achievement (%)	Total	
44	Tsai W, 2010[48], Taiwan ,	Cohort	Tuberculosis	-	•The tuberculosis no default rate (%)	24,056 (88.63)	27,142	11,138 (84.44)	13,191	1.44 (1.35, 1.53) P < 0.00001 4.9621% increase
45	Lai C, 2013[49]a, Taiwan	Cross sectional	Diabetes	•Record of Lipid test (%)	-	117,174 (80)	146,467	49,614 (28)	177,192	10.29 (10.12,10.46) P < 0.00001 185.71% increase
46	Lai C, 2013[49]b, Taiwan	Cross sectional	Diabetes	• Record of serum Creatinine (sCr) test (%)	-	136,214 (93)	146,467	131,122 (74)	177,192	4.67 (4.56, 4.78) P < 0.00001 25.68% increase
47	Lai C, 2013[49]c, Taiwan	Cross sectional	Diabetes	• Record of urinalysis test (%)	-	112,780 (77)	146,467	42,526 (24)	177,192	10.60 (10.43,10.78) P < 0.00001 220.83% increase
48	Lai C, 2013[49]d, Taiwan	Cross sectional	Diabetes	• Eye test (%)	-	98,133 (67)	146,467	33,666 (19)	177,192	8.66 (8.52, 8.80) P < 0.00001 252.63% increase
49	Kontopantelis E, 2013[50]a, The United Kingdom	Cohort	Diabetes	• Record of Body Mass Index (%)	-	14,307 (90.9)	15,739	9,900 (74.4)	13,306	3.44 (3.22, 3.67) P < 0.00001 22.1774% increase

No.	1st Author Year[Ref] Setting	Study design	Condition	Outcome measured		Results				Odds Ratio M-H, (95%CI)
				Clinical information recorded	Quality indicators reached	P4P		Non-P4P		% Change between P4P and nonP4P
						Number Achievement (%)	Total	Number achievement (%)	Total	
50	Kontopantelis E, 2013[50]b, The United Kingdom	Cohort	Diabetes	• Record of smoking status (%)	-	14,370 (91.3)	15,739	7,052 (53.0)	13,306	9.31 (8.72, 9.93) P < 0.00001 72.26% increase
51	Kontopantelis E, 2013[50]c, The United Kingdom	Cohort	Diabetes	• Record of smoking cessation advice (%)	-	6,012 (38.2)	15,739	1,783 (13.4)	13,306	3.99 (3.76, 4.24) P < 0.00001 185.07% increase
52	Kontopantelis E, 2013[50]d, The United Kingdom	Cohort	Diabetes	• Record of HbA1c or equivalent level (%)	-	14,637 (93.0)	15,739	11,017 (82.8)	13,306	2.76 (2.56, 2.98) P < 0.00001 12.32% increase
53	Kontopantelis E, 2013[50]e, The United Kingdom	Cohort	Diabetes	• Record of retinal screening (%)	-	11,348 (72.1)	15,739	7,411 (55.7)	13,306	2.06 (1.96, 2.16) P < 0.00001 29.44% increase
54	Kontopantelis E, 2013[50]f, The United Kingdom	Cohort	Diabetes	• Record of presence/absence of peripheral pulses (%)	-	12,213 (77.6)	15,739	6,493 (48.8)	13,306	3.63 (3.46, 3.82) P < 0.00001 59.02% increase
55	Kontopantelis E, 2013[50]g, The United Kingdom	Cohort	Diabetes	• Record of neuropathy testing (%)	-	10,010 (63.6)	15,739	479 (3.6)	13,306	46.79 (42.47,51.55) P < 0.00001 1666.67% increase
56	Kontopantelis E, 2013[50]h, The United Kingdom	Cohort	Diabetes	• Record of blood pressure (%)	-	15,157 (96.3)	15,739	11,882 (89.3)	13,306	3.12 (2.83, 3.45) P < 0.00001 7.8387% increase

No.	1st Author Year[Ref] Setting	Study design	Condition	Outcome measured		Results				Odds Ratio M-H, (95%CI) % Change between P4P and nonP4P
				Clinical information recorded	Quality indicators reached	P4P Number Achievement (%)	P4P Total	Non-P4P Number achievement (%)	Non-P4P Total	
57	Kontopantelis E, 2013[50]i, The United Kingdom	Cohort	Diabetes	• Record of micro-albuminuria testing (%)	-	8,169 (51.9)	15,739	1,850 (13.9)	13,306	6.68 (6.30, 7.08) P < 0.00001 273.3813% increase
58	Kontopantelis E, 2013[50]j, The United Kingdom	Cohort	Diabetes	• Record of serum creatinine testing (%)	-	14,653 (93.1)	15,739	9,554 (71.8)	13,306	5.30 (4.93, 5.70) P < 0.00001 29.67% increase
59	Kontopantelis E, 2013[50]k, The United Kingdom	Cohort	Diabetes	• Treated with ACE inhibitors (%)	-	12,686 (80.6)	15,739	10,498 (78.9)	13,306	1.11(1.05, 1.18) P = 0.0003 2.1546% increase
60	Kontopantelis E, 2013[50]l, The United Kingdom	Cohort	Diabetes	• Record of total cholesterol (%)	-	14,559 (92.5)	15,739	10,086 (75.8)	13,306	3.94 (3.67, 4.23) P < 0.00001 22.03% increase
61	Kontopantelis E, 2013[50]m The United Kingdom	Cohort	Diabetes	• Record of Influenza immunisation (%)	-	11,930 (75.8)	15,739	8,915 (67)	13,306	1.54 (1.47, 1.62) P < 0.00001 13.13% increase
62	Kontopantelis E, 2013[50]n, The United Kingdom	Cohort	Diabetes	-	• HbA1C ≤ 7.4% (%)	8,751 (55.6)	15,739	6,680 (50.2)	13,306	1.24 (1.19, 1.30) P < 0.00001 10.76% increase
63	Kontopantelis E, 2013[50]o, The United Kingdom	Cohort	Diabetes	-	• HbA1C ≤ 10 % (%)	14,574 (92.6)	15,739	12,082 (90.8)	13,306	1.27 (1.17, 1.38) P < 0.00001 1.9824% increase

No.	1st Author Year[Ref] Setting	Study design	Condition	Outcome measured		Results				Odds Ratio M-H, (95%CI) % Change between P4P and nonP4P
				Clinical information recorded	Quality indicators reached	P4P Number Achievement (%)	P4P Total	Non-P4P Number achievement (%)	Non-P4P Total	
64	Kontopantelis E, 2013[50]p, The United Kingdom	Cohort	Diabetes	-	• Blood pressure ≤145/85 mm Hg (%)	10,813 (68.7)	15,739	7,451 (56.0)	13,306	1.72 (1.64, 1.81) P < 0.00001 22.68% increase
65	Kontopantelis E, 2013[50]q, The United Kingdom	Cohort	Diabetes	-	• Total cholesterol ≤5 mmol/l (%)	11,663 (74.1)	15,739	7,531 (56.6)	13,306	2.19 (2.09, 2.31) P < 0.00001 30.92% increase
66	Colais P, 2013[51], Italy	Cohort	Hip fracture	• The patients with hip fracture that had surgery within 48 hours (%)	-	1,419 (22.20)	6,390	707 (11.70)	6,043	2.15 (1.95, 2.38) P < 0.00001 89.74% increase
67	Rubinstein A, 2009 [52]a, Argentina	Cross sectional	Diabetes	• Record of HbA1c (%)	-	2,035 (91)	2,236	1,431 (64)	2,236	5.70 (4.81, 6.74) P < 0.00001 42.19% increase
68	Rubinstein A, 2009 [52]b, Argentina	Cross sectional	Diabetes	-	•HbA1c < 8% (%)	1,766 (79)	2,236	1,431 (64)	2,236	2.11 (1.85, 2.42) P < 0.00001 23.44% increase
69	Rubinstein A, 2009 [52]c, Argentina,	Cross sectional	Diabetes	• Record of LDL-C (%)		2,124 (95)	2,236	1,722 (77)	2,236	5.66 (4.57, 7.01) P < 0.00001 23.38% increase
70	Rubinstein A, 2009 [52]d, Argentina	Cross sectional	Diabetes		• LDL-C < 130 mg/% (%)	1,655 (74)	2,236	939 (42)	2,236	3.93 (3.47, 4.46) P < 0.00001 76.19% increase
71	Rubinstein A, 2009 [52]e, Argentina	Cross sectional	Cardio-Vascular Disease	• Record of LDL-C (%)		3,077 (93)	3,309	2,713 (82)	3,309	3.93 (3.47, 4.46) P < 0.00001 13.41% increase

No.	1st Author Year[Ref] Setting	Study design	Condition	Outcome measured		Results				Odds Ratio M-H, (95%CI)
				Clinical information recorded	Quality indicators reached	P4P		Non-P4P		% Change between P4P and nonP4P
						Number Achievement (%)	Total	Number achievement (%)	Total	
72	Rubinstein A, 2009 [52]f, Argentina,	Cross sectional	Cardio-Vascular Disease		• LDL-C < 130 mg/% (%)	1,654.5 (50)	3,309	827 (25)	3,309	2.91 (2.48,3.42) P < 0.00001 100% increase
73	Rubinstein A, 2009 [52]g, Argentina,	Cross sectional	Hypertension	• Record of blood pressure (%)	-	11,246 (90)	12,495	8,372 (67)	12,495	4.43 (4.14, 4.75) P < 0.00001 34.3284% increase
74	Rubinstein A, 2009 [52]h, Argentina,	Cross sectional	Hypertension	-	• Blood pressure < 140/90 mm Hg (%)	7,997 (64)	12,495	4,373 (35)	12,495	3.30 (3.14, 3.48) P < 0.00001 82.8571% increase
75	Millett C, 2009[54]a, The United Kingdom	Cross sectional	Coronary Heart Disease	-	• Blood pressure < 150/90 mm Hg (%)	1,826 (58.9)	3,101	1,480 (51.2)	2,891	1.37 (1.23, 1.51) P < 0.00001 15.04% increase
76	Millett C, 2009[54]b, The United Kingdom	Cross sectional	Coronary Heart Disease	-	• Total cholesterol ≤5 mmol/l (%)	2,289 (73.8)	3,101	1,199 (65.7)	2,891	3.98 (3.57, 4.44) P < 0.00001 12.33% increase
77	Calvert M, 2009[56]a, The United Kingdom	Cohort	Diabetes	• Record of body mass index (%)	-	33,131 (84.6)	39175	16,559 (51.7)	32,037	5.12 (4.95, 5.31) P < 0.00001 63.64% increase
78	Calvert M, 2009[56]b, The United Kingdom	Cohort	Diabetes	• Record of smoking status (%)	-	37,107 (94.7)	39,175	24,535 (76.6)	32,037	5.49 (5.21, 5.78) P < 0.00001 23.63% increase

No.	1st Author Year[Ref] Setting	Study design	Condition	Outcome measured		Results				Odds Ratio M-H, (95%CI)
				Clinical information recorded	Quality indicators reached	P4P		Non-P4P		% Change between P4P and nonP4P
						Number Achievement (%)	Total	Number achievement (%)	Total	
79	Calvert M, 2009[56]c, The United Kingdom	Cohort	Diabetes	• Record of smoking cessation advice (%)	-	5,243 (87.7)	5,978	1,793 (34.5)	5,196	13.54 (12.30,14.90) P < 0.00001 154.2029% increase
80	Calvert M, 2009[56]d, The United Kingdom	Cohort	Diabetes	• Record of HbA1c or equivalent level (%)	-	34,627 (88.4)	39,175	24,554 (76.6)	32,037	2.32 (2.23, 2.42) P < 0.00001 15.40% increase
81	Calvert M, 2009[56]e, The United Kingdom	Cohort	Diabetes	• Record of retinal screening (%)	-	28,985 (74.3)	39,004	14,976 (46.8)	32,037	3.30 (3.19, 3.40) P < 0.00001 58.76% increase
82	Calvert M, 2009[56]f, The United Kingdom	Cohort	Diabetes	• Record of presence or absence of peripheral pulses (%)	-	29,613 (75.8)	39,059	7,195 (22.5)	32,037	10.82 (10.45,11.21) P < 0.00001 236.89% increase
83	Calvert M, 2009[56]g, The United Kingdom	Cohort	Diabetes	• Record of neuropathy testing (%)	-	29,419 (75.3)	39,059	5,282 (16.5)	32,037	15.46 (14.89,16.05) P < 0.00001 356.36% increase
84	Calvert M, 2009[56]h, The United Kingdom	Cohort	Diabetes	• Record of blood pressure (%)	-	37,037 (94.5)	39174	28,334 (88.4)	32,037	2.27 (2.14, 2.39) P < 0.00001 6.9005% increase

No.	1st Author Year[Ref] Setting	Study design	Condition	Outcome measured		Results				Odds Ratio M-H, (95%CI)
				Clinical information recorded	Quality indicators reached	P4P		Non-P4P		% Change between P4P and nonP4P
						Number Achievement (%)	Total	Number achievement (%)	Total	
85	Calvert M, 2009[56]i, The United Kingdom	Cohort	Diabetes	• Record of micro-albuminuria testing (%)	-	25,686 (67.5)	38,053	7,447 (23.5)	31,658	6.75 (6.53, 6.98) P < 0.00001 187.234% increase
86	Calvert M, 2009[56]j, The United Kingdom	Cohort	Diabetes	• Record of serum creatinine testing (%)	-	35,414 (90.4)	39,175	22,225 (69.4)	32,037	4.16 (3.99, 4.33) P < 0.00001 30.26% increase
87	Calvert M, 2009[56]k, The United Kingdom	Cohort	Diabetes	• Record of micro albuminuria and treated with ACE inhibitors (%)	-	1,942 (82.3)	2,360	356 (70.2)	507	1.97 (1.59, 2.45) P < 0.00001 17.24% increase
88	Calvert M, 2009[56]l, The United Kingdom	Cohort	Diabetes	• Record of total cholesterol (%)	-	34,832 (88.9)	39,175	23,438 (73.2)	32,037	2.94 (2.83, 3.06) P < 0.00001 21.45% increase
89	Calvert M, 2009[56]m, The United Kingdom	Cohort	Diabetes	• Record of Influenza vaccination (%)	-	28,260 (82.1)	34,435	19,458 (62.1)	31,341	2.79 (2.70, 2.90) P < 0.00001 32.21% increase
90	Calvert M, 2009[56]n, The United Kingdom	Cohort	Diabetes	-	• Total cholesterol ≤5 mmol/l (%)	23,055 (66.7)	34,562	12,213 (41.7)	29,275	2.80 (2.71, 2.89) P < 0.00001 59.95% increase

No.	1st Author Year[Ref] Setting	Study design	Condition	Outcome measured		Results				Odds Ratio M-H, (95%CI)
				Clinical information recorded	Quality indicators reached	P4P		Non-P4P		% Change between P4P and nonP4P
						Number Achievement (%)	Total	Number achievement (%)	Total	
91	Calvert M, 2009[56]o, The United Kingdom	Cohort	Diabetes	-	• Blood pressure ≤145/85 mm Hg (%)	23,020 (64.9)	35,457	14,481 (49.2)	29,413	1.91 (1.85, 1.97) P < 0.00001 31.91% increase
92	Calvert M, 2009[56]p, The United Kingdom	Cohort	Diabetes	-	• HbA1C ≤ 7.5% (%)	19,756 (56)	35,271	12,997 (44.2)	29,413	1.61 (1.56, 1.66) P < 0.00001 26.70% increase
93	Calvert M, 2009[56]q, The United Kingdom	Cohort	Diabetes	-	• HbA1C ≤ 10% (%)	29,412 (83.4)	35,271	20,933 (71.2)	29,413	2.03 (1.96, 2.11) P < 0.00001 17.13% increase
94	McGovern M, 2008[57]a, Scotland	Cross sectional	Coronary Heart Disease	• Record of Angina patient exercise test (%)	-	19,561 (66.2)	29,548	7,962 (63.9)	12,460	1.11 (1.06, 1.16) P < 0.00001 3.60% increase
95	McGovern M, 2008 [57]b, Scotland	Cross sectional	Coronary Heart Disease	• Record of smoking status (%)	-	71,747 (95.7)	74,971	35,095 (69.5)	50,496	9.77 (9.38, 10.16) P < 0.00001 37.70% increase
96	McGovern M, 2008 [57]c, Scotland	Cross sectional	Coronary Heart Disease	• Record of smoking cessation advice (%)	-	31,881 (96.2)	33,140	9,904 (81.0)	12,227	5.94 (5.53, 6.38) P < 0.00001 18.76% increase
97	McGovern M, 2008 [57]d, Scotland	Cross sectional	Coronary Heart Disease	• Record of blood pressure (%)	-	72,887 (97.2)	74,987	38,203 (75.7)	50,466	11.14 (10.62,11.69) P < 0.00001 28.40% increase

No.	1st Author Year[Ref] Setting	Study design	Condition	Outcome measured		Results				Odds Ratio M-H, (95%CI)
				Clinical information recorded	Quality indicators reached	P4P		Non-P4P		% Change between P4P and nonP4P
						Number Achievement (%)	Total	Number achievement (%)	Total	
98	McGovern M, 2008 [57]e, Scotland	Cross sectional	Coronary Heart Disease	• Record of total cholesterol (%)	-	63,393 (85.8)	73,885	22,253 (44.1)	50,460	7.66 (7.45, 7.87) P < 0.00001 94.56% increase
99	McGovern M, 2008 [57]f, Scotland	Cross sectional	Coronary Heart Disease	•Record of anti-platelet or anticoagulant therapy (%)	-	66,625 (90.3)	73,782	33,225 (65.8)	50,494	4.84 (4.69, 4.99) P < 0.00001 37.23% increase
100	McGovern M, 2008 [57]g, Scotland	Cross sectional	Coronary Heart Disease	•Record of ß-Blocker therapy (%)	-	41,371 (70.0)	59,101	20,688 (42.6)	48,563	3.14 (3.07, 3.22) P < 0.00001 64.32% increase
101	McGovern M, 2008 [57]h, Scotland	Cross sectional	Coronary Heart Disease	•Record of ACE inhibitor (%)	-	5,764 (77.9)	7,399	1,838 (66.4)	2,768	1.78 (1.62, 1.96) P < 0.00001 17.3193% increase
102	McGovern M, 2008 [57]i, Scotland	Cross sectional	Coronary Heart Disease	•Record of Influenza vaccination (%)	-	57,170 (85.5)	66,866	28,289 (57.4)	49,284	4.38 (4.26, 4.50) P < 0.00001 48.95% increase
103	McGovern M, 2008 [57]j, Scotland	Cross sectional	Coronary Heart Disease	-	• Blood pressure < 150/90 mm Hg (%)	58,342 (80)	72,928	30,280 (79.3)	38,184	1.04 (1.01, 1.08) P = 0.0059 0.88% increase
104	McGovern M, 2008 [57]k, Scotland	Cross sectional	Coronary Heart Disease	-	• Total cholesterol $\leq$ 5 mmol/L (%)	47,872 (75.5)	63,407	19,197 (86.3)	22,245	0.49 (0.47, 0.51) P < 0.00001 **12.51 % decrease**

No.	1st Author Year[Ref] Setting	Study design	Condition	Outcome measured		Results				Odds Ratio M-H, (95%CI) % Change between P4P and nonP4P
				Clinical information recorded	Quality indicators reached	P4P Number Achievement (%)	P4P Total	Non-P4P Number achievement (%)	Non-P4P Total	
105	Simpson CR, 2006 [59]a, Scotland	Cross sectional	Stroke	• Record of smoking status (%)	-	27,019 (90.6)	29,822	8,990 (41.1)	21,873	13.81 (13.18,14.48) P < 0.00001 120.44% increase
106	Simpson CR, 2006 [59]b, Scotland	Cross sectional	Stroke	• Record of smoking cessation advice (%)	-	13,021 (95)	13,706	3,081 (79)	3,900	5.05 (4.53, 5.63) P < 0.00001 20.25% increase
107	Simpson CR, 2006 [59]c, Scotland	Cross sectional	Stroke	• Record of total cholesterol (%)	-	24,067 (80.7)	29,823	6,247 (28.6)	21,843	10.44 (10.02,10.88) P < 0.00001 182.18% increase
108	Simpson CR, 2006 [59]d, Scotland	Cross sectional	Stroke	• Record of blood pressure (%)	-	28,470 (95.4)	29,843	13,554 (62)	21,861	12.71 (11.96,13.50) P < 0.00001 53.87% increase
109	Simpson CR, 2006 [59]e, Scotland	Cross sectional	Stroke	• Record of Anti-platelet or anticoagulant therapy (%)	-	24,674 (88.2)	27,975	11,744 (55.9)	21,009	5.90 (5.64, 6.17) P < 0.00001 57.7818% increase
110	Simpson CR, 2006 [59]f, Scotland	Cross sectional	Stroke	• Record of Influenza vaccination (%)	-	21,068 (81.3)	25,914	10,113 (47.1)	21,471	4.88 (4.69, 5.09) P < 0.00001 72.62% increase

No.	1st Author Year[Ref] Setting	Study design	Condition	Outcome measured		Results				Odds Ratio M-H, (95%CI)
				Clinical information recorded	Quality indicators reached	P4P		Non-P4P		% Change between P4P and nonP4P
						Number Achievement (%)	Total	Number achievement (%)	Total	
111	Simpson CR, 2006 [59]g, Scotland	Cross sectional	Stroke	• Record of Body mass index (%)	-	24,607 (88.2)	27,899	11,725 (53.6)	21,875	6.47 (6.19, 6.77) P < 0.00001 64.5522% increase
112	Simpson CR, 2006 [59]h, Scotland	Cross sectional	Stroke	• Record of MRI/CT scan (%)	-	6,231 (71.8)	8,678	1,062 (28.3)	3,753	6.45 (5.93, 7.02) P < 0.00001 153.71% increase
113	Simpson CR, 2006 [59]i, Scotland	Cross sectional	Stroke	-	• Total cholesterol ≤ 5 mmol/L (%)	16,018 (68.1)	23,521	2,551 (65.8)	3,877	1.11 (1.03, 1.19) P = 0.0045 3.4954% increase
114	Simpson CR, 2006 [59]j, Scotland	Cross sectional	Stroke	-	• Blood pressure ≤ 150/90 mmHg (%)	20,658 (76.5)	27,004	10,145 (75.2)	13,491	1.07 (1.02, 1.13) P = 0.0038 1.73% increase
115	Kiran T, 2012[66] , Canada	Cohort	Diabetes	•Record of 1 retinal eye test, 4 hemoglobin A1c tests, and 2 cholesterol tests (%)	-	15,910 (27)	58,927	9,428 (16)	58,927	1.94 (1.89, 2.00) P < 0.00001 68.75% increase
116	Inoue Y, 2011[70], Japan	Cohort	Rehabilitation	-	• Patient return home rate (%)	305 (81.80)	373	427 (80.60)	530	1.08 (0.77, 1.52) P = 0.6495 1.4888% increase
117	Adaji A, 2013[85]a, Australia	Cohort	Diabetes	• Record of HbA1c (%)	-	70,528 (86.11)	81,904	313,600 (77.21)	406,165	1.83 (1.79, 1.87) P < 0.00001 11.53% increase

No.	1st Author Year[Ref] Setting	Study design	Condition	Outcome measured		Results				Odds Ratio M-H, (95%CI)
				Clinical information recorded	Quality indicators reached	P4P		Non-P4P		% Change between P4P and nonP4P
						Number Achievement (%)	Total	Number achievement (%)	Total	
118	Adaji A, 2013[85]b, Australia	Cohort	Diabetes	•Record of HDL (%)	-	39,928 (54.05)	73,873	170,630 (42.01)	406,165	1.62 (1.60, 1.65) P < 0.00001 28.66% increase
119	Adaji A, 2013[85]c, Australia	Cohort	Diabetes	•Record of micro albumin (%)	-	55,162 (71.27)	77,398	191,866 (57.97)	330,974	1.80 (1.77, 1.83) P < 0.00001 22.94% increase

SUMMARY OF THE SEPARATE PERFORMANCE ESTIMATES (Continued)

No.	1st Author Year[Ref] Setting	Study design	Condition	Outcome measured	Interrupted Time Series Analysis
120	Alshamsan R, 2012 [64]a, The United Kingdom	Interrupted Time Series (ITS)	Diabetes Diabetes	• Mean HbA1c level, % (95% CI) • Mean total cholesterol level, mmol /L (95% CI)	Non-P4P trend: -0.21(-0.23 to -0.18), P < 0.01 Level change P4P: 0.04 (-0.04 to 0.12) P4P trend: 0.19 (0.15 to 0.22), P < 0.01
121	Alshamsan R, 2012 [64]b, The United Kingdom	Interrupted Time Series (ITS)	Diabetes	• Mean systolic blood pressure, mm Hg (95% CI)	Non-P4P trend: -0.13 (-0.15 to -0.11), P < 0.01 Level change P4P: -0.12 (-0.18 to -0.06), P < 0.01 P4P trend: 0.03 (0.01 to 0.05), P < 0.05
122	Alshamsan R, 2012 [64]c, The United Kingdom	Interrupted Time Series (ITS)	Diabetes Diabetes	• Mean HbA1c level, % (95% CI) • Mean total cholesterol level, mmol /L (95% CI)	Non-P4P trend: -0.03 (-0.31 to 0.25), Level change P4P: -1.95 (-2.87 to -1.02), P <0.01 P4P trend: -1.04 (-1.42 to -0.64), P <0.01
123	Alshamsan R, 2012 [64]d, The United Kingdom	Interrupted Time Series (ITS)	Diabetes	• Mean systolic blood pressure, mm Hg (95% CI)	Non-P4P trend: -0.84 (-1.00 to -0.67), P <0.01 Level change P4P: -0.51 (-1.05 to 0.01) P4P trend: 0.19 (-0.03 to 0.41)

APPENDIX F

POSTER PRESENTATION AT THE 22ND COCHRANE COLLOQUIUM, HYDERABAD IN INDIA IN 21-26 SEPTEMBER 2014

The effects of pay-for-performance in health care systems with universal health coverage: a systematic review

Sermsuti-anuwat,N.[1], Pavananunt,P.[2], Rattanamongkolgul,S.[3], Wibulpolprasert,S.[4]

1 Post-graduate student, Department of Public Health Administration, Faculty of Public Health, Mahidol University, Thailand
2 Department of Public Health Administration, Faculty of Public Health, Mahidol University, Thailand
3 Department of Preventive and Social Medicine, Faculty of Medicine, Srinakharinwirot University, Thailand
4 International Health Policy Program (IHPP), Thailand

Mahidol University International College

Background:
Currently, there has been clear evidence about the effectiveness of P4P on the achievement of universal health coverage (UHC) in the healthcare system. Previous systematic reviews did not provide information for the context of UHC.

Objectives:
The main purpose of this systematic review was to examine the effectiveness of P4P for improving health service quality and the accessibility of care in the UHC context.

Methods:
Searches were carried out in five electronic databases: Cochrane Library, MEDLINE, PUBMED, EBSCO, and CINAHL. Only the research papers published in English between 2000 and 2013 were included in this review with the following inclusion criteria: (1) the studies were conducted in the countries that provided UHC; (2) P4P were implemented at the supply side; (3) the quality performances of outcomes were reported. The quality of the studies was then assessed by using the modified version of EPOC and NOS.

Results:
The electronic search obtained 2,264 publications of which 23 papers met all the inclusion criteria. Most of the studies reported that the achievement of service quality outcome reached the set targets in the period of P4P implementation than in the non-P4P period. Particularly, P4P scheme implemented in context of the lower baseline of performance shown more improvements than in the higher baseline, as shown in Figure A. In addition, the P4P scheme could enhance a steady increase in both quality of services and accessibility in the first three years of implementation; after this, the growth rates were declined, but still shown improvements, as depicted in Figure B.

Conclusion:
Pay-per-performance (P4P) incentive scheme was found to be more efficient in the low productive health care service areas such as health services for the underprivileged population. Furthermore, the effectiveness of P4P last for a reliable period; after that a new incentive scheme may be considered to boost the outcomes.

P4P is most effective in the unproductive health service areas.

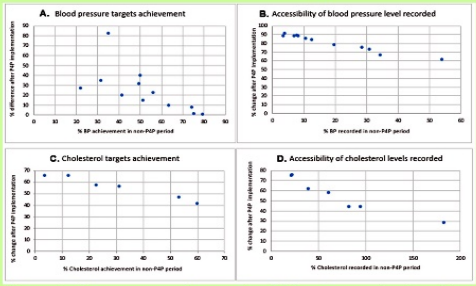


Figure A: The percentage change after P4P implementation (y-axis) when non-P4P was baseline (x-axis).

P4P is effective in the first three years of the implementation.

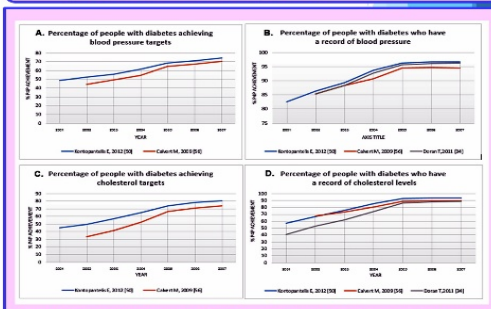


Figure B: The percentage of diabetes patients achieving quality targets.

BIOGRAPHY

NAME	Nithimar Sermsuti-anuwat
DATE OF BIRTH	10 October 1975
PLACE OF BIRTH	Bangkok, Thailand
INSTITUTION ATTENDED	Chulalongkorn University, 1993-1999 DDS.
NAME ADDRESS	119/887 Soi Rattanathibet18 Rattanathibet Road, Bangkasao, Nonthaburi, 11000
EMPLOYMENT ADDRESS	Tuntapootorn And Phuean Company Limited, 119/887 Soi Rattanathibet18 Rattanathibet Road, Bangkasao, Nonthaburi, 11000
PRESENTATION	**Oral Presentation** Implementation of Oral Health Program in School for Mentally Challenged Children: a Case Study in Southern Thailand: Asia-Pasific Community Based Rehabilitation convention 13-15 November 2010, Kuala Lumpur, Malaysia **Poster Presentation** Oral health network for disabled people in Thailand: iADH2012, 21st Congress of the International Association for Disability and Oral Health 28-31 October 2012, Melbourne, Australia **Poster Presentation** The effects of Pay-for-performance in health care systems with universal health coverage: a systematic review: 22nd Cochrane Colloquium; Evidence-Informed Public Health: Opportunities and challenges, 21-26 September 2014, Hyderabad, India