Advantages, disadvantages and feasibility of the introduction of 'Pay for Quality' programmes in Belgium

KCE reports 118C

Federaal Kenniscentrum voor de Gezondheidszorg
Centre fédéral d'expertise des soins de santé
Belgian Health Care Knowledge Centre
2009

The Belgian Health Care Knowledge Centre

Introduction : The Belgian Health Care Knowledge Centre (KCE) is an organization of public interest, created on the 24th of December 2002 under the supervision of the Minister of Public Health and Social Affairs.
KCE is in charge of conducting studies that support the political decision making on health care and health insurance.

Administrative Council

Actual Members : Pierre Gillet (President), Dirk Cuypers (Vice-president), Jo De Cock (Vice-president), Frank Van Massenhove (Vice-president), Yolande Avondtroodt, Jean-Pierre Baeyens, Ri de Ridder, Olivier De Stexhe, Peter Degadt, Daniel Devos, Jean-Noël Godin, Floris Goyens, Jef Maes, Pascal Mertens, Raf Mertens, Marc Moens, François Perl, Marco Schetgen, Yves Smeets, Patrick Verertbruggen, Michel Foulon, Myriam Hubinon

Substitute Members : Rita Cuypers, Christiaan De Coster, Benoît Collin, Lambert Stamatakis, Karel Vermeyen, Katrien Kesteloot, Bart Ooghe, Frederic Lernoux, Anne Vanderstappen, Paul Palsterman, Geert Messiaen, Anne Remacle, Roland Lemeye, Annick Poncé, Pierre Smiets, Jan Bertels, Catherine Lucet, Ludo Meyers, Olivier Thonon.

Government commissioner : Roger Yves

Management

Chief Executive Officer a.i. : Jean-Pierre Closon

Information

Federaal Kenniscentrum voor de gezondheidszorg - Centre fédéral d'expertise des soins de santé –
Belgian Health Care Knowlegde Centre.
Centre Administratif Botanique, Doorbuilding (10th floor)
Boulevard du Jardin Botanique 55
B-1000 Brussels
Belgium
Tel: +32 [0]2 287 33 88
Fax: +32 [0]2 287 33 85
Email : info@kce.fgov.be
Web : http://www.kce.fgov.be

Advantages, disadvantages and feasibility of the introduction of 'Pay for Quality' programmes in Belgium

KCE reports 118C

LIEVEN ANNEMANS, PAULINE BOECKXSTAENS, LIESBETH BORGERMANS, DELPHINE DE SMEDT,
CHRISTIANE DUCHESNES, JAN HEYRMAN, ROY REMMEN, WALTER SERMEUS,
CARINE VAN DEN BROEKE, PIETER VAN HERCK, MARC VANMEERBEEK,
SARA WILLEMS, KRISTEL DE GAUQUIER

Federaal Kenniscentrum voor de Gezondheidszorg
Centre fédéral d'expertise des soins de santé
Belgian Health Care Knowledge Centre
2009

KCE reports 118C

Title :	Advantages, disadvantages and feasibility of the introduction of 'Pay for Quality' programmes in Belgium
Authors :	Lieven Annemans (UGent), Pauline Boeckxstaens (UGent), Liesbeth Borgermans (KULeuven), Delphine De Smedt (UGent), Christiane Duchesnes (Université de Liège), Jan Heyrman (KULeuven), Roy Remmen (Universiteit Antwerpen), Walter Sermeus (KULeuven), Carine Van Den Broeke (KULeuven), Pieter Van Herck (KULeuven), Marc Vanmeerbeek (Université de Liège), Sara Willems (UGent), Kristel De Gauquier (KCE)
Reviewers:	Christian Léonard (KCE), Imgard Vinck (KCE)
External experts:	Hugo Casteleyn (AZ Sint-Blasius Dendermonde), Pierre Chevalier (UCL), Xavier de Béthune (Alliance Nationale des Mutualités Chrétiennes), Pierre Gillet (Université de Liège), Margareta Haelterman (Federal Public Service Health, Food Chain Safety and Environment), Annelies Van Linden (Domus Medica)
Acknowledgements:	Belgium: Bert Aertgeerts (CEBAM, KULeuven), Dirk Broeckx (APB), Geneviève Bruwier (ULg, SSMG, Forum des Associations de Médecins Généralistes), Piet Calcoen (DKV), Xavier de Béthune (Alliance Nationale des Mutualités Chrétiennes), Jo De Cock (RIZIV/INAMI), Christiaan Decoster (Federal Public Service Health, Food Chain Safety and Environment), Peter Degadt (Zorgnet Vlaanderen), Didier de Laminne de Bex (DKV), Jan De Maeseneer (UGent), Ri De Ridder (RIZIV/INAMI), Daniel Désir (CHU Brugmann), Jos Desmedt (Domus Medica), Jacques De Toeuf (CHIREC, ABSYM), Alain De Wever (ULB), Pierre Drielsma (Maisons Médicales), Guy Durant (UCL), Micky Fierens (LUSS), Pierre Gillet (ULg), Johan Hellings (Ziekenhuis Oost-Limburg), Marc Justaert (Landsbond der Christelijke Mutualiteiten), Johan Kips (UZLeuven), Jean-Marc Laasman (Union nationale des mutualités socialistes), Roger Lonfils (Communauté française Wallonie-Bruxelles, Direction générale de la Santé), Catherine Lucet (Union nationale des mutualités socialistes), Pascal Meeus (RIZIV/INAMI), Marc Moens (BVAS), Dominique Pestiaux (UCL), Michel Roland (ULB, CEBAM), Robert Rutsaert (ASGB), Marco Schetgen (ULB), Erik Schokkaert (KULeuven), Carine Serano (Ligue des Usagers des Services de Santé), Piet Vandenbussche (UGent), Walter Van Den Eede (Agentschap Inspectie Welzijn, Volkgezondheid en Gezin), Chris Vander Auwera (Flemish Agency for Care and Health), Philippe Vandermeeren (Groupement Belge des Omnipraticiens), Michel Van Hoegaarden (Federal Public Service Health, Food Chain Safety and Environment), Bernard Vercruysse (UCL, Forum des Associations de Médecins Généralistes), Michel Vermeylen (Association des Médecins de Famille, ABSYM), Arthur Vleugels (KULeuven), Ilse Weeghmans (Vlaams Patiëntenplatform) International: Mark Ashworth (King's College, London, UK), Jozé Braspenning (Scientific Institute for Quality of Health Care, the Netherlands), Alyna Chien (University of Chicago, US), Tim Coleman (University of Leeds, UK), Sheryl Damberg (RAND, US), Stephen Duckett (Queensland Health Government, Australia & Alberta Health Services, Canada), Adams Dudley (University of California, US), Robert Fleetcroft (University of East Anglia, UK), Christine Gardel (Haute Autorité de Santé, France), Ruth McDonald (University of Manchester, UK), Gary McLean (University of Glasgow, Scotland, UK), Peggy McNamara (AHRQ, US), Christopher Millett (Imperial College, London, UK), Karen Murphy (Temple University, Philadelphia, US), Martin Roland (University of Manchester/ Cambridge, UK), Meredith Rosenthal (Harvard University, US), Ian Scott (Princess Alexandra Hospital, Queensland, Australia), Colin

Simpson (University of Aberdeen, Scotland, UK), Abd Tahrani (University of Birmingham, UK)

External validators: Marc Jegers (VUB), Matthew Sutton (University of Manchester, UK), Cordula Wagner (Nivel & Free University Amsterdam medical centre, the Netherlands)

Conflict of interest: Hugo Casteleyn regularly speeches at seminars or at training sessions for health care institutions. Xavier de Béthune works for an organisation that may gain or lose financially when P4Q programmes are implemented.

Disclaimer: The external experts collaborated on the scientific report that was subsequently submitted to the validators. The validation of the report results from a consensus or a voting process between the validators. Only the KCE is responsible for errors or omissions that could persist. The policy recommendations are also under the full responsibility of the KCE.

Layout: Ine Verhulst

Brussels, 16th November 2009

Study nr 2008-12

Domain: Health Services Research (HSR)

MeSH: Salaries and Fringe Benefits; Reimbursement, Incentive; Fees and Charges; Quality Assurance, Health Care; Quality Control; Health Services Accessibility

NLM classification: W 84.4

Language: English

Format: Adobe® PDF™ (A4)

Legal depot: D/2009/10.273/52

Any partial reproduction of this document is allowed if the source is indicated. This document is available on the website of the Belgian Health Care Knowledge Centre.

How to refer to this document?

Annemans L, Boeckxstaens P, Borgermans L, De Smedt D, Duchesnes C, Heyrman J, Remmen R, Sermeus W, Van Den Broecke C, Van Herck P, Vanmeerbeek M, Willems S, De Gauquier K. Advantages, disadvantages and feasibility of the introduction of 'Pay for Quality' programmes in Belgium. Health Services Research (HSR). Brussels: Belgian Health Care Knowledge Centre (KCE). 2009. KCE Reports 118C. D/2009/10.273/52



Executive summary

INTRODUCTION AND SCOPE

This study has to be seen as a logical sequel to previous KCE reports on quality, i.e. report 41 on clinical quality indicators in hospitals and report 76 on quality improvement in general practice. In this study, however, the focus is on financial incentives related to quality achievement.

Pay for Performance (P4P) or Pay for Quality (P4Q), which focalizes exclusively on the quality component of performance, is the mechanism which directly relates the remuneration of delivered care to the achieved result on structure, process and/or outcome indicators.

There is an increasing interest in P4Q, abroad as well as in our own country. Potential future Belgian P4Q initiatives are, however, more likely to reach their goal when they are based on lessons learned abroad and on a conceptual framework applicable to the Belgian setting. Therefore, this project aims to answer the following research questions:

1. What can be learned from international P4Q models on
   - evaluation: what are the beneficial effects and unintended consequences of P4Q programmes?
   - design and implementation: which conceptual framework can be applied for the Belgian healthcare system; how should financial incentives be designed; what are the critical success factors?
2. What are the necessary conditions required for applying international P4Q models or for adding P4Q components to Belgian quality improvement initiatives?
   - What are the current initiatives in Belgium? Is there any evidence of their impact on quality?
   - To what extent are the current financing schemes, databases and other tools (guidelines, quality indicators) appropriate to implement P4Q in the Belgian health care setting? What are the most important facilitating and hindering factors?

The scope of this study is restricted to primary care (general practitioners and specialist care) and hospital care. In addition, P4Q is distinguished from other quality improvement interventions, such as accreditation and public reporting of differences in quality between practitioners, since these initiatives do not necessarily include a financial incentive and therefore are not part of this report.

METHODOLOGY

For the first research question, the following methods are applied:

- construction of a conceptual framework, based on theoretical literature;
- systematic review of the evidence on the effects of the use of a P4Q programme;
- revision of the conceptual framework on the basis of the literature findings;
- consultation of international experts involved in P4Q in their country (i.e. USA, UK, the Netherlands and Australia).

In this executive summary, the results of the latter two aspects are integrated in the conclusion section.

Both the conceptual framework and the international evidence are based on a simultaneous systematic review of both theoretical and empirical peer reviewed scientific literature. The review was performed in a two-phased approach: the first focus was on systematic reviews, which were subsequently completed with primary more recent studies.

For the second research question, the methodology consists of:

- listing Belgian P4Q initiatives;
- theoretical evaluation of the feasibility of P4Q implementation;
- consultation of Belgian stakeholders.

CONCEPTUAL FRAMEWORK

The P4Q conceptual framework was developed to represent all relevant aspects of P4Q and its application in practice. Central in the model is the relationship between the desired quality achievement and the financial incentive. The nature of the incentive and the way quality is defined and measured can be quite different between programmes. Equally important is the relationship between the payer (i.e. provider of incentive) and the health care provider. Here again, the characteristics of these stakeholders and their relationship will affect the success of a P4Q programme. In addition, the model emphasizes the need to account for characteristics of the patients, as well as the overall health system (i.e. social security or NHS, type of prevailing physician payment system, etc…). The implementation of the programme must follow a "Plan Do Check Act" (PDCA) logic, in which room and efforts are foreseen to regularly seek input from all stakeholders involved and undertake a continuous evaluation of the programme's effects.

EVIDENCE FROM PEER REVIEWED LITERATURE

STUDY CHARACTERISTICS

The international literature is dominated by evidence from Anglo-Saxon countries. The majority of studies focus on the primary care setting. In the US, however, there are also reports on P4Q in the hospital setting. Australia, the Netherlands, Germany, Italy and Spain have only just started to report the use of P4Q as a health system intervention. Details on these start-up countries can be found in the scientific summary.

CONTEXTUAL ASPECTS

Most P4Q studies have to be interpreted within the context of the country at stake. This context is described in the scientific summary.

PATIENT CHARACTERISTICS

Medical conditions targeted by P4Q mainly include preventive care (e.g. screening) and chronic care (e.g. diabetes care). P4Q initiatives in acute care focus on conditions such as myocardial infarction and pneumonia. Some studies are not medical condition or patient group specific, but focus on a wider use of disease management or care management processes.

Inclusion of patients is often based on specific clinical selection criteria to define a clinically coherent patient group. Additional patient selection criteria included in a number of studies are: patient age (children, adults, elderly), the continuity of the patient-physician relationship, patient language, socio-economic status and ethnic background.

P4Q INTERVENTIONS

In the UK, P4Q is embedded in the national contract between the NHS and the general practitioners (GPs) as the Quality and Outcomes framework (QOF). In 2004, quality is measured against a set of 146 quality indicators relating to clinical care, organization of care and patient experience. Practices can earn points for each indicator; each point representing £ 125. Up to 25% of the GPs' income is determined by the QOF programme. Data on quality of care from each practice is being automatically extracted from the electronic health record of the practices. The programme allows practices to exclude patients who they deem inappropriate from specific indicators (i.e. exception reporting). Recently, a P4Q demonstration programme in hospital care was launched.

In the USA, the characteristics of different P4Q interventions are very diverse. Most USA programmes define thresholds to be met. The indicators used are mostly structural, process and/or outcome measures. The incentive determines a certain percentage of the providers' or the organizations' income, this percentage ranges from 0 to 12%. The base for the incentive calculation and the target level vary between programmes. Many P4Q programmes use a bonus given only to the best performing providers.

IMPLEMENTATION AND COMMUNICATION

A first attempt in the mid-1980's to implement a P4Q programme in the UK, was rejected by the GPs. During the 1990s, evidence based medicine was introduced and health professionals gradually accepted the possibility to define and measure the quality of care. In a second attempt, the government decided to provide a substantial increase in health care funding and negotiations took place between government, professionals and academics. This has led to the implementation of QOF in 2004. Participation in the QOF programme is voluntary.

In the USA, implementation and communication support is provided in different ways: involvement of providers/peers, involvement of patients, office staff and educators, leadership support or the use of a small core work group. In addition, there are surveys, reminders and presentations. The level of awareness and acceptance of the programme varies between programmes. Participation is mostly voluntary.

EFFECT OF P4Q PROGRAMMES

Effects in terms of clinical effectiveness range from negative, or absent up to positive or very positive, depending on the target and programme. Negative results are only found in a minority of cases. Table 1 lists the targets/indicators for which a positive effect of at least 5% was found. According to the study design, the evidence is labelled as strong (randomized studies and concurrent plus historical comparison design studies without randomization) or weak (historical comparison design studies with multiple time points).

Table 1: Indicators on which P4Q programmes had a positive clinical effect

Patient group	Indicator	Evi-dence	(Range of) Effect size
Preventive care			
Influenza	Immunization rate	strong	6.8-8.4%
Acute care			
Myocardial infarction	Timely (i.e. within 120 min after arrival) percutaneous intervention rate	strong	5.4%
Heart failure (acute)	Provision of discharge instructions	strong	25.5%
CABG	Discharged to home rate	weak	10%
Community acquired pneumonia	Pneumococcal screening and/or vaccination rate	strong	9.5-44.7%
Chronic care			
Diabetes	Hba1c below threshold rate	strong	13.2-19.3%
	Lipid or cholesterol below threshold rate	strong	29.9%
	Blood pressure below threshold rate	strong	1.6-18%
	Nephropathy testing rate	strong	10%
	Foot exam rate	strong	2.7-45.4%
	Peripheral pulse testing rate	weak	4.9-59%
	Provision of smoking cessation advice	weak	12-35.5%
	Pneumococcal vaccination rate	weak	24.3%
	Overall diabetes performance	weak	7.5%, 6.9%
Heart failure (chronic)	Use of ACE inhibitor or angiotensin receptor blocking	weak	23.4%
Smoking cessation	Smoking status recording	strong	7.9-24%
	Referral rate to a tobacco cessation service	strong	6.2%

EQUITY

The impact of P4Q on equity in health care has only been investigated in the QOF. Since no information on access was found, the main focus of this report was on equity in treatment and treatment outcomes.

The extent to which different patient groups benefit from P4Q tends to vary and be highly dependent on the type and complexity of the indicator(s) under study, the observed patient groups, the characteristics of the study and the level of detail of the indicators. Hence, it is difficult to draw firm conclusions regarding equity. In general, all citizens benefit from the improvements in quality of care and the extent to which they benefit determines whether the existing health gap narrows (when the worse off have a larger growth than the best off) or increases. More studies suggest a narrowing of the gap (for age, socio-economic status and ethnicity), however, for some indicators (i.e. gender) new gaps arise. Further research is needed to understand the mechanisms behind these observations.

COST-EFFECTIVENESS

Only three studies were identified that focus on cost-effectiveness of P4Q although it is crucial to know whether the money spent in P4Q was well spent. From the twelve QOF-indicators, investigated in the UK study, only paying for increased diabetes retinal screening seemed not cost-effective. A US study focussing on P4Q programmes in primary diabetic care showed a positive return on investment. In the third study which evaluated an American hospital P4Q programme focussing on heart care, the programme seems to be cost-effective even in a worst case scenario.

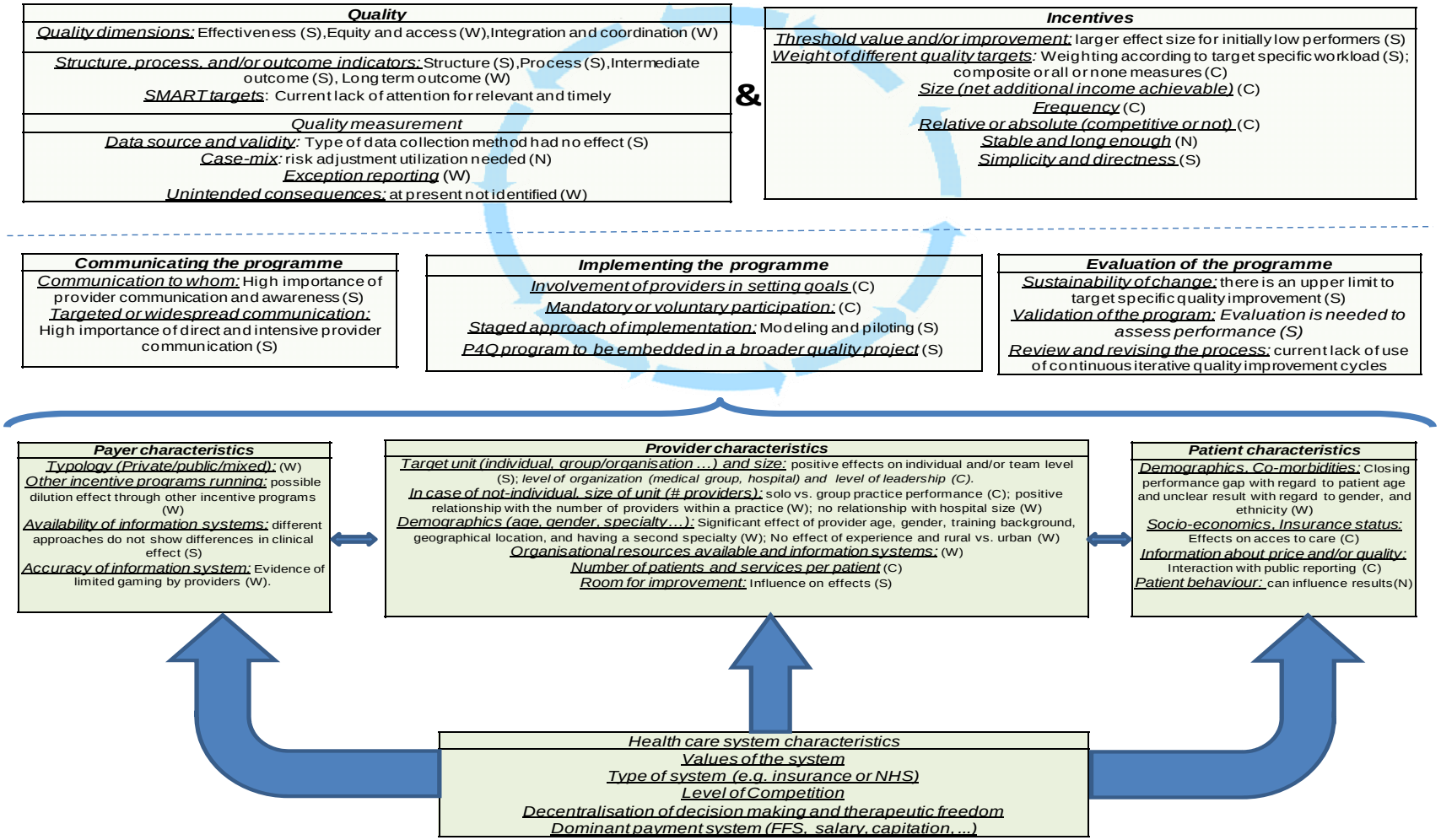
OTHER EFFECTS

Comparable to existing payment systems, P4Q can have unintended consequences such as gaming, i.e. finding ways to maximize the measured results without actually accomplishing the desired objectives. Patient selection is one example of gaming. Other unintended consequences can be diversion of attention away from important aspects of care that are not included in the P4Q program. On the other hand, positive consequences have been demonstrated such as a positive spillover effect and a decrease in the gap in performance between providers.

REVISED CONCEPTUAL FRAMEWORK

Figure 1 incorporates the results from the literature study in the conceptual model. For each item from the model, the strength of the evidence and the direction of the evidence are indicated ('S' = strong evidence, 'W' = weak evidence, 'C' = conflicting evidence, 'N' = no evidence).

Figure 1: P4Q conceptual framework



PAY FOR QUALITY IN BELGIUM

DESCRIPTION OF QUALITY INITIATIVES IN BELGIUM

Only a few of the current Belgian programmes involve a kind of financial incentive. None, however, fits the definition of P4Q which is based on the actual measurement of quality and the linkage of those measurement results to the allocation of a financial incentive.

A total of 14 programmes are described:

- in primary and hospital care: care itineraries, clinical pathways, providers' accreditation;
- only in primary care: breast cancer screening prevention bonus, capitation funding revaluation in the medical houses, European Practice Assessment (EPA) tool, global medical record, prescription feedback of the National council for Quality Promotion, preventive module in global medical record;
- only in hospital care: centres of reference, hospital accreditation, hospital benchmarking, Quality and Patient Safety Framework for hospitals, reference payment.

There are still some weaknesses in these initiatives. First, providers are sometimes free to participate or free to choose the indicators they want to comply with. Consequently, many of these are not evidence based. Second, indicators are often not measured and, finally, unintended consequences are not sufficiently monitored.

A strength of many initiatives is their dynamical approach. Many of them are meant to evolve and change over time, both regarding targets as regarding the quality improvement process itself. Another strength is the fact that twelve initiatives do not exclusively address efficiency and therefore offer a P4Q opportunity.

FEASIBILITY OF P4Q IN BELGIUM

The analysis of existing initiatives learned that there is a substantial body of knowledge and experience that can be leveraged to assist in P4Q implementation.

As already pointed out in previous KCE reports, practical feasibility depends on the level of data validity, which is currently considered insufficient in Belgium, especially in primary care.

At present, quality of care is often considered to be the provider's responsibility only. P4Q, however, will imply a shift toward an external quality auditing approach.

Another issue of utmost importance for a successful P4Q implementation is the support by all stakeholders i.e. providers, payers, patients and policy makers.

BELGIAN STAKEHOLDERS' VIEW ON P4Q

Most stakeholders suggest a staged approach. As a first step, a broad range of potential quality improving initiatives and projects (that have proven their success in the past) should be listed. It is suggested that domains for which the needs are urgently displayed, the benefits clearly documented and the targets widely accepted should be considered first. Hospital care and primary care are both considered a priority.

There is, however, no consensus on the type of initiatives that should be launched first. Many stakeholders propose to start with initiatives that have a traditional disease-related scope, but some advocate a practice-related scope. Chronic diseases seem to be the most obvious choice.

Stakeholders mentioned the importance of a good balance between a centrally-led institute and locally managed initiatives.

CONCLUSIONS AND RECOMMENDATIONS

The main conclusion of the literature is that there is a strong variation in results between the programmes: some P4Q programmes have a large positive effect on the quality indicators but these are rather rare in comparison with the programmes with a moderate effect. Until now, negative results in terms of effect were hardly found and unintended consequences seem limited.

When Belgian policy makers should decide to apply P4Q principles to existing initiatives, they first should take into account the recommendations formulated in the KCE reports 41 and 76, which remain valid in the P4Q context.

In addition, the following is recommended:

- ***Quality*** can be measured by structure, process as well as intermediate outcome indicators provided they are supported by evidence. Several examples are given in previous KCE reports (e.g. diabetes, medical houses, clinical quality indicators).
- We recommend to reward everybody who achieves the quality targets and not only those with the best results. The ***incentive*** should be targeted at all providers, at an individual as well as organizational level.
- The ***implementation*** of the P4Q programme has to be done gradually (i.e. first an assessment of potential cost-effectiveness) and with **pilot programmes** as an add-on to existing quality improvement initiatives.
- Accurate, validated and already available ***data*** are preferred. This implies:
  - an investment in **IT development** to set up a system in which data are extracted automatically from the electronic health record.
  - an **audit system** to assure the quality of data, with penalties whenever fraud is discovered.
- A ***monitor system*** of overall impact, potential unintended consequences, effectiveness and cost-effectiveness with feedback to health professionals must be provided from the start.

Scientific summary

Table of contents

1 BACKGROUND AND OBJECTIVES 4

2 METHODS 6

2.1 MIXED METHODS APPROACH 6
2.1.1 Conceptual framework and international evidence 6
2.1.2 Selection of and data-extraction from articles related to equity 11
2.1.3 Revision of the conceptual framework into an evidence based state of the art framework 11
2.1.4 Consultation of international experts as P4Q involved country representatives 11
2.1.5 Conducting a systematic review of Belgian P4Q initiatives, and evaluating the feasibility of P4Q implementation, based on the comparison of current Belgian policy vs. findings from literature. 13
2.1.6 Consultation of Belgian stakeholders as involved country representatives 13

3 A THEORETICAL FRAMEWORK FOR P4Q 16

3.1 P4Q CONCEPTS 16
3.1.1 Quality 16
3.1.2 Incentives 21
3.2 THE P4Q CONTEXT 24
3.2.1 The health care system 24
3.2.2 The payer-provider relationship 25
3.2.3 The patients 28
3.3 IMPLEMENTING AND COMMUNICATING THE PROGRAM 29
3.4 P4Q PRINCIPLES FOR DESIGN AND IMPLEMENTATION 31

4 THE EVIDENCE BASE FOR P4Q: A SYSTEMATIC REVIEW OF THE PEER REVIEWED LITERATURE 35

4.1 INTRODUCTION 35
4.2 DESCRIPTION OF STUDIES 35
4.2.1 General description 35
4.2.2 Contextual aspects 36
4.2.3 Patient characteristics 41
4.2.4 Description of P4Q interventions 43
4.2.5 Implementing and communicating P4Q programmes 48
4.2.6 Existing and concurrent quality improvement initiatives 51
4.3 GENERAL EVIDENCE 54
4.3.1 Reported effect of Pay for Quality programmes 54
4.4 REPORTED COST EFFECTIVENESS AND MODELLING EFFECTS OF P4Q PROGRAMMES 59
4.4.1 Cost-effectiveness of Pay for Quality 59
4.4.2 Modelling costs or effectiveness 61
4.5 EVIDENCE RELATED TO EQUITY 66
4.5.1 Reported evidence on equity 66

5 REVISING THE CONCEPTUAL FRAMEWORK BASED ON EVIDENCE 71

5.1 INTRODUCTION 71
5.1.1 Quality goals and targets 71
5.1.2 Quality measurement 72
5.1.3 P4Q incentives 74
5.1.4 Implementing and communicating the programme 75
5.1.5 Evaluation of the programme 76
5.1.6 Health care system characteristics 77
5.1.7 Payer characteristics 79
5.1.8 Provider characteristics 80

5.1.9 Patient characteristics 83
5.2 DISCUSSION AND CONCLUSIONS 83
5.2.1 Reported effect of Pay for Quality programmes 83
5.2.2 Reported cost effectiveness and modelling effects of P4Q programmes 87
5.2.3 Reported impact of P4Q on equity 89
5.2.4 Revising the conceptual framework based on evidence 92
6 INTERNATIONAL COMPARISON 96
6.1 INTRODUCTION 96
6.2 DESCRIPTION 96
6.2.1 Health care system characteristics 96
6.2.2 Existing P4Q interventions 96
6.2.3 P4Q concepts 99
6.2.4 Implementing and communicating the programme 102
6.3 EVIDENCE 104
6.3.1 Likely effect according to the experts 104
6.3.2 Unintended consequences according to the experts 104
6.4 DISCUSSION AND CONCLUSION 105
6.4.1 Key recommendations for design and implementation 105
6.4.2 What the future holds 107
6.4.3 Conclusions 107
7 P4Q IN BELGIUM 109
7.1 DESCRIPTION OF (PAY FOR) QUALITY INITIATIVES IN BELGIUM 109
7.1.1 Introduction 109
7.1.2 Methods 111
7.1.3 Results 111
7.1.4 Discussion 126
7.2 FEASIBILITY STUDY OF IMPLEMENTATION OF P4Q IN BELGIUM 127
7.2.1 Introduction 127
7.2.2 Methods 127
7.2.3 Results 127
7.3 DISCUSSION AND CONCLUSIONS 137
8 PERCEPTIONS AND OPINIONS OF STAKEHOLDERS IN BELGIUM 139
8.1 PART I: THE BELGIAN HEALTH CARE SYSTEM FACING QUALITY OF CARE. 139
8.1.1 Health care system characteristics 140
8.1.2 Payer characteristics 142
8.1.3 Provider characteristics 142
8.1.4 Patient characteristics 143
8.2 PART II: IMPROVING QUALITY THROUGH P4Q 144
8.2.1 Quality and its dimensions, as seen by Belgian stakeholders 144
8.2.2 Patient populations and goals for quality projects 145
8.2.3 Incentives 146
8.2.4 Target audience for P4Q programmes 149
8.2.5 Quality measurement in P4Q programmes 149
8.2.6 Evaluation of the programme 151
8.2.7 Implementing and communicating the programme 152
8.3 PART III: CONCRETE PLANNING IN THE BELGIAN SITUATION 154
8.3.1 The need for an "evidence collecting institute" 154
8.3.2 A new role for a revised National Council on Quality Promotion (NCQP) 154
8.3.3 Think global, act local: the crucial need for organizing at the local level, and paying for local support initiatives 154
8.3.4 How and where to start in Belgium 155
8.3.5 Schematic view 155
9 DISCUSSION 156
9.1 QUALITY 157
9.1.1 Defining quality 157

9.1.2 Measuring quality 158
9.2 INCENTIVE 159
9.3 THE OVERALL HEALTH CARE SYSTEM 160
9.4 PAYER CHARACTERISTICS 160
9.5 PROVIDER CHARACTERISTICS 160
9.6 PATIENT CHARACTERISTICS 161
9.7 IMPLEMENTATION 161
9.8 TOWARDS P4Q IN BELGIUM? 162
10 REFERENCES 165

1 BACKGROUND AND OBJECTIVES

Research into the quality of healthcare is producing increasing amounts of evidence about treatment underuse, overuse and misuse. [1,2,3,4,5,6,7]

Unintended variability in processes and outcomes were also reported in Belgium, [8,9] and current strategies to tackle these problems do not show clear results. [10,11] One of the most important characteristics of any health system are the financial driving forces. However, 'Fee for service', 'capitation' and 'prospective payment' may influence mainly quantity instead of quality. As a result several studies indicate that current payment systems insufficiently reward the delivery of good care. [12,13,14]

One proposed intervention is to directly relate the remuneration of delivered care to the achieved result on structure, process and/or outcome indicators. This mechanism is known as 'pay for performance' (P4P) or 'pay for quality' (P4Q) (when focusing exclusively on the quality component of performance). The Institute of Medicine (2007) (IOM) explains P4Q as 'the systematic and deliberate use of payment incentives that recognize and reward high levels of quality and quality improvement' [2]. Quality is defined as: 'The degree to which health services for individuals and populations increase the likelihood of desired health outcomes and are consistent with current professional knowledge'. It consists of different dimensions, namely clinical effectiveness of care, interpersonal aspects of care, patient safety, access and equity of care, continuity and/or coordination of care and cost-effectiveness of care.

There is an increasing amount of evidence related to P4Q programmes. However, these programmes are very heterogeneous with regard to the type of incentive, the target health care providers, the applied criteria for quality, the way the programme is implemented and evaluated and the contextual aspects related to the programme.

Several systematic reviews have already been published aiming at finding evidence on what works and what does not. At the same time different authors have tried to create a conceptual and theoretical framework, which could serve as a basis for the design of new programmes.

In Belgium, the interest in P4Q is observed as well, as witnessed by some initiatives.

However, it is felt that without learning from lessons abroad, and the absence of a clear conceptual framework applicable to the Belgian setting, such initiatives may not reach the desired goals for which they are (or should be) intended.

Therefore, the purpose of this project is to answer the following key research questions:

What can be learned from the international P4Q literature regarding the design, implementation and evaluation?

- *Design and implementation.* What are the current system components of a P4Q programme? How are financial incentives designed? What are the critical success factors for the implementation of a P4Q programme?
- *Evaluation.* What are the effects of the use of P4Q programmes, focusing on all relevant quality domains (including access, coordination, equity, cost-effectiveness,…) and also taking into account unintended consequences?

What are the necessary conditions for applying P4Q in Belgium in the context of already existing quality initiatives?

- Design, implementation, evaluation. What are the current initiatives
- in Belgium in the public and private sector to enhance quality of care by way of financial incentives linked to quality of care measures? Is there any evidence of their impact on quality?
- Conditions. To what extent are the current financing scheme, databases and other tools (guidelines, quality indicators) appropriate to implement P4Q in the Belgian health care setting? What are the most important facilitating and hindering factors?

The research questions are in principle relevant to the diverse types of healthcare: general and mental health care, primary care and hospital care. The IOM report on P4Q gives some examples of programmes in skilled nursing facilities and home health agencies [2]. Initiatives in these settings are however still rare and often lack the availability of quality of care data. Being a first explorative study on P4Q in Belgium, our study scope in terms of setting is restricted primary care and medical or surgical hospital care. The similarities and differences between both settings in terms of P4Q will be addressed as a part of the study.

In this study we clearly distinguish P4Q from other quality improvement interventions, such as accreditation and public reporting. These tools do not necessarily contain a financial incentive. However, their influence as a possible co-intervention with P4Q will be discussed.

The report consists of nine chapters. Chapter 2 explains the general methodology that was followed. The next chapters each address a specific research question. Chapter 3 presents a theoretical framework of P4Q. In Chapter 4 the evidence base of P4Q is systematically reviewed. In Chapter 5 a revision of the framework based on the evidence is discussed. Chapter 6 comprises an international comparison, based in interviews with international experts. Chapter 7 gives a description of the current existing P4Q initiatives in Belgium and includes a feasibility study of the implementation of P4Q in Belgium. In Chapter 8 the perceptions and opinions of key stakeholders in Belgium regarding P4Q are addressed. Finally Chapter 9 provides a roundup of the results of the previous chapters, leading to general conclusions and recommendations.

2 METHODS

This Chapter describes the methods used to answer the research questions as stated in Chapter 1. A mixed methods approach is presented in terms of data collection and analysis. Also, the ethical aspects of this study are discussed.

2.1 MIXED METHODS APPROACH

The study makes use of a Mixed Methods design, relating quantitative and qualitative data as part of the research process. These methods are structured as followed:

Research question 1: What can be learned from the international P4Q literature on the design, implementation and evaluation?

- The construction of a conceptual framework, based on theoretical literature. This part aims at answering the following questions: *how should P4Q be modelled, taking into account all theoretically relevant factors? What may be the practical implications that can be derived from theory to support P4Q design and implementation? (See chapter 3)*
- Conducting a systematic review of the evidence base. This part aims at answering the following question: *what is the effect of P4Q on the different quality domains (effectiveness, equity, safety, coordination, cost effectiveness, etc.), describing both intended and unintended consequences? (see chapter 4)*
- Revision of the conceptual framework into an evidence based state-of-the-art framework. *(see chapter 5)*
- Consultation of international experts as P4Q involved country representatives. This part aims at answering the following questions: *how is P4Q applied and how is it influenced by market, payer, provider and other healthcare system characteristics. (see chapter 6)*

Research question 2: What are the necessary conditions for applying P4Q in Belgium, given the already existing quality initiatives?

- Conducting a systematic review of Belgian (P4)Q initiatives. This part aims at answering the following question: what are the current initiatives in Belgium in the public and private sector to enhance quality of care by way of financial incentives linked to quality of care measures? Is there any evidence of their impact on quality?(see chapter 7)
- Evaluating the feasibility of P4Q implementation, based on the comparison of current Belgian policy vs. the evidence coming from research question 1 (see chapter 7)
- Consultation of Belgian stakeholders as involved country representatives. This part aims at answering the following question: to what extent are the current financing scheme, databases and other tools (guidelines, quality indicators) appropriate to implement P4Q in the Belgian health care setting? What are the most important facilitating and hindering factors?(see chapter8)

Each of these sub methods is described in the subsequent sections.

2.1.1 Conceptual framework and international evidence

Both the conceptual framework and the international evidence base are based on a simultaneous systematic review of both theoretical and empirical peer reviewed scientific literature.

The systematic review process was performed in a two-phased approach by three independent reviewers, each a member of different partner research institutions. In the first phase we focussed on existing systematic reviews. Subsequently we completed the evidence by a focused study of the primary studies (phase 2). Figure 1 presents a flow chart of the methods used and their subsequent results.

Searching, application of in- and exclusion criteria and quality appraisal was performed by two reviewers independently (DDS and PVH). In case of non corresponding results, consensus was sought, led by the third reviewer (RR).

To describe evidence about the equity dimension of P4Q models a systematic review process on a selection of publications from the main database was performed by two other reviewers (SW and PB).

2.1.1.1 Search strategy

The sources consisted in both phases of the following electronic databases: Medline, Embase, the Cochrane Library, Web of Science, PsycInfo and Econlit.

An iterative procedure was followed. Publications were first judged on the basis of title review, and excluded if clearly irrelevant. Subsequently, abstract review was performed for all articles that remained after title selection and for the articles of which eligibility was unclear. Finally a full text analysis was performed for all articles for which relevance remained doubtful.

Appendix 1 and 2 give an overview of the search strategy including search string, limits, and number of retrievals per database, respectively during phase 1 and 2. A combination of MeSH terms and non MeSH terms was used wherever possible. A time period of 2000-2008 was applied in the search for systematic reviews (see Appendix 1). The period 2004-2008 or 2005-2008 was used in searching additional primary studies, depending on the date and the used database in the latest existing relevant high quality systematic reviews. (see Appendix 2). In chapter 4 the reviews were not discussed. They only served as a means to set the date in the search for additional primary studies. Hence, only the primary articles included in the selected reviews, and the additional primary articles were taken into account for this study.

After completion of the report, an additional search from January 2009 to July 2009 was performed, to identify more recent relevant studies that could be of value to our report. The results of this additional search are described in appendix 16C.

In addition, all references of relevant publications were screened for additional material. Forward citation tracking was applied on all relevant publications. More than sixty international experts were asked to provide any not yet retrieved relevant publication, using a standard template. The experts were selected based on their own number of publications with regard to P4Q, with a minimum of two publications.

As a final check the following journals were hand searched through their online archives: the New England Journal of Medicine, the Journal of the American Medical Association, the British Medical Journal, the Lancet, Medical care, Health Affairs, Health Policy and Health Economics. This last step didn't yield any unretrieved publication and therefore confirmed a high level of completeness of the search strategy.

The decision to include peer reviewed literature only is based on various reasons. During the first phase of the review (search for existing systematic reviews) a broad comprehensive approach including grey literature was tested through the screening of specific topic related websites and the use of a Google, Scirus and Sumsearch search engine. The websites included the Institute of Medicine, the AHRQ, the Joint Commission, the Leapfrog Group, the RAND corporation, the John Woods Richardson Foundation, etc. next to governmental websites such as the CMS in the USA, NHS resources in the UK and others in Australia. Although this identified many documents concerning P4Q, almost none of these passed the phase of quality appraisal. Almost all of these documents are aimed at a broad audience of P4Q users and therefore omit detailed methodological specifications. Many grey literature sources about P4Q even lack the use of references.

In addition, sources managed by health plans or employer groups showed a conflict of interest, which was exemplified by a marketing style approach of P4Q. It was deemed unreliable to include sources that almost sell P4Q as a product.

Because of these reasons and because of the high number of retrieved publications in peer reviewed scientific literature as an input in the study, grey literature is excluded from the systematic review.

After completion of the report, an additional search from January 2009 to July 2009 was performed, to identify more recent relevant studies, which could be of value to our report. The results of this additional search are described in appendix 16C.

2.1.1.2 Relevance screening

The following inclusion criteria were defined:

- Participants included health care providers in general primary and/or hospital care, being a provider organization, team of providers or an individual provider.
- The intervention is defined as policies, including laws, rules, financial or administrative orders, made by governments, non-governmental organizations (health funds, provider organizations,…), public or private insurers, that specifically and directly intend to affect the quality of care, by means of financial incentives. The intervention can be combined with other interventions, conform the McKinley model [15]. It can comprise a financial incentive directed at a person's income or directed at further investment in quality improvement. The financial incentive can be either positive or negative. Target payments, paying the practice of professionals only if they provide a minimum level of care, are considered a form of P4Q. Even if these targets imply a volume or quantity (e.g. the % of patients vaccinated), these indicators are still based on evidence based guidelines, intended to maximize quality of care, making it distinct from an uncontrolled fee for service approach not based on evidence based standards.
- Comparison is standard practice, i.e. without the presence of the above described intervention.
- At least one objective process or (intermediate) outcome measure must be reported on clinical effectiveness of care, interpersonal aspects of care, patient safety, access and equity of care, continuity and/or coordination of care or cost-effectiveness of care.

The following exclusion criteria were applied:

- Studies situated in mental health or behavioural health are excluded.
- Financial incentives aimed at patients do not comply with the P4Q concept. This is a whole other array of research, mainly within the field of prevention and changing lifestyle behaviour [15, 16].
- Papers only treating implicit financial incentives, which might influence quality of care, but are not specifically intended as such to promote explicitly quality in the first place, are excluded. Examples are studies on Fee For Service, capitation, salary use, etc. As noted by Kane et al (2004): "Any provider fee or reimbursement system within healthcare is by definition an economic incentive" [15].
- Subjective outcomes (qualitative research results, patient satisfaction, etc.) are not included, unless they were measured using standardized validated instruments.
- Papers only focusing on improving cost containment or productivity using financial incentives, without a primary quality objective, do not pass the PICOD requirements and are therefore excluded (see Lewandowski et al (2006) as an example [17]).

Appendix 3 and 4 list the reasons for excluding publications during full text review.

2.1.1.3 *Quality appraisal*

Quality appraisal of systematic reviews was based on the criteria of the Dutch Cochrane Center (Form Va).[18] Validity rating by use of this Cochrane tool was based on 7 items (namely research question, search strategy, relevance selection, quality appraisal included studies, data extraction, study description, heterogeneity and pooling). If more than three items were assessed as unclear or insufficient, a publication was excluded. However, for the majority of excluded reviews the reason was the narrative nature of the review or the lack of specification of methods used (Appendix 5).

Quality appraisal of primary studies was performed using a specifically constructed tool, based on a combination of existing tools [19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30]. The tool consists of ten generic items (namely research question, patient population and setting, intervention, comparison, outcome, design, sample size, statistics, generalisability, confounders addressed) and four design specific items (randomization, blinding, clustering effect, number of data points) (Appendix 6). For each primary study, a score has been given to each of the 14 items, ranging from -1 to +1. A score of +1 on a certain item corresponds with 'a good result', a score of -1 corresponds with 'a bad result', and a score of 0 corresponds with 'this item has not been discussed in the article'. To pass the quality appraisal, an article has to gain a score of minimum 8 out of 14. An article can be excluded on the basis of an 'overrule' argument when a certain criteria was not met, hence the overall validity has to be questioned (for example no significance testing, too small sample size etc…).

Quality appraisal of the studies reporting cost-effectiveness and modelling effects of P4Q programmes, was also performed by using two specifically constructed tools (See appendix 7). The quality appraisal tool for the modelling studies is based on the modelling guidelines from the KCE and on the ISPOR guidelines by Weinstein et al.[31, 32] It consists of the following items: model design (transparent and as simple as possible); assumptions and data input tested in extensive sensitivity analysis; original data set provided; sources used presented and described in detail; scenarios (for models that extrapolate to longer time periods); calibration (results should be logically consistent with real-life observations and data); face validity (the results of the model should be intuitively correct) and cross-validation (transparent enough to allow an explanation of the differences with other models for the same intervention). The tool used for the cost-effectiveness studies is based on the pharmaco-economic guidelines from the KCE [31]. It consists of the following items: literature review; perspective of the evaluation; target population; comparator; analytic technique; study design; calculation of costs; valuation of outcomes; data sources; incremental cost-effectiveness; time horizon; modelling; handling uncertainty; discount rate.

For each item one of the following scores have been assigned: +, +/-, -. A '+' score on given item corresponds with a good result, a '-' score corresponds with a bad result, a '+/-' score corresponds with an in between result. Given the limited amount of modelling and cost-effectiveness studies, the quality appraisal is only performed to inform the reader about the quality of the studies, but was not used to exclude any study. Overall the quality of the modelling and cost-effectiveness studies was acceptable.

Comparison between reviewers (DDS and PVH) identified seven non corresponding review publications out of 5480 potentially relevant reviews (Cohen's Kappa 99.9% interrater reliability) and 18 non corresponding primary publications out of 5517 potentially relevant primary publications (Cohen's Kappa 99.7% interrater reliability). Appendix 8 and 9 list the citations that were included in full text analysis.

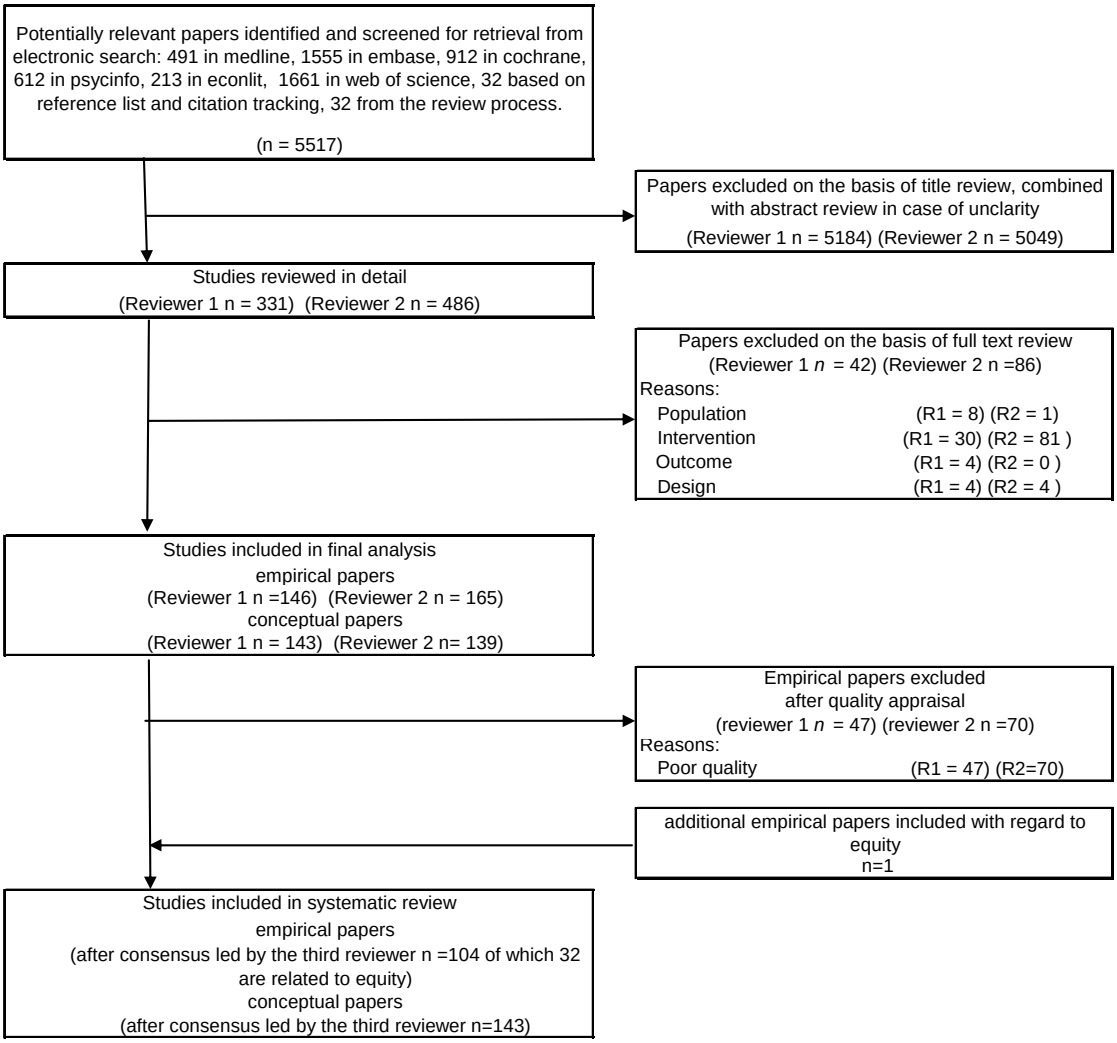
As a result 103 primary P4Q evaluation articles, 5 modelling articles and 3 cost-effectiveness articles, resulting from the primary publication search (phase 2) and from the review search (phase 1), are included in this review.

2.1.1.4 Data Extraction and construction of evidence tables

Subsequently 2 reviewers (DDS and PVH) performed the data-extraction process. The following data were extracted and summarized in evidence tables: citation, country, primary vs. hospital care, health system characteristics, payer characteristics, provider characteristics, patient characteristics, quality goals and targets, P4Q incentives, implementing and communicating the programme, quality measurement, study design, sampling/response/drop out, comparison, analysis, effectiveness evaluation, safety evaluation, access and equity evaluation, cost effectiveness evaluation, continuity and integration evaluation, co interventions, relationship results with regard to health system, payer, provider and patient.

The results of this process are described extensively in chapter 3 (conceptual grounding of P4Q) and chapter 4 (the evidence base of P4Q). Meta analysis of these results is not performed, because of a high level of clinical heterogeneity (differences in setting, population, intervention, and outcome measures used).

Figure 1: Flow chart for the identification of relevant, high quality primary conceptual and empirical papers



2.1.2 Selection of and data-extraction from articles related to equity

To describe the equity dimension of P4Q models a selection of publications from the main database was selected from the above evidence table. Afterwards two reviewers (SW and PB) performed an independent full text analysis. Publications labelled as "doubtful relevance concerning the impact on equity" by one of the two reviewers, were discussed between the reviewers until consensus was reached. In addition, all references of relevant equity articles were once again screened for additional material and one publication was added.

In a second phase the researchers (SW and PB) assessed how equity is conceptualized in the studies (to what extent did the selected papers address different domains of equity and did they conceptualize equity as vertical or horizontal or both?). To do this systematically, the researchers developed a grid which includes all equity-related concepts identified as essential (see appendix 10).

2.1.3 Revision of the conceptual framework into an evidence based state of the art framework

After the results were available from the conceptual framework construction and from the systematic review of the P4Q evidence base (equity dimension as well as all other quality dimension), both were integrated by relating the evidence base to the theoretical assumptions. The initial framework was assumed to be a set of hypotheses to be confirmed or disconfirmed by empirical findings.

When confronting theory with evidence, it is important to consider both the internal validity as the external validity of the findings. The internal validity is guarded by the previous 'hard' approach using a systematic review process. However, as suggested by Leontien et al (2008), this is combined with a Realist Evaluation Approach (see publications by Pawson & Tilley) to address external validity [33]. This approach addresses the question of what works for whom in what circumstances? Measures are expected to vary in their impact depending on the conditions in which they are introduced. Realistic evaluation assumes that mechanisms supporting an intervention are activated by conditions and can explain observed patterns in outcomes. These configurations are also known as context mechanism regularities which can be explored to uncover the 'black box' behind an intervention as P4Q.

Both underlying mechanisms (more practically than fundamentally behaviourally approached) and contextual factors are identified in the conceptual framework. During the revision process these were specifically related to evidence on outcomes. Missing links are also identified. The results can be consulted in Chapter 5.

2.1.4 Consultation of international experts as P4Q involved country representatives

For the international comparison study (chapter 6) four countries are taken into account, namely the USA, the UK, the Netherlands and Australia. 6 experts were chosen, based on their expertise with P4Q. For the USA and the UK, 2 key experts per country were invited to participate. For the Netherlands and Australia, two countries which are still in a starting phase of implementing P4Q, with only a few P4Q schemes operational, only 1 key expert per country was invited to participate.

Prof. Rosenthal and Prof. Damberg acted as US experts. The last 10 years Prof. Rosenthal has been working as a health care economist at the Harvard school of public health. Most of her work is related to provider-payer incentives. Her research focuses primarily on Pay For Quality and other payment mechanisms like capitation and FFS.

Dr. Damberg is a health services researcher, who works at a non profit US think-thank, named RAND. In the past she has worked for both the federal government doing health policy work, as well as for the private industry. Her main areas of interest are measuring quality and performance of the health system, patient safety, consumer-directed health plans, and assessing the impact of health care reform initiatives on quality and costs. Since 1992 most of her research has focused on Pay For Quality.

In the UK Prof. Roland and Dr. Millet acted as key experts. Prof. Roland is a family practitioner for over 30 years. He currently works as a professor of Health Services Research at the University of Cambridge. He has been involved in the development and the evaluation of P4Q in the UK. His areas of expertise include the development of methods for measuring quality of care, and evaluating interventions to improve care.

Dr. Millet is a consultant in public health at the Imperial College and completed a PhD on the impact of the Quality and Outcomes Framework in the GP contract on variations in diabetes management between ethnic groups. Currently, he is investigating the health care in the United States as a Harkness Fellow in Health Care Policy and Practice. His main current research interests are situated in the Health care quality improvement with a special attention on inequities in care.

Prof. Braspenning was invited as key expert for the Netherlands. She works for the Scientific Institute for quality of health care (IQ healthcare) situated within the UMC Radboud. The research unit is specialized in research related to quality and quality improvement in health care. Prof. Braspenning's research focuses on how to measure quality of care.

Prof. Duckett acted as key expert for Australia. He is an economist who has been an academic at La Trobe University for several years. The last three years he has worked for the Queensland government at the Centre of Health Care Improvement. Currently, he is working as CEO at Alberta health services in Canada.

With each key expert a 1 hour semi-structured interview was performed by two interviewers (PVH, DDS, or RR). The interview included the following 10 main questions:

1. Is there any form of pay for quality present in the health care system of your country, as you are aware of?
2. How did 'pay for quality' arise in your country?
3. What are/ were necessary cornerstones for the implementation of pay for quality, in the culture of your health care system?
4. How is 'pay for quality' developed and implemented in your country?
5. What are the reported or likely effects of 'pay for quality' in your country?
6. What is your opinion about P4Q programmes?
7. Which specific health system characteristics have largely influenced or will influence P4Q design, implementation and effects in your country?
8. What does the future hold for 'pay for quality' in your country?
9. What is your view on the current P4Q research status and its future evolution?
10. Which key recommendations do you formulate as an advice to a country at the first initial stage of considering the implementation of P4Q?

A more comprehensive description of the questions used for the interviews can be found in appendix 11

An overview of the interview content, was provided to each expert as a preparation document. Each interview was conducted by means of the questionnaire and the interviews were recorded. The interview contents are reported extensively in chapter 6. The draft version of this chapter was send to all experts which gave them the possibility to change or clarify things wherever they wanted to.

2.1.5 Conducting a systematic review of Belgian P4Q initiatives, and evaluating the feasibility of P4Q implementation, based on the comparison of current Belgian policy vs. findings from literature.

The output of the systematic review on P4Q (see chapter 3 and 4) nor an additional search strategy using Dutch and French entry terms within the predefined search string identified any Belgian study.

These limitations highlighted the need to collect the information as broadly as possible on a local scale by direct contact of all stakeholders involved. Because interviews of forty persons were planned as part of the chapter 8 data collection, these meetings were used as an opportunity to seek additional information on existing (P4)Q initiatives. They were asked to refer the research team to other persons involved in (P4)Q initiatives to provide additional information. These persons were contacted by phone and email, using a standardized template to collect (P4)Q initiative data. We write on purpose "(P4)Q" because, as will be shown in Chapter 7, the large majority of the current Belgian initiatives do focus on quality improvement, however do not apply P4Q. Nevertheless, a number of programmes did involve a kind of financial incentive. But this incentive wasn't in any programme *directly* related to the measured performance of participants with regard to predefined quality targets. In Chapter 7 we provide an overview of all Belgian quality initiatives that were mentioned by the experts. These programmes provide possibilities to develop pay for quality initiatives, as will be further analyzed in section 7.2.

The feasibility study makes use of the empirically revised conceptual framework (Chapter 5) to analyze strengths and weaknesses in current quality circle components to define P4Q threats and opportunities. The level of correspondence for both the quality initiative independent approach and the quality initiative dependent approach with the set of 'to do's' is used as the central parameter. The feasibility of modifications is assessed (see Appendix 12).

2.1.6 Consultation of Belgian stakeholders as involved country representatives

The assessment of the difficulties and possibilities to introduce Pay for quality in Belgium was performed following different steps.

2.1.6.1 Selection of stakeholders

In a first step stakeholders were identified from national health care organizations, institutions or organisational bodies that represented a potential interest in pay for quality systems. The long list is presented in appendix 13.

For each stakeholder, we identified mother tongue, the institution where the person is working and the stakeholder subpopulation the stakeholder belongs to. We aimed for 40 persons to be interviewed.

In a second step, the list of stakeholders was submitted to different scorers (all academic experts), i.e. persons who know many people in the Belgian healthcare system. They were asked to evaluate whether the different stakeholders on the long list were persons who were sufficiently aware of and interested in the pay for quality topic.

For each name, 4 options were possible (corresponding with a point):

- Yes (1 point)
- No (minus 1 point)
- I do not know the person (no point)
- I do not know whether it is worthwhile to interview this person (no point).

All values were computed resulting into a classification of stakeholders. The scorers had the opportunity to comment and add another name. The latter was considered important in the case a stakeholder did not accept the interview or in case it was necessary to complete the list of a particular subgroup.

The persons that acted as scorers to the long list were:

For French-speaking stakeholders:

- Geneviève Bruwier (ULg)
- Michel Roland (ULB)
- Pierre Gillet (ULg)
- Marc Vanmeerbeek (ULg)

For the Dutch-speaking stakeholders

- Jan Heyrman (KUL)
- Lieven Annemans (VUB, UGent)
- Walter Sermeus (KUL)
- Roy Remmen (UAntwerpen)

Based on the aggregated scores given by the scorers a ranking was made for every respective category of stakeholders. A total and maximum number of about 40 stakeholders were finally selected, representing a balance between Dutch and French speaking persons. The selection was made in close cooperation with the KCE.

2.1.6.2 Design of the questionnaire

A questionnaire was developed in parallel with the selection of the stakeholders. This questionnaire:

- Was based on state of the art literature with regard to pay for quality.
- Included open-ended questions
- Was reviewed by experts from the USA, the UK, The Netherlands and France.

The questionnaire (appendix 14) was developed following the usual approach of qualitative methodology. It started with an explanation for agreement on basic terminology about pay for quality concepts. It comprised introductory questions to facilitate the communication and allow the free expression of personal opinions; more narrow questions to cover the whole topic; closing questions so that the interviewee can develop some previous answers or some new ideas about the topic. Questions were open-ended to allow the stakeholder's free expression. A list of detailed questions was provided one week on beforehand.

We proposed each interviewee to imagine a P4Q programme being developed in Belgium. Without further details, we asked him/her to state his/her opinion on advantages, disadvantages, pitfalls and expected resistances to such a programme.

Then, we discussed the above more in detail following all elements of the theoretical framework (see Chapter 3).

Final questions were related to the budget and open to more economic view of the impact of such a programme. The last question asked about the stakeholder's knowledge of other experts in the field or persons having developed such a programme. This allowed the team to be sure to cover the whole expertise in the domain in Belgium and eventually abroad. It goes without saying that the interviewee was allowed to complete or put the emphasis on one point of interest by the usual prompt: "Do you want to add something, do you think we have forgotten one important point?"

2.1.6.3 *Protocol of the interviews*

As a protocol to this questionnaire the stakeholders first received a letter from the KCE to present the study and to ask if they knew some P4Q studies already developed in Belgium. Then, they were contacted by phone to arrange an appointment.

The interviews were conducted by two academic teams, including three Flemish-speaking researchers (KUL: JH, LB and CVdB) and two French-speaking researchers (ULg: MV and CD).

The Flemish-speaking interviewers were assisted by a student who took notes of what was said whilst the interview was tape-recorded. The French-speaking interviewers were usually alone but recorded the whole interview. Interviews lasted for 50 minutes to more than 2 hours.

2.1.6.4 *Analysis of the content of the interviews*

To guarantee as much as possible the objectivity and the reproducibility of the conclusions within the chosen qualitative approach, the following steps where used to come from interview to integrated conclusions:

- Basically the interview structure followed the conceptual framework, so that the answers could maximally fit into and contribute to the framework reflection.
- All interviews were audio taped or written down.
- Answers were translated from French or Dutch into English and transformed to more condensed "long list of quotes" by each of both interviewers, and put to mutual agreement. All quotes where marked with the name of the interviewee and his/her adherence to the stakeholders' typology (government, academics, unions, primary care, hospitals, etc…) and language group. The total list of quotes was made available to each of the two academic team members.
- As a first analysis, the quotes were assigned as close as possible to the theoretical framework and its different subheadings.
- The exercises of omitting unproductive quotes, eliminating overlap and repetition, and condensing different quotes to one statement was done by half of the academic team, under control of the other half. The decisions about the final "reduced list of quotes" were taken during a common open meeting of the whole team.
- A final reduced list of "interesting quotes, including their origin" was constructed and is still available at the courtesy of the research team. An agreement was made with the stakeholders not to publish names and affiliation. From this "reduced list of quotes" half of the academic team made a readable and prioritized text version of the final conclusions emerging from these quotes, the other half controlled this exercise. Statements of stakeholders were, if appropriate, put in contrast or in support with the messages from the international literature study and the conceptual framework. Only these condensed messages are reported, referring to categories as "some / many / most of / all stakeholders", "primary care versus hospital care stakeholders", "union versus government stakeholders" and "French region versus Flemish region stakeholders".
- Two meetings were necessary to extract main messages and to bring them together with the general conclusions of the whole project in chapter 9.

3 A THEORETICAL FRAMEWORK FOR P4Q

The theory behind P4Q can be viewed from different perspectives: the definition of "quality", the definition of "incentive", the relationship between the payer and the health care provider, the consequences of incentives on the provider's behaviour, and finally the way in which the health care context influences programmes and their results.

We identified in the literature several comprehensive conceptual frameworks that attempt to take into consideration most of the above. [15, 34-43] Yet, in our opinion, none of these provides a full comprehensive picture of all P4Q elements.

Interestingly, a general appeal is made in the recent literature to refocus P4Q from effectiveness towards the inherent combination with other quality domains such as coordination of care, and reducing the fragmentation of care [35], but also patient safety, equity and cost effectiveness [44, 45, 46]. While this allows encompassing more aspects of care, it may obviously complicate the concept and its implementation.

This chapter aims at finding an answer to the following questions:

How should P4Q be modelled, taking into account all theoretically relevant factors? What may be the practical implications that can be derived from theory to support P4Q design and implementation?

The emphasis thereby is more on implementation rather than on the pure theoretical grounding. Note, however, that the design and the question with regard to domain and implementation is still approached in conditional terms at this point ("what may be …"); later in this report, based on observed evidence from the literature, more firm recommendations will be given regarding do's and don'ts in P4Q design and implementation.

The P4Q conceptual framework that we present at the end of this chapter forms the basis for evaluating existing P4Q applications and provides a first set of information based on which programmes could be addressed within the Belgian context. Many of the conceptual findings which are presented below are also founded on psychological and economical theories, as applicable to healthcare [47, 48, 49, 50, 51]. In addition, most elements are also identified in other sectors which resemble healthcare on key characteristics, such as teaching and legal professions [52, 53]. Similar to healthcare these professions are part of public service with mainly independent professional actors and the presence of asymmetrical information within the client/patient, provider and payer relationships. The findings of non healthcare sciences and sectors are integrated, while recognizing the unique nature of the healthcare environment in its own respect.

The methods of this literature review supporting this conceptual analysis were described in Chapter 2.

3.1 P4Q CONCEPTS

3.1.1 Quality

3.1.1.1 Definition of Quality

The basic principle of a P4Q programme is to offer explicit financial incentives health care providers in order to achieve predefined *quality* targets [54].

If the ultimate goal is to achieve predefined quality targets, then the obvious question is to define quality. Since the '90s it became increasingly possible both to define high-quality care and to provide methods that could be used to measure some aspects of the quality of care [55].

As described in our introduction, quality of health services has been defined as 'The degree to which health services for individuals and populations increase the likelihood of desired health outcomes and are consistent with current professional knowledge'.

A former KCE report by Vlayen et al. (2006) proposed the following dimensions of quality of care [56].

- **Safety**: avoiding injuries to patients from the care intended to help them;
- **Clinical effectiveness**: the professionals giving care should be competent, provide services based on scientific knowledge to all who could benefit and refrain from providing services to those not likely to benefit;
- **Patient centeredness**: providing care that is respectful of and responsive to individual patient preferences, needs, and values and ensuring that patient values guide major clinical decisions;
- **Timeliness**: avoiding waits and potentially harmful delays;
- **Equity of care**: services should be available to all people and care should not vary in quality because of personal characteristics such as gender, ethnicity, geographic location, and socioeconomic status (this is elaborated more below);
- **Efficiency of care**: the society should get value for money by avoiding waste, such as waste of equipment, supplies, ideas, and energy;
- **Continuity and integrativeness**: all contributions should be well integrated to optimise the delivery of care by the same healthcare provider throughout the course of care (when appropriate), with appropriate and timely referral and communication between providers.

This approach goes much broader than the strict "clinical outcomes" approach. Especially the focus on equity could be considered as an implicit criticism to the neo-classical way of thinking as if effectiveness and efficiency should be the major values of any health care system.

In addition to the above mentioned dimensions of quality of care, quality on a global level also involves reducing variability in care [55, 57].

P4Q involves an evolution of payment systems from "pay to do things" towards "pay to do things right". However this view could further evolve to "pay to do the right things" and even simply "pay to do right".

Equity has received special attention in this project. From the beginning the research group took a special interest in assessing the impact of a P4Q programme on equity. Indeed from a Belgian point of view this seemed a rational choice. Inequities in health concern systematic differences in health status between different groups (men versus women, age groups, socio-economic groups, …). Inequities in health are systematic (not distributed randomly), socially produced (social processes and not biological processes produce these variations in health) and considered to be unfair because they are generated and maintained by unjust social arrangements. [58] In this context the concept of "substantive equity" refers to the minimization of disparities in health among subgroups. [59]

When striving to reduce inequities in health, policies should be designed in ways most likely to move toward equalizing the health outcomes of disadvantaged social groups with the outcomes of their more advantaged counterparts [60].This means that the only strategy to narrow the health gap in an equitable way is to bring up the health level of those who are worse off. This concept has also been called levelling-up [61, 62].

Inequity in health care is one of the many determinants of inequities in health. [58] (In)equity in health care is a central point of attention of many health care systems and tackling this inequity has been an important objective in the development and reorganization of health services. [63]

There is a large amount of literature on how to conceptualize and measure equity in health care.

Despite a current lack of consensus, some common ground can be found in literature to divide equity in three domains: equal *access* to care for people in equal need, equal *treatment* for people in equal need, and equal treatment *outcomes* for people in equal need. While this is a simplification of the nature of equity, it is useful in delineating the various domains in which inequities may arise. [64].

An important step when assessing equity in ***access to care*** is to formulate an operationally useful definition of access. As indicated by Goddard et al, the precise formulation of 'access' is highly contingent on the context within which the analysis takes place. E.g. in the U.S. access is often considered to refer merely to whether or not the patient is insured. Hereby nuances such as the level of insurance or quality levels in care are secondary. In Europe, where most citizens are insured, access is formulated in a more detailed and broader way. [63]

According to Goddard et al (2001) 'access' refers -at the most general level- "to the ability to secure a specified range of services, at a specified level of quality, subject to a specified maximum level of personal inconvenience and costs, whilst in the possession of a specified level of information". [63]

The aspect of *availability* and having equal access *to a specified range of services* or *availability of equal services* for people with equal need is found in almost every definition of equal access. [63] [59] It refers to the fact that age, sex, income, … should not dictate that people with similar needs enter different doors (e.g. public versus private providers) or be treated differently in terms of the type or intensity of services provided. [59]

In this definition *quality of service* is also an intrinsic element of access. Poor quality in terms of the structure, the processes of care or the outcomes might compromise the access to care. For example poor quality of the care process might lead to patient dissatisfaction and result in low compliance. [63]

Concerning the aspect of *personal inconvenience and cost, and of information*, Goddard and Smith (2001) indicate that there might be considerable variations in the personal costs of using services (user fees, transportation costs, …) and in the awareness of the availability and efficacy of services (e.g. because of language or cultural differences). Although completely equalizing personal costs of access and distribution of information, is infeasible, there must be some point when differences in costs and information distribution become unacceptable. [63 , 65]

In the assessment of ***equity in treatment and treatment outcomes***, the interaction between patient and provider plays a major role: variations in treatment and treatment outcomes are considered to arise from this interaction which depends on the knowledge, skills, preferences, perceptions, attitudes, prejudices, … of both patient and health care provider [63]. Also the wider social determinants of health such as the social circumstances in which people live and work, might contribute to inequity in treatment and treatment outcomes. For example recovery rates after an operation in different social groups, can occur even when there was no inequity in the access or the treatment that has been provided. [58] For these reasons, analyzing equity in treatment and treatment outcome is complex, and not always feasible. [58]

Equity works on the central principle of equal access, treatment, … for people in equal ***need***. [58]. The 'taxonomy of need' identifies 4 domains of need. [64, 66] The first domain is 'normative need' which is need defined by an expert or professional according to his/her own standards e.g. a guideline (defining e.g. which group is at risk for lung cancer). The second need is the 'felt need': the need in which people identify what they want. Important is that felt needs may be limited or inflated by people's awareness and knowledge about what could be available, so, for example, people will not have a felt need for knowing their blood cholesterol level if they have never heard that such a thing is possible. 'Expressed need', the third domain, is felt need which has been turned into an expressed request or demand and can therefore be conceptualised as demand for care and – if the demand is fulfilled – care utilisation. Finally, the forth domain, 'comparative need', is defined by comparing the user rates of care of different groups of people e.g. screening rates for breast cancer. The group who uses less (in the example with the lowest screening rates) is then defined as being in need.

This approach simply compares and makes no judgments about the appropriateness or the adequacy of the use in the group with the highest rates, e.g. it is possible that the screening rates in the group with the highest rates, are still not at an adequate level or that there is an overscreening or overtreatment in certain groups. [64, 66]

3.1.1.2 Quality criteria and indicators

If the above dimensions of quality are generally accepted, then ideally, each of them should be made evaluable in terms of quality *criteria*. Thereby, any criterion selected, as well as the desired changes in this criterion, must be policy *relevant*. [67]

In practice, the above definition of quality can be further completed with Donabedian's distinction of *structure*, *process* and *outcome* quality. But these criteria are clearly not applied to the same extent.

Indeed, the quality of ***structure*** is often omitted [35]. Structural elements are intended to provide the infrastructure, staff and material resources (including time scheduling) to enable quality improvement processes. These are sometimes also referred to as Care Management Processes (CMPs). The availability of an incident reporting system in patient safety management, the availability of an integrated IT infrastructure with a decision support function, a reminder function, an automatic quality data extraction function, etc. are considered necessary conditions to support the use of specific quality improvement strategies (education, benchmarking, clinical pathway use, etc.) [68].

Moreover, a perceived shortcoming of many P4Q efforts has been the lack of focus on demonstrable benefit — including both health outcomes and spending — as opposed to process-of-care measures [69]. This is likely to be explained by the fact that providers may be more confident that they can control processes of care than outcomes [38]. Numerous external factors (e.g. patient lifestyle, patient compliance, and many non healthcare related factors) influence health outcome. It is therefore hard, and according to some authors unethical or unacceptable, to assign full responsibility for patient long term health outcomes to a provider or team of providers. However, the relationship between structure and process measures and long term patient outcome measures is scarcely grounded in scientific evidence [70, 71, 72, 73]. Therefore, overly relying on structure and process outcomes threatens the credibility of a P4Q system. It instils doubt that the programme will do what it is intended to do: to maintain or improve quality of care, which is mostly patient outcome and patient experience based (the global health of the patient).
This choice between structure/process indicators and outcomes indicators corresponds with the distinction between efforts versus result based assignment of responsibilities. As in most P4Q design decisions a combined approach is also possible, and used in most programmes.

As important as the choice of criteria is the final *number* of criteria applied in the P4Q programme. Rosenthal et al (2005) point out that too few criteria may lead to drawing the attention away of many aspects of medical practice that are not covered by the selected criteria, while too many criteria will lead to organisational complexity [74]. With this regard, some authors refer to the term "multitasking" [75, 76]. If the goal of the payer is multidimensional, but not all dimensions lend themselves to measurement, then rewarding performance based on available measures will distort the efforts away from the immeasurable objectives. This has also been described earlier by Conrad et al (2004)[35].

Once criteria have been selected, the desired value for each of the criteria reflects the specific *goals* (or *targets*) of the P4Q programme, the achievement of which can be measured by *quality indicators*.

Quality targets can be directed at improving current underuse, overuse or misuse of treatment. The definition of appropriate vaccination rates is an example of the first, the definition of appropriate lab testing and medical imaging use is often an example of the second. We will see that almost all current P4Q programmes focus on underuse.

It has been noted by Duckett et al. (2008) that some programmes may adversely encourage overutilization or unnecessary diagnostic testing [67]. Hence, before setting criteria related to one or more quality dimensions, it must be fully documented that these criteria are in line with the current knowledge base regarding the optimisation of these dimensions.

Regardless of the above issue, targets must be SMART (specific, measurable, achievable, relevant, and timely) [67, 77].

Rather than using explicitly the SMART terminology, Dudley and Rosenthal (2006) state that following questions should be asked when deciding on quality targets and hence indicators. [78]

1. Does the indicator measure care that is a priority for quality improvement?
2. Does the indicator reflect technical competency or patient experiences with care?
3. Is the indicator actionable?
4. Is there a valid source for the data needed to calculate the indicator? What is the cost of acquisition and validation of those data [79]?
5. Is the indicator accepted by the medical community?

With regard to "achievable" or "actionable", the challenge is to find an optimal achievement level thereby considering that there must be sufficient room for improvement on the one hand, but that the target must be realistic and achievable on the other hand [38, 67]. The "Measurable" part of the SMART concept is discussed in the next paragraph.

3.1.1.3 Quality measurement

The desired behaviour must also be *measurable*, which involves that valid and comprehensive management information systems to track performance against the goals must be available, and must also be easy to apply by both the payer and the provider (see payer characteristics, page 25) [67]. With this regard, Conrad and Christianson (2004) point out that there may be strong differences between perceived and actual accuracy of the underlying database for the incentive [35].

Moreover, correct measurement also involves that the provider's case-mix is taken into account, using risk adjustment for outcome measures (see provider and patient characteristics) [67, 80, 81, 82, 83]. Some P4Q programmes make also use of a procedure called 'exception reporting' [84]. This procedure enables providers to exclude individual patients from the calculations for specific targets, because there was a valid reason for not reaching the target in that individual patient, which was not quality of healthcare related. Exception reporting is mostly restricted to the use of a predefined set of exclusion criteria as an acceptable rationale. These can include the level of patient compliance or willingness to cooperate in treatment, the maximal degree of treatment which can be applied (e.g. when a maximal dose of a drug has been prescribed, without further alternatives, and still the quality target is not met) and the influence of co morbidities on the appropriate care in an individual patient (the typical example being not to provide an eye examination to a diabetes patient who is blind). Procedures like exception reporting are used as a safeguard to protect professional autonomy and therapeutic freedom [85, 86].

The discussion about exception reporting points our attention to the fact that incentive schemes always risk generating unintended consequences [87, 88]. According to Dudley and Rosenthal (2006) there are three important negative effects to look for: patient selection, diversion of attention away from other important aspects of care, and widening gaps in performance among providers. [78]

The principal type of unintended consequence is patient selection or more generally "gaming", where participants find ways to maximize the measured results without actually accomplishing the desired objectives [75]. Diversion of attention has been earlier described as the risk associated with multitasking (see above for quality targets) [89]. Widening gaps in performance may be the result of differences in motivation and response to the programme (which are in turn influenced by the type of incentive, communication etc. - see also further) and should therefore be systematically monitored. At present there is no evidence supporting these effects [90].

3.1.2 Incentives

3.1.2.1 Introduction

Incentives are part of everyday life in all its aspects. People do things for various reasons both of a non financial and a financial nature. A drive to act, behave, change, etc. in a professional context is related to personal work satisfaction, to recognition and status and to receiving financial and non financial resources and opportunities in exchange for professional healthcare delivery (the effort) and performance (the results of the effort)[a]. There is a wealth of evidence available that financial incentives have an impact on professional behaviour, both in healthcare as in other sectors [92]. The, often unconscious and not deliberate, effect of financial incentives can impact behaviour both in a positive as in a negative way [93].

Whereas some financial incentives have always been used in the history of healthcare or have been introduced to maintain or improve target performance (salary, fee for service, capitation,…) [94], a P4Q programme does the same but is explicitly aimed at quality of care. New innovative forms of payment approaches are emerging rapidly [69].

The actual incentive is often considered as the core of the programme, since the overall aim of a P4Q programme could be considered as a better alignment of the incentives of providers with the health system goals [75]. To be clear, in this context, incentives are considered to be *explicit* (and not broad categories of "incentives" such as fee for service, salaries or capitation) [35]. In this paragraph, we focus on the possible structure and magnitude of this explicit incentive. [38]

3.1.2.2 Incentive structure

The following different possible incentive structures can be considered: [38, 54, 75]

- Bonuses, rewarding providers with additional payments for achieving the stipulated targets. This has been a popular approach in many programmes, likely because of its relative attractiveness towards providers. Yet, while such bonuses may be effective, they are not necessarily cost-effective. [75] Related to bonuses is the so called "pay for activities", whereby the merely presence of required activities (without a result commitment) is rewarded. Also related to bonuses are the so-called "shared savings contracts", whereby the goal of the programme is in the first place to save money and whereby the provider is then entitled to receive a share of the realized savings.
- Performance based fee-schedule; the only difference with bonuses, is that the payment is ongoing rather than one-time or periodic. Mark-ups to a usual fee (also called enhanced fee for service) are an example of this. [35, 38]
- Performance based withholds; since it has been shown that individuals place more value on losses than on equivalent gains, withholds or financial penalties may be more effective than bonuses. [78] Also, as Averill et al. (2006) state, paying bonuses comes down to rewarding what in fact should be standard of care. [78, 95-97]
- The empirical evidence with regard to the effect of penalties is however limited, as will be shown in Chapter 4. [78]

[a] Examples of the Global Health Workforce Alliance are provided in their guidelines report.[91]

- Regular payment increases linked to performance. This is an approach whereby some or all funding increases are placed at risk, so that future increases will only be realized depending on the performance level.
- Quality grants/ Financial awards/ Performance funds. In this approach, health providers receive funding to implement a quality related programme (either a project, or the infrastructure and means to enable the realisation of such a project), which is sometimes, but not necessarily done in a competitive way. In "non-competitive" markets, it may be a tool to introduce some level of competition and market forces. Quality infrastructure grants also fall within this category, as far as these are rewards that are directly related to the achievement of quality targets. [35]

Each of these incentive structures has strengths and weaknesses. Custers et al. (2008) designed a rather practical algorithm that could be of help in deciding which incentive type to use in which circumstances. [75]

3.1.2.3 Other incentive characteristics

Further questions however rise:

What should be the size of the incentive? The size of the incentive is obviously considered as a key factor influencing the effectiveness of the P4Q programme. Yet, Rosenthal and Dudley (2007) observe that some pay-for-performance schemes have paid as little as $2 per patient and had an impact, while others offering bonuses of up to $10 000 to a practice had no effect.[78] It is important to note that P4Q complements rather than substitutes existing payment systems. Together with other, more volume of care related incentive systems like Fee For Service and Capitation, it aligns activities in the healthcare system with its basic goals and targets. P4Q is therefore a component of a payment system that serves the quality related goals, next to other components. To provide sufficient security and a reliable income to providers, as in other sectors, the majority of the income or revenues are linked to healthcare delivery as such, and only a minority percentage is distributed according to P4Q principles. In general, an incentive size of 5% of income or revenues is considered to be a minimal amount to induce an effect on provider behaviour. Some authors indicate that 10% would be a more appropriate number above the minimum. This size corresponds fairly well with bonus sizes used in other sectors (e.g. 5 to 10% based on yearly profit or revenue, or a one month's wage as end of year additional income).

1. What will be the formulation of the incentive scheme?
2. As said before, both penalties/ withholds or bonuses can be applied.
3. In general, if bonuses are applied there are four possibilities [78]:

- rewarding only those providers that meet or exceed a single threshold of performance;
- differentially rewarding providers for achievements along a continuum of performance thresholds (those who achieve a higher threshold earn more than the others but the others receive some payment as well);
- rewarding providers that meet or exceed a single threshold of performance combined with and incentive rewarding of those that improve, regardless of whether they meet the threshold;
- rewarding providers in a continuous manner in proportion to their achievement (i.e. solely in function of the improvement).

A problem with the first approach (working with a threshold) is that high-quality providers may receive bonuses without making any improvements, while low-quality providers may find the single threshold too difficult to meet and opt not to engage, as has been shown by Rosenthal et al. (2005) [74] and other authors [98].

What is the frequency of payments (e.g. yearly, quarterly, monthly)? The answer to this question has some practical components and a theoretical component. The frequency is limited by the data processing capacities of the quality measurement system. This often causes a time lag between measurement, feedback and payment.

On theoretical grounds it can be stated that behaviour and consequences should be related in time as closely as possible. However, a continuous preoccupation with the incentive effects would also divert attention of providing care on a regular daily basis without an excessive additional programme related workload. The costs of a continuous reporting and payment system are likely to exceed the benefits due to time related gain in quality improvement. Therefore, like in other sectors, P4Q programmes often provide payment only on one or a few fixed time points such as at the end of the year. This also gives providers a sufficient amount of time to invest in intermediate quality improvement in between quality measurements. A related question is whether the incentive is stable and long enough [67]. Incentives which are given on predefined time points with a sufficient duration over time would offer the providers the assurance that their efforts in achieving the indicators, will be warranted [67].

What is the weight allocated to the different dimensions of quality and to the different criteria within each dimension [35]? Some P4Q programmes, like the Quality and Outcome Framework (QOF) in the UK, attach payment weights to specific quality targets as a function of the estimated related workload or time investment necessary to affect the targets. The weights can also express other criteria like the relative importance of specific targets in terms of public health. The QOF example also uses another weighting approach, based on the distinction of a clinical and organizational domain. The clinical domain contains process and intermediate outcome targets. The organizational domain contains a set of structural targets. An extreme form is the expectation that a provider should perform perfectly on all included quality targets. This is translated in some programmes into an all-or-none approach, meaning that the P4Q incentive is only provided if all targets have been met [99, 100].

Is the reward relative or absolute? In case of an absolute reward, anyone who performs well obtains this reward no matter how the others perform. If the reward is relative, providers compete against one another to obtain a bigger share of the available money. This is also called the "tournament approach". The tournament approach has some theoretical advantages: it is cheaper or at least the expenses are more under control, and there is also a continuous incentive (to outperform the others).[78] Disadvantages may include the uncertainty about what can be achieved, so that providers may judge investments in quality improvement to be unacceptably risky, and decide not to engage in the programme.[78] The choice between both options is related to what are considered to be the defining aspects of the level of performance to be achieved. It is easy to state that the percentage of HbA1c<7.4% should be a target in diabetes care, based on latest scientific evidence. But it is much more difficult to state for which percentage of all diabetes patients cared for by one provider or a team this target should be met. Many authors consider a 100% performance rate unrealistically high and unachievable. Based on evidence about what's achievable one can define a lower percentage as a threshold (e.g. 80 percent), based epidemiological data or on a consensus approach, and which can be gradually revised. A relative reward system makes use of short term comparison and the performance of the other providers becomes the point of reference. Because of the higher variability this is likely to induce more short term change in provider behaviour. However, too much pressure and uncertainty may also lead to a giving up response. People in general prefer to have a sense of control or security.

Tournament and ranking systems are often felt as a threat to these needs. In addition, some consider it to be unfair systems because the incentive gained will depend on other providers' achievement rate on the different indicators, which is beyond the control and responsibility of each individual provider.

Is the reward clear and simple enough to understand [67]? It is suggested by Conrad and Christianson that this factor may influence the effectiveness of a programme [35]. However, next to transparency and straightforwardness, which will strengthen the relationship between the incentive and behaviour, trying to keep all things simple also holds specific risks. The current payment system of healthcare is complex, because it provides resources and incentives in a very complex practice environment with influences of market, payer, provider and patient characteristics (see below for a detailed description of these characteristics).

Taking into account this complexity and the multiple goals which should be incentivized simultaneously, leads to the construction of more elaborate P4Q schemes. Some authors suggest solving this problem by separating the complex calculations, formula and adjustment procedures from the way in which quality performance feedback is provided on the key targets driving the incentive. According to what needs improvement other presentation methods can be used. Whereas the presentation of one composite measure, summarizing the performance on any preferred level, can be useful for general comparisons, the presentation of target specific data informs the provider on which targets to centre future quality improvement initiatives and often also which means can be used to accomplish them [101, 102, 103, 104, 105, 106].

3.2 THE P4Q CONTEXT

Hutchison et al. (in Frolich et al. 2004) point to the importance of considering the *context* in which financial incentives are designed or implemented to understand their potential effects [38]. Indeed, each provider's efforts in responding to incentives are mediated by characteristics of the local market, the medical organisation (if any) in which he or she practices, individual provider characteristics and on patient characteristics.

The following contextual aspects will be discussed in this section: the health care system, the payer-provider relationship (with theoretical grounding), their respective characteristics, and the patient's characteristics.

3.2.1 The health care system

Obviously, it is essential that incentive models are congruent with the *values* of the health care system. For instance, in Ontario, Custers et al. (2008) adopted 4 principles that were congruent with the health care system values [75]:

1. Be fiscally prudent (no new money);
2. Be simple to administer (no additional administrative concerns);
3. Support a culture of continuous improvement (no one-shot action);
4. Improve equity in and access to quality of health services.

General aspects of the system include the type of system (insurance or NHS; level of regionalisation), the public/private mix (% insured), the dominant payment system (fee for service, salary, capitation, etc.) and the level of therapeutic freedom among providers.

According to Conrad and Christianson (2004), these market and environmental conditions will, among other things, drive investment in structural quality (medical equipment, human capital) and could therefore be considered as exogenous determinants of incentive programmes [35]. It can be argued that the market and environmental characteristics will also drive process and outcomes related aspects of quality, and hence the success of P4Q programmes.

For instance, the extent of competition between providers may affect their response to incentives: a provider in a monopoly situation could maximise profits without improving quality [38]. This level of competition is in its turn related to other healthcare system characteristics, such as the degree of patients' free choice to consult with providers of their own choosing. Although P4Q is not directly related to the patient's choice of a provider, the number of providers one patient consults will influence responsibilities' allocation and the level of care continuity to support high quality healthcare.

Interestingly, the appropriate and timely referring of patients is in several P4Q programmes a quality target on its own. Furthermore, the level of decision making in P4Q policy will influence the uniformity, transparency, awareness and general acceptance of a P4Q programme design. The lower the level, the more risk for fragmentation and for variously competing approaches. This reduces programme awareness and acceptance of providers. It also reduces the impact size of the incentive (the effect of one of many simultaneous programmes, also known as a dilution effect versus the effect of one national programme).

Obviously P4Q design on a national level can, as mentioned before, be combined with involvement and local priority setting on other levels.

Salary as a general payment system is considered as a volume neutral payment system, and will therefore likely not to have any positive or negative influence on P4Q programmes targeting underuse or overuse. Salary provides mainly sufficient security and a reliable income, but no care quality or quantity incentive. P4Q may add an additional quality stimulus when combined with a dominant salary payment system. Fee For Service as an activity volume driver is likely to combine well with P4Q targets aimed at underuse, but also induces a risk to reduce or even to eliminate the effects of P4Q targets aimed at overuse. Capitation as a patient volume driver, but also per patient cost containment driver, is likely to combine well with P4Q targets aimed at overuse, but also induces a risk to reduce or even to eliminate the effects of P4Q targets aimed at underuse.

3.2.2 The payer-provider relationship

3.2.2.1 Principal-agent theory

According to Nahra et al.(2006), the conceptual foundation of providing an incentive to achieve a desired result from the receiver of the incentive can be found within the context of a principal-agent framework [107]. The agency theory describes the relationship between a principal (for instance the insurer or a national health service) and one or more agents (physicians, hospitals,...).

Under this theory, a principal must hire agent(s) to carry out an objective that the principal cannot carry out alone. To align the goals of the agent with those of the principal, rather than contracting with the agent solely for the provision of *effort*, the principal may contract with the agent, at least partially, on a measure of *outcome*. [108], [107]. Such part of a contract refers to pay for quality. Hence, principal-agent theory addresses relationships in which 1° both parties have different abilities (and it is therefore desirable that the first party delegates responsibility for performing a function to the second), 2° there is asymmetric information (for instance the insurer cannot monitor all the actions that physicians take), and 3° the parties have –to some extent- different goals (or other priorities within the diverse set of quality domains).[38]

In the relationship between the principal/payer and the agent/provider the latter can be both a potential ally and a potential source of resistance to P4Q. Regarding resistance, providers may have particular concerns about the quality of the data and the validity of measures created from the data.

They can also be very sceptical about data produced by outside stakeholders such as government agencies or employer coalitions. Finally, they are also concerned about their ability to influence many outcomes measures of quality because of the substantial role played by patient actions and preferences (see the discussion above regarding the control of providers over processes of care).[78]

In order to avoid the above to some extent, one may implement a voluntary programme wherein not all providers need be ready and willing to participate. However, voluntary programmes will be likely to attract those providers who expect to perform well — usually those that are already performing well — while the poor performers remain on the sideline.[78]

3.2.2.2 The payer

Several organizational and market mechanisms influence the way a payer can and will implement P4Q [109]. For instance, if there is already an existing policy of clinical guidelines endorsed by the payer, it will be easier to build further on this policy and add a P4Q dimension to it. Also, if a variable patient contribution in function of provider and/or technology performance is already in place, then again, it will be more acceptable to introduce P4Q [35]. In general the inclusion of elements of existing quality incentive schemes obviously will influence the success of a new programme [38].

On the structural side, the availability of management information systems is crucial for the success of P4Q.

Also, in a context of multiple payers, the question about coordinated action from different payers can be raised. On the one hand, a payer may be reluctant to work alone if the fruits from the programme are also of benefit for other payers (because the providers treat patients related to different payers); in other words, one wants to avoid a free-rider situation in which certain payers profit from the efforts of other payers. On the other hand, if payers compete with each other, it may be more interesting to obtain a competitive advantage through P4Q.[78]

Although the vision on the health care system and the typology have already been discussed before (see health care system, page 24) these may also be considered as payer specific since it obviously possible that a payer has a different vision as compared to the overall health system, and operates in his specific way (e.g. a private insurer within a NHS).

3.2.2.3 The provider

The health (care) provider can be considered as the target audience of a P4Q programme. The provider can be an individual physician (GP or specialist), a group of physicians, a hospital, a hospital department, a resting home, etc...

As said before, it is of importance that the programme is in line with the provider's culture. For instance, the emphasis of Custers et al. (2008) on continuous improvement, innovation and mutual learning (see above) was believed to fit well with the vision of the Canadian physicians [75].

In the following, we will discuss consecutively provider's motivation, the target unit, and other organisational aspects.

Motivation

How can providers be motivated to participate in a programme? Motivation of health professionals is often ignored in P4Q programmes [110, 111].

A rather simplistic view on this is that when explicit incentives are used to change behaviour, the motivating effect of money will channel the professionals to the policy defined goals.

In this view, each physician has a target income, and incentives that help to achieve that income will change behaviour. It also means that if the desired income has already been reached a P4Q programme will have less effect [38], or that providers whose performance has improved but does not reach the threshold, could become demotivated to make an effort [67].

More importantly, this view ignores the complex interplay of internal and external factors affecting the health professional's behaviour [75, 110]. As social beings and as agents for their patients, physicians are driven by important societal and professional norms and by altruism, in addition to net income.[35, 67]

Thus, the financial incentive might either enhance intrinsic motivation if it is viewed as being legitimating the internal or professional norms or reinforcing them; but it may as well diminish the strength of the intrinsic quality motivators [35]. In other words, an extrinsic motivation like the use of financial incentives can crowd out the intrinsic motivation by for example demotivating individual providers, or devaluating their altruistic motivation.

Motivation is possibly also related with the level of *trust* the physician has in the payer. Conrad et al. (2004) therefore cite trust as a key factor influencing the effectiveness of P4Q programmes [35].

It eventually comes down to "internalizing" the external regulation [110], i.e. to make the incentives instrumentally important for the personal goals. One could also argue that our systems should more externalize the intrinsic values of medicine, i.e. reward societal and professional norms.

Also, the role of medical leadership in supporting the P4Q programme (as one of the many roles that medical leadership fulfils) is described by Conrad et al. (2004) as potentially influencing motivation and therefore effectiveness of P4Q programmes [35].

Finally, the practicing physician's knowledge and understanding will contribute to the motivation to act in line with the goals [35, 38, 54].

The latter is immediately related with the level of *involvement* of the individual clinicians and their degree of autonomy. Here again, it could be argued that more involvement and more autonomy will increase motivation.

Note that Conrad et al. (2004) also refer to peers' knowledge of individual provider performance, as a variable potentially influencing effectiveness of P4Q programmes. Indeed, if peers are aware of the performance of an individual physician, this will definitely influence his/her behaviour [35].

According to Adams and Hicks (2001), the industry can have an important role in affecting physician professional behaviour [112]. The incentives given by pharmaceutical representatives to providers can effect providers prescribing and professional behaviour [113].

The role of the media in P4Q programmes is rather small. However in public reporting, where the quality of care provided by physicians or hospitals is made publicly, the media plays an important role. The availability of ranking lists or performance reports on the internet can influence physicians' behaviour [38].

The target unit

Another important question is related to the "target unit" , i.e. to whom to address the incentive.

According to Dudley and Rosenthal (2006) 3 factors determine the choice of the unit: 1° Where the largest benefit can be achieved; 2° the share of covered services delivered by the providers (providers treating rare diseases are in this view less interesting targets); and 3° available performance measures and existing data for each type of provider.[78]

A related question is whether the programme should be focussed on a manager of a department, an individual clinician, or a department or group of physicians [67].

Most studies on P4Q have not distinguished between the effects of incentives that target the physician organisation and those that target the individual physician.

Targeting incentives at the individual provider makes the accountability clearer and implies that the target provider is more in control of his actions.

Targeting incentives at the medical group or hospital system level can also be beneficial because it can encourage collaboration, coordination and interaction. Also, if the performance measurement system is subject to some variation, this variation is expected to be averaged out [78]. On the other hand, the free-rider phenomenon may occur here as well when targeting provider groups [35].

An automatic question that then rises is what the role of the "meso" level will be (e.g. the head of department): will this meso level play the role of a principal or of an agent [67]? Referring back to the agency theory, definition of the principal and the agent requires careful consideration. Suppose that in a P4Q programme, hospitals are the target audience (hence the agent): the incentive payments go to the hospital for performance according to the standards of the principal. To successfully improve quality of delivered care, the hospital as an agent must rely on the cooperation of their medical staff and other clinical people, who are often not employees of the hospital [107].

Physicians enjoy a monopoly in several major decision areas: the decision to admit patients to the hospital, the decision to perform procedures, the decision regarding which procedure to perform, and the decision to prescribe pharmaceuticals. This professional autonomy is reinforced in a fragmented financing system, paying physicians on a fee-for-service basis and hospitals on a prospective payment basis.

This dual split may create conflicting goals and is often cited as a major obstacle to effective collaboration. Financial incentives for doctors and hospitals to do the right things or to do better are often mismatched or even in conflict [114, 115]. Better alignment of incentives is one of the expectations in the pay-for-quality world [116]. A possibility is that the hospital may make a part of the incentive payment available to clinicians responsible for quality improvements, thereby to motivate their cooperation. An alternative form of this "gain sharing" can be developed to afford physicians direct payments as an incentive, not to improve efficiency but to improve overall hospital quality. Another form of shared gain (or risk) is the bundled payment in which the physician and hospital are paid together in one lump sum, which then must be divided among the different specialists participating in the patient's treatment. Finally some specific pay-for-quality models compensate physicians for clinical improvement that require collaboration with hospitals, or reward hospitals for improvements that may require physicians to collaborate. This kind of compensations encourages the needed collaboration between hospitals and physicians in joint quality improvement initiatives. Further research on Hospital-Physician relationships, who are at the centre of several policy proposals such as pay-for-quality, gains sharing and bundled payments is required in this regard.

Organisational aspects

Regardless of how the target unit is defined, organisational aspects at the provider's side need to be taken into account. When participating in a P4Q programme, providers may need to create patient registries, use support staff to monitor medical management and patient compliance with preventive and treatment protocols, and adopt information technology to improve access to patient data [54]. Hence, there may be little value in establishing ambitious performance targets based on process or outcome measures if providers have weak information systems and poor office systems for managing patient care [78].

Moreover, there may be costs associated with complying with the programme [75], and the response of providers is likely to be influenced by their costs of performing the tasks necessary to improve. This can be considered in economic terms as an opportunity cost [38]. Hence, the reward should address these additional costs in the design. Obviously one should also take into account the possible benefits. This relates to cost-effectiveness of P4Q and is being discussed later in this report.

Finally, the number of patients in a practice, the quantity of services per patient [35, 38], but also the physician's age, gender, specialty, years since completion of the training, etc. can influence the compliance of physicians with the guidelines. [37, 38, 78]

3.2.3 The patients

Several patient characteristics can influence the outcomes of a programme [38, 54]. For instance, age, education level, insurance status, socio-economic status, etc.

Also their awareness of the programme (are patients aware of prices, and financial aspects of the programme, do patients receive information about the provider's behaviour) is of importance. Several authors notice that such disclosures should be handled carefully to safeguard the patient provider relationship [117].

Especially the presence of co-morbidities in patients and how this affects best practice care is of concern to some [118, 119]. Finally, the patient has through his own behaviour a large influence on certain P4Q outcome targets. His lifestyle, cooperation and level of therapeutic compliance will co-determine his health evolution, next to provider action. A general principle is to safeguard P4Q purposes by assigning accountability only to a degree that corresponds with clear responsibility and control. Therefore patient behaviour has to be taken into account [120].

3.3 IMPLEMENTING AND COMMUNICATING THE PROGRAM

In this paragraph, we discuss, largely based on the above knowledge, which steps are to be involved when initiating a P4Q programme, and which elements require special attention in that process. Importantly, these steps should be considered in the context of a quality circle, whereby the measurement phase leads to adjusting the goals, design and implementation.

Understand the concepts and the context

A logic requirement when planning the introduction of a P4Q programme is to understand all the above after carefully considering all relevant mediators and reaching consensus in terms of the choices and options which have to be assessed and decided upon. Specifically for a semi-decentralized system like Belgium this means that the political context within which a possible programme is implemented must be clarified (responsibilities of the federal and the regional health authorities)

Making money available

Obviously, introducing a P4Q programme implies that money is made available.

Potential sources of funds for a P4Q initiative include (see Dudley and Rosenthal 2006 [78], and see above -incentives):

- New money
- Redirection of existing money
- Reallocation of payment among providers, e.g., through a combined bonus-penalty payment scheme.
- Cost savings resulting from improved quality.
- The latter seems to be sometimes wrongly conceived as a necessary consequence of P4Q programmes (see for instance Young et al, 2005 [54]). Yet, improving quality is not necessarily associated with net savings.

Stepwise introduction

Payers introducing a P4Q programme may consider a stepwise introduction, also called phasing [78]. Advantages for phasing in P4Q are that it permits testing of measures before full scale implementation, giving providers time to gear up for a P4Q initiative; and enables purchasers to evaluate the small scale impact before applying it to the larger group of providers.

Options for phasing in P4Q include the following:

- Pilot test a payment scheme in a limited geographic area.
- Focus on specific provider types or clinical areas.
- Begin with pre-existing, national target sets and add targets over time.
- Rely on existing data (most likely billing data) and incorporate additional data as needed over time.
- Begin with a voluntary system.
- Begin with (private) quality reports and introduce incentives over time.
- Begin with a modest benchmark for performance and raise the standard over time.
- Begin with requiring or rewarding data collection and reporting and introduce performance incentives over time.

Communication

According to Rosenthal (2008), earlier P4Q efforts are perceived to have had too little impact on provider behaviour [69]. The way the payer communicates with the provider may therefore be crucial in the success of a programme (see agency theory and provider motivation). Young et al. (2005) cite different communication approaches such as local and regional meetings, hardcopy and electronic mailings, and websites [54].

Several theories focus on effective communication aimed at changing individual attitudes and behaviours [121]. For instance, the Persuasion-Communication Model presents a stepwise model of persuasion: exposure to a message, attention to that message, comprehension of the arguments and conclusions, acceptance of the arguments, retention of the content, and attitude change (McGuire 1985 [122], in Grol et al, 2007 [121]).

A critical step in any P4Q programme is to involve providers early so that they can help payers to identify performance indicators or measurement systems that meet providers' standards for validity and by facilitating cooperative relations needed to maintain provider participation [78]. This also involves a planned negotiation process between payers and providers [35].

Yet, it is questioned in the literature to which extent the providers should be aware of *all* the details and subtleties of the programme. Perhaps, if the programme is addressed to an organisation or to a group of physicians (see target unit, page 27), the "leader" of that organisation or group needs to be aware of all these subtleties, but not the individual practitioner. On the other hand, one may argue that a programme designed without the active participation of all participants and without their strong understanding of the nature and the rationale of the changes is bound to fail (see motivation, page 26).

Identify sources of data to perform the evaluation

Learning about the impact of a P4Q programme can be particularly challenging because a multitude of additional forces simultaneously affect the quality of patient care and costs [78]. Therefore in any P4Q programme, the availability of correct data for evaluating the performance of providers is of crucial importance. These data must be reliable and integer, in order to allow correct evaluation and to install feed-back processes regarding the achievement of the targets [54].

Also, some care regarding the design of the evaluation process is needed to disentangle the effects of the programme from other trends. At a minimum, payers should collect baseline data on the targeted quality measures [78].

Particularly for the hospital context, two important requirements need to be met in order to allow for a P4Q programme related to avoiding in-hospital complications [95, 123]. the ability to distinguish diagnoses that are present at admission from diagnoses that develop post admission, and the ability to identify diagnoses that represent complications that are potentially preventable.

An essential part of evaluation is the evaluation of the programme itself: the validation of the programme (did it meet its overall objectives according to the planned process); the sustainability of the realised changes; and the financial impact and return on investment. As we will see in the next chapters, few papers have already tackled the latter issue: is investing in P4Q programmes potentially cost-effective, and under which circumstances? This involves a regular review of processes and content. Note thereby that cost-effective does not necessary means that the investment in the P4Q programme is completely recuperated by savings in the system. It may be that there is still a net cost (a net investment) which is then to be balanced with the health gain.

3.4 P4Q PRINCIPLES FOR DESIGN AND IMPLEMENTATION

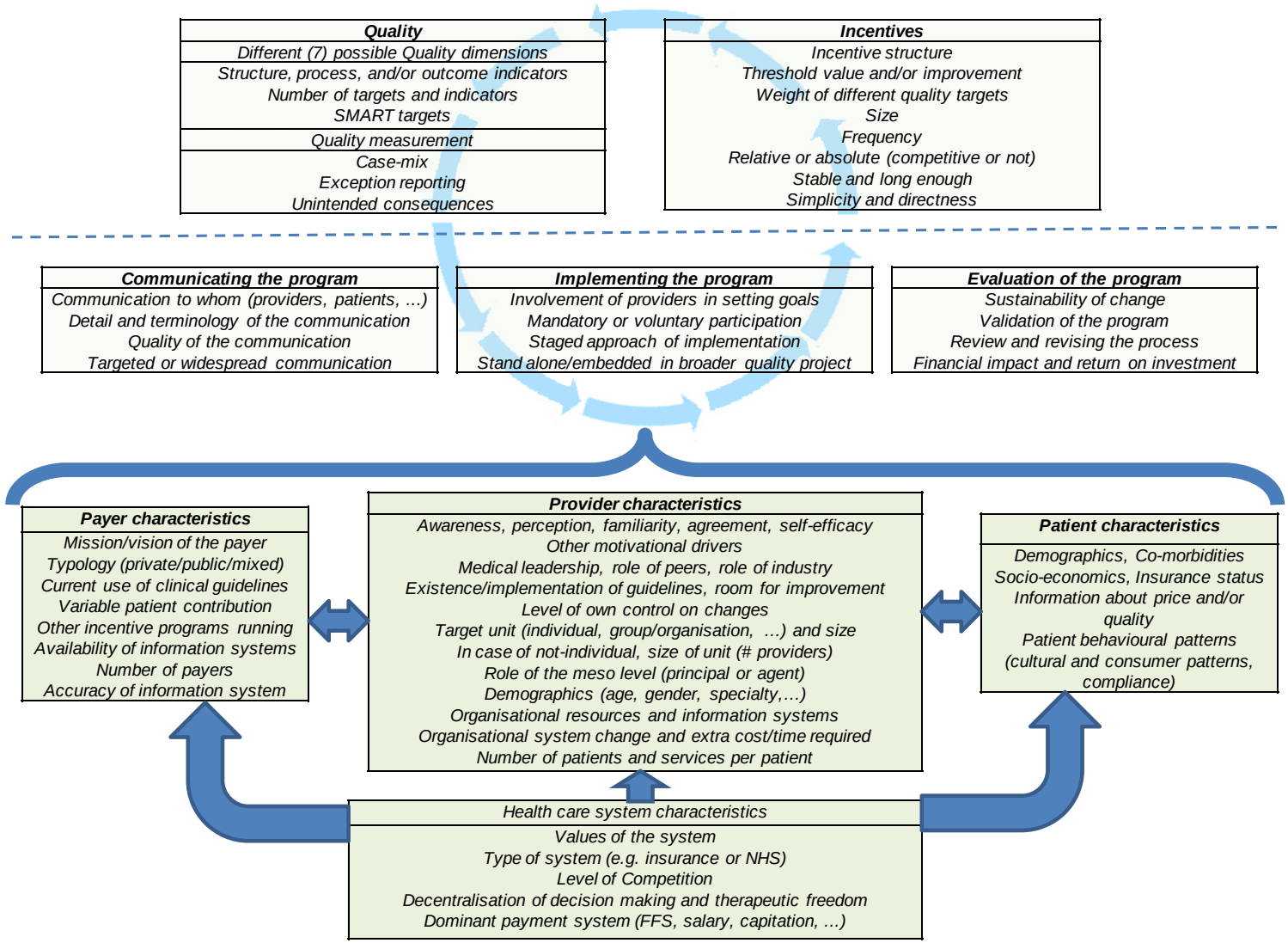
In this section, a multifactorial implementation framework (see Figure 2) is proposed that can be applied by designers of a new P4Q programme. The 'MIMIQ' model (Model for Implementing and Monitoring Incentives for Quality), which can also be interpreted as a checklist, is generally applicable, and contains different items which are not necessarily to be considered as mandatory elements of a P4Q programme, but rather as aspects to keep in mind and to challenge when designing a programme.

The framework is in our view comprehensive and complementary to previously published models/checklists. For instance, Dudley and Rosenthal (2006) published a checklist focussing on the content of P4Q programmes [78].

They distinguished 4 phases, thereby setting the basis for a time dimension in the design and implementation of P4Q: the contemplation phase, the design, the implementation and the evaluation. Based on other conceptual models identified in our review, we build further on this phased approach in our checklist. Figure shows how we made a distinction between questions related to the context and actors (health care system, payers, providers, patients) on the one hand and the programme on the other hand. Finally, the interplay between both creates the full P4Q picture.

Later in this report we will come back to this model on different occasions, since based on the existing evidence of what works and what doesn't (chapter 4) it will become more clear to which extent the different elements of the model need to be complied with in order to increase the chances for a successful programme.

Figure 2 : P4Q conceptual implementation framework: Model for Implementing and Monitoring Incentives for Quality (MIMIQ)



Key points

- **P4Q programmes are intended to offer explicit financial incentives to providers in order to achieve predefined quality targets. Quality goes further than the strict clinical outcome. It includes 7 dimensions: safety, clinical effectiveness, patient centeredness, timeliness, equity of care, efficiency of care and continuous and integrative care[b]. Furthermore quality can be expressed in structural, process and outcome quality criteria. Many P4Q programmes focus mainly on structure and process outcomes. However, given the scarcely grounded relationship between process measures and long term patient outcome measures, overly relying on structure and process outcomes threatens the credibility of a P4Q programme. As important as the choice of criteria is the final number of criteria. Too few criteria could draw the attention of providers away from the not incentivized criteria. Too many criteria could lead to organisational complexity. When setting these criteria, it is important to verify that the criteria are measurable. A valid and comprehensive management information system to track performance against the goals must be available. Correct measurement makes use of risk adjustment. Some P4Q programmes make use of exception reporting. Gaming, patient selection and diversion of attention are the most important types of unintended consequences.**
- **Financial incentives are considered as the core of a P4Q programme. Several possible incentives structures are possible: bonuses, performance based fee schedule, performance based withholds, regular payment increase linked to performance and quality grants/financial awards/performance funds. There is still some disagreement amongst researchers about whether bonuses or penalties should be applied. Most programmes make use of bonuses whereby ideally, an incentive size should amount up to 5% of income and according to some authors even up to 10% of income. The formulation of the incentive arrangement is of importance. Rewarding a threshold could discourage low-quality providers to engage in the P4Q programme. P4Q programmes often provide payment only on one or a few fixed time points. Some P4Q programmes attach payment weights to specific quality targets as a function of the estimated related workload, or to express the relative importance of a target in terms of public health. Some P4Q programmes make use of an absolute reward whereby anyone who performs well obtains the reward no matter how the other providers perform. Other P4Q programmes use a 'tournament approach', where providers compete against one another. The latter method has the advantage that the expenses are more under control, however the uncertainty about what can be achieved could provoke providers not to engage in the programme. Finally, a balance should be found between the simplicity of the programme on the one hand and sufficient attention for all the issues related to P4Q complexity on the other hand.**
- **Health care, payer, provider and patient characteristics are the main contextual factors and are of influence in various ways. Not taking these factors into account will compromise the success of a possible P4Q programme.**
- **Several market and environmental characteristics, like the private/public mix, the dominant payment system, the level of competition and the level of therapeutic freedom among providers, can drive quality, and thus P4Q success.**

[b] Quality on a global level also involves reducing variability in care

- **The principal agent theory broadly addresses relationships in which both parties have different abilities, in which there is asymmetric information and in which the parties have different goals. Related to a P4Q programme, the "agent" (the health care provider) can be both a potential ally and a potential source of resistance to P4Q.**
- **The use of clinical guidelines in current policies, variable patient contributions, other quality programmes, the availability of management information systems, the number of payers, the accuracy of the data system, the vision of the payer regarding health care goals and the typology (private/public/regional/…) are important mechanisms that influence the way a payer can and will implement P4Q.**
- **Internal and external motivational drivers, the specifically targeted "unit" (an individual or a group) and organizational aspects are of importance in the behaviour of health care providers.**
- **Finally, patient demographics, co-morbidities, their socio-economic and insurance status, information about price and quality and several patient behavioural patterns influence the outcome of P4Q programmes.**
- **When planning the introduction of a P4Q programme, all relevant concepts and contextual factors have to be understood and taken into account. Furthermore, when introducing P4Q, money has to be made available, either new money or by disinvestments elsewhere or planned savings within the programme. A stepwise introduction (phasing), permits testing the targets and indicators, gives the providers the chance to gear up for a P4Q initiative and enables purchasers to evaluate the small scale impact before applying it to the larger group. The way of communicating the programme to the providers is seen as crucial in the success of the programme. The evaluation of the programme itself is as important as communication and implementation.**

4 THE EVIDENCE BASE FOR P4Q: A SYSTEMATIC REVIEW OF THE PEER REVIEWED LITERATURE

4.1 INTRODUCTION

In this chapter the evidence from peer reviewed literature for P4Q is reported. This chapter starts from a systematic review of the already published primary P4Q evaluation studies.

The aim is to come to a comprehensive overview of results in terms of P4Q effects in relationship with their context, design and implementation process. Due to the amount of newly available research findings, there is a need for a synthesis that strengthens the applicability of theoretical P4Q concepts.

The following research question is addressed:

What is the effect of P4Q on the different quality domains (effectiveness, equity, safety, coordination, cost effectiveness, etc.), describing both intended and unintended consequences?

The methods used for this review are described in Chapter 2. Chapter 4 focuses on the empirical evidence, based on the experience with or observation of P4Q programmes.

Equity related aspects were described in a separate paragraph (see paragraph 0).

4.2 DESCRIPTION OF STUDIES

4.2.1 General description

Appendix 15 provides an overview of the 104 studies that were included in the review.

The number of P4Q evaluation studies increases steadily. In the nineties only a few studies were published each year. Since 2000, this number is increasing rapidly to more than 20 studies per annum in 2007 and 2008.

The majority of studies are from American or British origin. In the period 1990-2008, 54 studies were published in the USA and 45 in the UK. Two studies were conducted in Australia while Italy, Germany and Spain each published one evaluation study on P4Q. There are also a limited number of Canadian, Swedish, Czech, Estonian, Chinese and Taiwanese P4Q publications. However, none of these described an empirical P4Q evaluation and were therefore excluded from further analysis in this Chapter.

Of a total of 104 studies, 91 evaluate P4Q use in a primary care setting while 25 studies focused on a hospital setting. Some of these studies take place in both settings.

In terms of study design, studies are classified in six groups:

- 'Randomized' : randomized controlled trials and cluster randomized controlled trials with randomization of group allocation;
- 'Concurrent + Historic' : studies in which evaluation over time is combined with a concurrent control group without randomization;
- 'Concurrent' : studies with a concurrent control group, without an evaluation over time and without randomization;
- 'Historic, Multi' : studies with evaluation over time with more than two time points of data collection, without a concurrent control group and without randomization;
- 'Historic, before-after' : studies with evaluation over time with two time points of data collection, without a concurrent control group and without randomization;

- 'Cross section' : studies with data collection at one time point, without a control group and without randomization.

Nine randomized P4Q evaluation studies were published between 1990 and 2008. Fifteen studies make use of a 'concurrent + historic' design. One study uses a 'concurrent' comparison design. Eighteen studies are based on a 'historic multi' design. This number rose during more recent years in the UK. Sixteen studies use a 'historic before-after' design and 45 studies use a cross sectional design.

Three aspects of sample size can be reported, as described in

- number of practices or organizations (i.e. family practice, medical group, independent practice association, hospital, etc.);
- number of providers (i.e. individual physicians);
- number of patients.

Most studies only focus on one of these three aspects as the unit of analysis and therefore often omit to report the sample size of the other aspects.

Key points on study characteristics

In terms of general study description the key points are:

- **The number of P4Q evaluation studies is increasing rapidly.**
- **The US and the UK are building a large body of P4Q knowledge and experience. Australia and the Netherlands are starting to report the use of P4Q as a health system intervention.**
- **The majority of studies focus on the primary care setting. However, in the US there are also reports on P4Q in the hospital setting.**
- **Although there are a number of randomized studies, most P4Q study designs are of an observational nature.**
- **Sample sizes of the studies depend on the unit of analysis within a patient, provider and organization hierarchy. The sample size is often not reported on all three levels.**

4.2.2 Contextual aspects

4.2.2.1 United Kingdom

All UK based studies take place in the National Health Service (NHS) in which there is almost universal coverage of healthcare expenses, provided by the state as a direct payer. Insurer and payer choice and competition are limited but patients are free to choose a provider. There are no out of pocket expenses for the patient (free at the point of access, free prescription drugs). P4Q in the UK is focused on primary care by general practitioners (GPs). Most studies are conducted at the practice level (solo or group), not on the individual provider level. The 'Payment by results' scheme recently introduced to reimburse hospital care is based on a case (Diagnosis Related Groups) centred prospective payment system with new productivity and efficiency incentives. This falls outside the P4Q definition scope.

The NHS periodically negotiates a national standard contract with GPs who provide care independently. There are various regional levels which assist in local prioritization of healthcare policy. The first is the level of UK semi independent regions: Scotland, [124-137] Wales, [124, 129, 130, 138] Northern Ireland, [130, 139] and England. [42, 84, 124, 130, 131, 133, 140-168]

Some differences in health policy are described between these regions [130]: In England there is a stronger focus on a more centralized performance management, on a regulated healthcare market around patient choice, on increased plurality of providers and an increasing role for the private sector. Scotland makes more use of vertically integrated health boards without a purchaser provider split: the state is responsible both for healthcare insurance and payment as for healthcare delivery, without a separation of these responsibilities. In addition, Scotland shows a strong emphasis on professionalism and clinical leadership in managed clinical networks.

Wales has a stronger focus on public health and local partnerships. In Northern Ireland, no major reforms have recently been conducted [130].

These regions are further divided into a number of strategic health authorities, each responsible for a primary care sub region. Many studies are conducted at this level to define their study scope in terms of region and population. A number of UK studies explicitly focus on an urban area [126, 138, 140, 149, 150, 151, 153, 154, 155, 156, 157, 158, 161], others on a rural area [129], or both [153, 165]. In the other UK studies this characteristic is not reported. Some studies investigate P4Q in a more socio-economical deprived area [125, 136, 149, 150, 153, 154, 155, 156, 157, 158, 161], while others are conducted in a more affluent area [126]. Some authors explicitly state that the study region and population is nationally representative in terms of these characteristics. Other UK studies do not report on these characteristics. Similar comparisons are made in terms of ethnicity [150, 154, 155, 156, 157, 158, 161] and population age distribution of the study area as a whole [136, 150, 154, 155, 156, 157, 158]. A few authors report the level of representativeness of the study area characteristics for the whole health system under consideration [151, 158]. The dominant payment system for UK general practitioners is capitation (based on the number of patients, adjusted for characteristics of the patients and the area such as deprivation level).

The following providers were the focus in UK studies:

- General practitioners are the main target group, at a general practice level. Only a few authors include other provider specialties such as community child health doctors when describing the pre QOF (Quality and Outcomes Framework) P4Q schemes [132]. See section 4.2.4.1 (page 43) for more information on QOF.
- Inclusion of providers is often based on data availability in national or regional clinical databases using automatic data extraction from electronic patient records.
- Only a few UK studies exclude practices based on additional criteria such as a minimal yearly number of patients per practice, missing disease registries, practice relocation, and change in practice population size [84, 130, 139, 142, 148, 167, 169].

4.2.2.2 United States of America

The market, payer and provider characteristics in P4Q evaluation studies in the USA are very heterogeneous. Providers have multiple payers, private as well as public. Each of these is involved in purchasing health care resulting in about 300 health insurance plans. The dominant payment system in the US often is a mix, but Fee For Service (FFS) is the most often used for physician services.

Some studies focus on a privately insured population [69, 170, 171, 172], others on a publicly insured population [173], often provided by Medicare for more socio economically deprived patients [44, 95, 174, 175, 176, 177, 178, 179]. P4Q evaluation studies that focus on Medicaid were also included [145, 180, 181, 182, 183, 184, 185, 186]. Only a few studies describe for which percentage of patients of an included provider or provider organization a health plan accounts [74, 171, 187].

The P4Q programmes in the USA were directed at primary care and, to a more limited extent, at hospital care (see Appendix 15). Some programmes include both primary and hospital care. This can be a regional provider network [145, 188] or an integrated healthcare delivery network [189, 190, 191, 192, 193].

Primary care in USA studies is often defined more broadly than in non USA studies. In addition to family physicians, many studies also include paediatricians [171, 172, 173, 181, 184, 194], specialists in internal medicine [69, 170, 171, 172, 194, 195] or geriatricians [194] as primary care providers. Only a few studies specify whether solo [181, 187], group or both types [178] of practices are included.

Healthcare insurance and payment are, in addition to Centers for Medicare & Medicaid Services (CMS), mostly organized on a regional or local level, with a diversity of forms and systems.

Based on free market competition a number of commercial for profit Health Maintenance Organizations (HMOs) include P4Q in their insurance packages, contracting by negotiation with providers as a form of managed care [188]. Only a few studies report the alignment and integration of different health plans' strategies [196].

Most providers are further organized into medical groups [74, 172, 192, 196, 197, 198, 199, 200, 201, 202, 203] or independent practice associations (IPAs) [171, 183, 184, 187, 198, 199, 202, 203]. These organizations often represent hundreds of providers and act as negotiation and contracting partners with the insurers. In medical groups this goes often further in terms of clinical healthcare organization and coordination. Besides the network and group HMO models of managed care, some authors specify indemnity, point of service, and staff HMO models [204]. The following attributes are assigned to the managed care models in terms of gate keeping:

- Network HMO (IPAs): healthcare is not covered outside the network
- Group HMO and Staff HMO: healthcare is not covered outside the group unless recommended
- Point of service: there is a preferred network without imposing restrictions [205]
- Indemnity: free choice of healthcare providers [206]

Some studies focus on a rural area [177], an urban area [44, 180, 181, 182, 184], or both [171]. Area representativeness is rarely tested in USA studies for other characteristics, with the exception of some studies that focus on a more deprived area [145, 180, 181, 182, 184, 185, 195]. However, provider characteristics are reported upon in more detail. The size can be large [74, 145, 183, 187, 189, 190, 191, 193, 196, 201] or small [44, 188, 195]. Some studies exclude smaller sites to reach sufficiently large sample sizes [184]. Other use such restrictions on the medical group/IPA level [172, 186, 198, 199, 200, 202, 207, 208]. This can lead to a 90-95% exclusion rate [209]. Ownership can be private [182, 188, 197] versus public [210]. Teaching status can be academic versus non academic [44, 188, 195].

The general payment system is often a mix. In many studies capitation dominates [74, 171, 179, 183, 184, 187, 188, 203], but sometimes fee for service is dominant [175, 178, 206]. Medical groups often make use of capitation payment. Indemnity models make more use of fee for service, but have a low presence in the systematic review sample. Fee for service providers are excluded in some studies [185]. Salary as a dominant payment system is rare [145]. Diagnosis Related Groups (DRG) based lump sum payment is often a component in hospital reimbursement [211].

Often a selection is made in terms of provider in- and exclusion criteria. These criteria are:

- An 'active provider' indicator: providing care to a minimal number of patients [171]. Possible thresholds are: 25 patients [184], 50 patients [177, 178], 20 hours of patient care per week [194, 212]
- An 'active provider' relative to a specific patient group: providing care to a minimal number of target patients. Thresholds are: 50 patients per month per provider [195], 2500 claims submitted [181], 15 or more denominator patients [206], at least 30 cases per condition annually [44], at least 20 patients in each time span [193], have 10 target patients continuously enrolled [187], having at least 200 episodes with 10 patients eligible for a certain measure [69]
- A 'stable patient panel' indicator: exclusion of rapidly growing practices [177]
- An 'intervention experience' indicator: inclusion of clinics that had applied the intervention for four years or longer [210], been a member physician for at least 24 months [187]
- Exclusion of residents and fellows [194]
- Exclusion of specific specialty groups such as radiologists and pathologists [194, 199, 202, 209]
- No refusal to cooperate with study procedures [195]
- No moving or retiring of the provider involved [195]

- Patient attribution based on which physicians had the greatest number of claims and their eligibility for responsibility for measures based on specialty [69]

4.2.2.3 Australia, the Netherlands, Germany, Italy, and Spain

In Australia there is a national health insurance arrangement which is supplemented by private health insurances (about 40). Two major branches can be distinguished in Australian health care, namely the medical setting (=primary care), which consists mainly of general practitioners, and hospitals settings (=secondary care). The responsibility for medical services rests with the commonwealth government, the responsibility for hospital services rests essentially with the states. Medical services are provided on a FFS basis.
One Australian P4Q evaluation study is performed in emergency departments of public hospitals. Private hospitals and small country hospitals were excluded [213]. The other study focused on family day care provided by child care centres and councils in a metropolitan area [214]. The market, payer and provider characteristics are not further specified in these studies.

In the Netherlands, a new law relating to care insurance has been introduced in 2006, which resulted in a major switch from supply-driven care to demand-driven care. Consequently, managed competition has been introduced in health care. Currently five major private health insurances can be distinguished. These insurances are responsible for purchasing health care based on quality and price. General Practitioners are largely remunerated in a mixed model: capitation augmented by FFS. Hospitals are increasingly commercial and payers negotiate with providers for prices. This allows competition between providers and even across primary and secondary care.
One Dutch P4Q study was found, aimed at general practitioners with availability of complete records in practices with at least 500 patients, and covered by one local insurance company. No further information on these characteristics is provided [215]. However this Dutch study does not explicitly recognizes and rewards high levels of quality and quality improvement, therefore this programme can not be seen as a P4Q programme as defined in this report, consequently the results of this study will not be taken into account in the evidence section.

The included German P4Q study is also aimed at general practitioners, without further specification of market, payer and provider type [216]. These authors used the degree of initial interest in participation as a practice inclusion criterion.

In Italy, one P4Q study was performed in a National Health Service environment. Six local health authorities were included as payers. The providers consisted of public maternity hospitals and immunization clinics [217].

Finally, in Spain, the Catalan Institute of Health offers primary healthcare services to 80% of the population in the region. The targeted primary care teams consist of physicians and nurses [218]. No further details are provided on these characteristics.

Key points on contextual aspects

United Kingdom

- **In the UK, P4Q is implemented in a national NHS regulated healthcare system, focusing on primary care.**
- **Market and payer characteristics include uniformity of the system and the payment scheme, universal health insurance, low competition, absence of gate keeping, free care at the point of access and case mix adjusted capitation as the general GP payment system.**
- **P4Q is arranged by a national contracting approach with participation of providers at the practice level.**
- **Many studies focus on the differences of care between urban and/or deprived regions in the UK.**
- **Inclusion of providers in studies is based on data availability within the national electronic data extraction system. Exclusion of providers in the studies is limited.**

USA

- **Contextual characteristics in the USA are very heterogeneous.**
- **Insurance is based on a public-private mix, with non-universal coverage. Health Maintenance Organizations (HMO) and Preferred Provider Organizations (PPO) contract with providers on a more regional or local level from a managed care approach. Gate keeping is present to a varying degree.**
- **Providers are often organized into medical groups or IPAs. Any provider's payment is based on multiple schemes from different health plans. The general payment system consists of a large part of fee for service, although capitation and fixed payment use is growing. This contrasts with the relatively high number of studies focusing on capitation based healthcare delivery. There is a high level of competition between professionals.**
- **P4Q in the US is diversely implemented in both primary and hospital care settings. Primary care is more broadly defined in the US than in most European countries, i.e. outpatient care in hospitals is considered as primary care.**
- **US studies focus less on area and population characteristics compared to the UK, but more on provider characteristics such as the size of an organization, ownership, teaching status and payment components.**
- **Provider selection criteria used in studies consist of indicators such as providers' activity level, panel size and stability. Because P4Q is organized on various levels in the USA, patient attribution to providers based on responsibility, is sometimes unclear. This is in contrast with the UK where patients are assigned to a certain physician who is responsible for the well-being of his patients.**

Australia, the Netherlands, Germany, Italy and Spain

- **One Australian study focused on public hospitals' emergency departments, the other on family day care in a community setting.**
- **One German study focused on general practitioners.**
- **One Italian study focused on public maternity hospitals and immunization clinics in a NHS environment.**
- **The Spanish study involved primary care teams of physicians and nurses. No further details are provided on these characteristics.**

4.2.3 Patient characteristics

In terms of type of disorder, the focus of P4Q can be preventive, acute and chronic care.

- Preventive care includes screening and immunization, prevention of Sexually Transmitted Diseases (STDs), child health, promotion of breast feeding, etc.
- Acute care includes emergency care, acute cardiac events and cardiac surgical procedures, stroke, pneumonia, surgical joint replacement, a number of infectious diseases and healthcare complications.
- In comparison with preventive and acute care, P4Q in chronic care focuses on chronic conditions like diabetes, asthma, chronic obstructive pulmonary disease (COPD), chronic heart failure, hypertension, coronary heart disease, depression, cancer care, cataract, epilepsy, hypothyroid diseases, chronic kidney diseases, osteoarthritis, osteoporosis, smoking cessation, urinary incontinence, hearing problems, pain management and falls.

In addition, a number of P4Q studies focus on other than patient related indicators. This includes patient and/or provider satisfaction studies and care management process (CMP) implementation studies. CMPs include the use of IT, guidelines, clinical pathways, etc. to support healthcare quality improvement.

Both the specific patient group results as the generic study results are summarized in section **Error! Reference source not found.** (page **Error! Bookmark not defined.**) and complemented with references to the relevant P4Q studies.

In addition to acute, chronic and preventive care, the following patient inclusion and exclusion criteria are used:

- Clinical selection criterion: some studies specify further clinical criteria to refine the diagnosis based patient group selection:
  - Acute sinusitis. Identification of episodes of acute sinusitis without sinus surgery, based on ICD-9-CM codes and claims up to a 'clean' period of 60 days of no related service [171].
  - Asthma. Persistent, not exercise or cold induced [191].
  - Coronary artery bypass grafting (CABG). Exclusion of non elective surgery [189].
  - Joint replacement. Exclusion of non elective surgery [211].
  - Diabetes: had diabetes for at least one year and were not pregnant [219], no gestational diabetes [144, 150, 155, 156, 158, 191], no women receiving treatment for polycystic ovarian syndrome [150, 155, 156, 158]. A number of studies explicitly focus on diabetes mellitus type 2 [193]. Some include both type 1 and type 2 diabetes mellitus [128, 131, 150, 155, 158]. The other studies did not specify the diabetes diagnosis further. To account for undiagnosed diabetes patients some authors use additional inclusion criteria to detect all diabetes patients: receiving medications for diabetes or HbA1c > 7.4% [150, 155, 156, 158]. Concerning the stage and timing within the diabetes evolution: diabetes on problem list in last 12 months or in ICD 9 entered four times in last 24 months [193].
  - Myocardial infarction. Acute non ST segment elevation [174].
  - Coronary heart disease. CHD diagnosis, or a repeat prescription for nitrates, or a positive angiography test, cardiac bypass surgery or coronary angioplasty confirmed by diagnosis in medical record [157].
  - Hypertension. Patients with cardiovascular co morbidities [154].
  - Smoking cessation. Smoked at least 10 cigarettes a day, irrespective of their intention to stop smoking [216], currently smoking cigarettes, intended to quit within the next 30 days [196], patients who use tobacco [220].
  - Vaccination. Exclusion of patients presenting with contra indications [181].

- Age selection criterion: children as target group [125, 132, 173, 180, 181, 182, 184, 188, 191, 210, 214, 217], the elderly as target group [177], patients older than 50 years [162, 170, 183, 195, 202], inclusion of adult patients only (sometimes depending on measure selection) [124, 127, 128, 144, 154-156, 157, Morrow, 1995 #385, 159, 191, 193, 196, 201, 204, 219, 220], 36 to 75 years of age [216].
- A 'regular patient' criterion: patient seen in practice during six months [184, 188], more than one 'well check' performed [188], minimal one year as a patient [188], minimal three times an office visit [180, 188], minimal one visit during the last year [181, 182, 188], minimal one office visit during the study period [178], having an office visit during the three months prior to study period [201], no mention of patient moving or leaving the practice [181, 182], patients with a regular personal physician [204], regular attendance [214], receiving care for a total of at least six months [179], continuously enrolled for at least 18 months [219], used at least one service during 18 months [219], patients who have been in the practice's care for at least one year [69], seeing no more than two physicians or two physicians equally often [144]. One study focuses explicitly on patients not in regular contact with their provider as a criterion for inclusion: patients with no visits for more than four months [179].
- A language criterion: patient speaks English or Spanish [219], an informed consent ability criterion: patient could give informed consent [219], informed consent provided [163], a physical or functional status criterion: functional limitations or increased need for healthcare services or dependence on medications or home medical equipment [173], a socio-economical status criterion: low income members of minority groups [144], an ethnicity criterion: patients with an African American ethnicity [176], a home setting criterion: non nursing home patients [177], lived in the community [219], patients in private households [162] and exclusion of loss due to death or transfer out [195].
- A number of UK P4Q studies based on the QOF, exclude exception reported patients from further analysis [127, 128, 134, 141]. Many studies do not specify how exception reporting has been dealt with. Some studies do an additional control or subgroup analysis for as exception reported patients or interventions [84, 127, 128, 130, 132, 133, 135, 147, 148, 160, 163, 165, 167, 171].

Key points on patient characteristics

- **Medical conditions targeted by P4Q include mainly preventive care (e.g. immunization, screening) and chronic care (e.g. diabetes care, asthma care). The lower number of P4Q initiatives in acute care focus on conditions as myocardial infarction and pneumonia. Interest in P4Q is growing in many medical specialties, including surgery, but without evaluation results at present. Some studies are not medical condition or patient group specific, but focus on a wider use of disease management or care management processes. Inclusion of patients is often based on medical condition specific clinical selection criteria to define a clinically coherent patient group.**
- **Additional patient selection criteria included in a number of studies are: patient age (children, adults, elderly), the regularity and stability of the patient-physician relationship, patient language, socio-economical status, ethnic background and clinical parameters (e.g. functional status, loss due to death or transfer).**
- **Exception reporting of patients in the UK is often treated ambiguously in the included studies, with some including these patients and some excluding them. Often this remains unclear. A few studies report specific subgroup analyses.**

4.2.4 Description of P4Q interventions

4.2.4.1 United Kingdom

Primary care: Quality and Outcomes Framework (QOF)

British family practitioners derive their income from NHS patients. Almost all of the citizens of the UK are registered with a family practitioner, and family practitioners have registered lists of patients for whom they are responsible. Their work within the NHS is governed by a national contract, known as the General Medical Services contract that is negotiated between professional representatives of the British Medical Association and the central government. This contract is revised at infrequent intervals; since the NHS started in 1948, major contract revisions have been made only in 1966 and 1990 [55].

At the core of the general practitioners contract was the Quality and Outcomes Framework (QOF), which links financial incentives to the quality of care that is provided by practices. The contract for providing medical care and all quality payments relate to the practice rather than to the individual physician. Under the 2004 contract, responsibility moves from the individual family practitioner to the practice, which is a group of, typically, one to six physicians. In 2004, the National Health Service committed £ 1.8 billion ($3.2 billion) in additional funding over a period of three years for a new pay for performance programme for family practitioners.

The quality-of-care payments make up approximately 20 percent of the government's total family practice budget.

These figures are gross numbers and therefore include the additional costs (e.g., for the employment of nurses) that physicians may need to incur in order to deliver high standards of care. The percentage increase in available resources for individual physicians thus depends on the extent to which they have already invested in high-quality systems in their practices.

This programme was intended to increase family practitioners' income by up to 25 percent, depending on their performance with respect to 146 quality indicators relating to clinical care for 10 chronic diseases, organization of care, and patient experience.

1. Quality is measured against a set of clinical activity indicators relating to aspects of care for several common chronic diseases, with practices rewarded according to the proportion of eligible patients for whom each target is achieved. Family practitioners earn more points if higher proportions of these patients have undergone "process measures" and further points for "intermediate outcomes" (i.e. management of these risk factors within certain limits). Generally, more points are available for the intermediate outcomes than for the process measures, which reflect the increased difficulty of achieving these standards. The number of points that can be earned for each indicator was determined partly by the academic advisory group and partly by a formal scoring process undertaken by groups of family practitioners in England and Scotland. The intention behind this process was to allocate points on the basis of the workload required to provide care to the relevant standard.

   For the clinical indicators, practices claim points that generate payments according to the proportion of patients for whom they achieve each target. Points are awarded on a sliding scale within the payment range. For example, for asthma indicator number 6, practices gain points for clinically reviewing at least 25 percent of patients with asthma in the previous 15 months. The maximum of 20 points is gained if at least 70 percent of patients with asthma are reviewed.

   For 2004–2005, payment was limited to £ 76 ($133) per point, adjusted for the relative prevalence of the disease (payment is multiplied by the square root of the prevalence of the disease among the patients served by the practice and divided by the square root of the mean national prevalence of the disease). A maximum of 1 050 points was available, which was equivalent

to an average of £ 79 800 ($139 400). From 2005–2006 onward, practices have been earning £ 125 ($218) per point.

There are an additional 36 points for Papanicolaou tests, childhood immunizations, maternity services, and contraceptive services and 50 points for the achievement of high standards of access. An additional 130 points can be awarded to practices that score highly in all areas.

2. Organizational indicators were included in five categories: records and information about patients, communication with patients, education and training, management of medicines, and management of physicians' practices.
3. The rewards in the section that covers the experience of patients relate to the use of surveys in the doctor's own practice, which earns up to 70 points, and to the length of consultations, which earns up to 30 points. For the latter, there is an incentive for practices whose routine booking interval is 10 minutes or more. The introduction of patient surveys was a point of contention. Until recently, surveys have not been commonly used in the NHS, and doctors were suspicious of their use. Partly because of this, and partly because of limited experience with available instruments for evaluating patients' opinions, an early decision was made to reward doctors for surveying their patients and for acting on the results; however, payments would not be linked to the scores from actual questionnaires. Family practitioners may choose from one of two approved questionnaires, which cover categories such as access to and interpersonal aspects of care. In addition to making plans for acting on survey results, there is an additional incentive for family practitioners to involve their patients in these discussions.

The payments are in addition to the practices' core funding, which is based on the number of patients, adjusted for characteristics of the patients and the area (Doran et al., 2006).

There is a single common electronic medical record for all practices, and data on quality of care are extracted automatically from clinical computing systems of practices and are collated in the central National Health Service Quality Management and Analysis System (QMAS) database. The central collection of claims data allows the government to monitor overall implementation of the quality framework, and mechanisms are established to update the indicators when required.

The scheme allows practices to exclude patients who they deem inappropriate from specific indicators (exception reporting). Patients can be excluded based on one or more of the following reasons [84]:

- The patient has received at least three invitations for a review during the preceding 12 months but has not attended.
- The indicator is judged to be inappropriate for the patient because of particular circumstances, such as terminal illness, extreme frailty, or the presence of a supervening condition that makes the specified treatment clinically inappropriate.
- The patient has recently received a diagnosis or has recently registered with the practice.
- The patient is taking the maximum tolerated dose of a medication, but the levels remain suboptimal.
- The patient has had an allergic or other adverse reaction to a specified medication or has another contraindication to the medication.
- The patient does not agree with the investigation or treatment.
- A specified investigative service is unavailable to the family practitioner.

There is no limit on the number of patients whom family practitioners may exclude, although the physicians' decisions may be questioned at annual inspection visits by Primary Care Trusts, which are NHS organizations with managerial responsibility for primary care in geographic areas that contain up to 100 practices.

Family practitioners already had some experience with financial incentives from the limited use of incentive programmes for cervical cytological testing (i.e. Papanicolaou) and immunization that were initiated in 1990. This preceding P4Q system is evaluated by some of the included studies [125, 132, 140, 177]. A second incentive that was introduced in the 1990 contract provided generous remuneration for family practitioners to establish "health promotion clinics" to encourage preventive screening and lifestyle interventions. Some evaluation studies are available [153]. The incentives schemes can be seen as a precursor of the current QOF scheme. The problems experienced with gaming and data manipulation in those early years are countered as much as possible in the new incentives system by the use of a thorough audit process, with peer participation, and severe penalties whenever fraud is discovered. All included UK studies concern the latest QOF contract as described above unless they are explicitly referenced for the 1990 contract. Most authors omit a detailed description of P4Q as studied intervention.

Hospital care

Since the first of October 2008, the NHS has launched a demonstration project concerning P4Q in hospital care. The programme, called 'Advancing Quality', which is implemented in the North West region of the UK, makes use of incentives based on three types of targets: evidence based process measures, patients' quality of life after surgery and the patient's experience of provided care. The programme is supported by peer review and public reporting. Because the first evaluation results of this project are expected to be available in 2010, this project is not included in this review of primary P4Q evaluation studies.

4.2.4.2 United States of America

The characteristics of P4Q as intervention in the USA are very diverse, as is the case for the contextual characteristics. Almost all described studies concern bonus systems.

Similar to the UK QOF programme most USA programmes make use of thresholds (e.g. 20 up to 90%) to be met in terms of the percentage of patients eligible for a certain indicator. In most studies this figure represents the level of targeted achievement in patient number with indicator met (the higher thresholds) or the level of targeted improvement in patient number with indicator met (the lower thresholds).

The indicators used are of a structural, process and/or outcome measure nature. Whenever patient outcome measures are used, these are mostly intermediate measures instead of long term outcome measures (e.g. the tracking of HbA1c levels instead of the tracking of long term diabetes complications).

The size of the incentive varies:

- A fixed amount, e.g. $5 [144], $50 [195], $50-80 [69], $100 [220], $750-1 250 [193], $1 000-5 000 [181], $1 000-7 500 [182], $5 000-10 000 [201].
- A certain percentage of the providers' or provider organization yearly revenues, e.g. 0-5.5% [205], 0-7% [188], 0.8% [74], 0.5-1% [190], 1-2% [44, 174, 175, 176, 211], 2.2% [172], 1-7.5% [206], 4% [221], 5% [187], 10% [171, 177, 178, 183, 184], 12% [222] or 20% [171, 177, 178, 183, 184, 189, 191].
- Use of a bonus pool with a floor and ceiling amount [197], use of a maximum i.e. one bonus procedure per encounter [144].

Only a few authors report on programmes that base the incentive amount on the level of estimated savings due to quality improvement [69].

The base for incentive calculation and the target level vary between programmes. The base for calculation can be (sometimes combined):

- An incentive as a percentage of capitation per patient eligible [69, 183, 184].
- An incentive per patient with target met [74, 177].

- An incentive per intervention as incentivized quality indicator, e.g. per vaccination [144, 177, 178, 179, 180, 181, 182, 196].
- An incentive per composite measure level. The composite measure is the aggregation of separate indicator measures into one overarching measure [44, 174, 175, 176, 193, 206, 211, 222, 223].
- An incentive amount according to the number of providers per clinic [201].
- An incentive based on the DRG based hospital payment amount [44, 174, 176, 211, 221].
- An incentive based on an all-or-none target performance [189, 193].

The target level can be:

- The individual provider level [69, 177, 180, 182, 187, 193].
- The practice or organization level [172, 183, 184, 188, 196, 201, 210], the physician organization or medical group level [74, 197, 203], the hospital level [44, 174, 176, 211, 221].
- The team level of per patient responsible individual providers combined [171, 205, 206].
- Some authors direct the incentive at the different levels of a network [191].
- The supervisors' level [223].

In addition to additional bonus approaches, many USA systems competition and redistribution between providers or organizations based on their quality performance [44, 174, 175, 176, 187, 188, 191, 211].

This implies in most cases the use of a bonus and withholds combination, imposing reward and punishment. In a few cases it concerns a bonus given to the best performing providers, without a withhold or punishment at the other end [183, 205, 206].

P4Q in the USA is sometimes directly combined with explicit incentives aimed at productivity, utilization management and/or cost containment [69, 144, 171, 188, 212].

A minority of USA P4Q programmes provides an incentive based on the level of improvement through time instead of or combined with a fixed threshold level [181, 183, 184]. This longitudinal approach uses combinations of multiple thresholds for one indicator linked to the incentive amount [223]. For example, the bonus amount goes up gradually as a function of reaching a 20%, 40%…100% threshold.

Exception reporting is used rarely in the evaluated USA P4Q programmes. One study uses it as a goal for reduction in the context of clinical pathway use [171]. The concept and approach differs from the UK QOF approach of exception reporting.

The frequency of related quality measurements and the timing of the incentive provision vary. The most common frequencies include a yearly rate [170, 188, 195, 210, 222], a semi-annual rate [183, 184, 193], a four month interval rate [181, 182], a quarterly rate [74, 195], a weekly rate [177, 178]. Only a few authors describe the degree of time delays between healthcare provision, data collection, feedback (if present) and incentive payment. This can vary from no delay up to six months of delay [74].

Feedback and P4Q payment do not always occur within the same time frame. It is for example possible to provide monthly feedback and award an incentive each six months [193].

Most authors do not discuss the differential weighting of targets to be met, with some exceptions [197]. In most cases this implies the use of an equal weighting system, although this is not always clear.

Although sometimes stated in early P4Q literature, only a few authors indicate that the proprietary nature of the P4Q schedule prohibits a detailed description [170].

4.2.4.3 Australia, Germany, Italy, and Spain

One P4Q programme studied in Australia is based on bonus payment varying according to case mix adjusted throughput, made to the hospital at the beginning of the financial year. The total bonus pool equals $7.2 million per year, and increased to $17 million for the last year. Bonus funds are absorbed within the general revenue of the hospital, with some funds directed to improving the emergency department and resources for bed management. Bonus reductions are applied if targets are not met in terms of percentages (1%, 5%, 20%, etc) [213]. The characteristics of the P4Q intervention are not specified in the other included Australian study [214].

The German study makes use of direct physician payments for every patient not smoking 12 months after recruitment. A physician receives $130 after study completion for each study participant they recruited who was proved smoke free. The time frame is not specified [216].

In Italy, one study tested the use of penalties according to the breastfeeding rate in a hospital setting. A deduction of 0,5% of the DRG annual revenues was applied, aimed at the intermediate regional health authority [217].

Finally, in Spanish Catalonia target payments were used with a maximum - per annum- of 5 200€ per physician and €6 200 per Primary Care Team manager. In addition, a professional development scheme was implemented and evaluated as part of Human Resource Management. Positive assessments implied an additional annual increase in individual salaries of €3 000 (physician) and €1 400 (nurses) [218].

Key points on P4Q as intervention

United Kingdom

- **The Quality and Outcomes framework in the UK links financial incentives to quality of care that is provided by primary care practices. Quality is measured against a set of 146 quality indicators relating to clinical care, organization care and patient experience. These indicators are linked to a number of quality points. In general, more points are assigned at intermediate outcomes than at process measures. For clinical indicators, practices claim points, which generate payments according to the proportion of patients for whom they achieved each target. Up to 25% of the family practitioners income is determined by the QOF programme. Data on quality of care from each practice is being automatically extracted from the clinical computing systems of the practices. To protect patients from inappropriate care, the programme allows practices to exclude patients who they deem inappropriate from specific indicators.**
- **Recently, a Pay for Quality demonstration programme in hospital care, named 'Advancing Quality' has been launched. Incentives are related to three types of target groups, namely evidence based process measures, patients' quality of life after surgery and patients' experience of provided care.**

USA

- **The characteristics concerning P4Q intervention in the USA are very diverse. Most USA programmes make use of thresholds to be met. The indicators used, are mostly structural, process and/or outcome measures. The incentive determines a certain percentage of the providers' or the organizations' income, this percentage ranges between 0% a 12%. The base for the incentive calculation and the target level vary between programmes. Many P4Q programmes make use of a competitive bonus.**

Australia, Germany, Italy and Spain

- **In Australia one P4Q programme is based on bonus payment. If targets are not met, bonus reductions are applied.**
- **The German P4Q programme makes use of a direct physician payment for every patient not smoking 12 months after recruitment.**
- **The Italian P4Q programme concerning breastfeeding makes use of penalties.**
- **One Spanish P4Q programme makes use of incentives for physicians, nurses and team managers.**

4.2.5 Implementing and communicating P4Q programmes

4.2.5.1 United Kingdom

A previous attempt to provide financial incentives for high-quality care in the mid-1980s, named 'Good Practice Allowance', was rejected by the general practitioners profession. At that time, there was little acceptance of the finding that wide variations existed in the practice of medicine, or that such variation might have a negative effect on patient care [55].

Part of the reason that doctors rejected the Good Practice Allowance in the mid-1980s was that they rejected the assumption that quality could be measured. In addition, there was a substantial degree of professional protectionism that took the form of denying the existence of poor practice. During the 1990s evidence-based medicine was introduced. Health professionals gradually accepted that there were differences in performance in the delivery of healthcare and that there were justifiable limits to individual freedom in the clinical setting. This was also the decade when researchers in health care demonstrated that there were widespread variations in healthcare delivery and that many patients were receiving care that fell short of what was considered good practice.

The effect of these developments was that it became increasingly possible both to define high-quality care and to provide methods that could be used to measure some aspects of the quality of care. The change in the culture of the profession that occurred during this decade in the UK was enormous [55]. The last piece missing was the political will to implement the proposed changes. To tie a substantial proportion of physicians' income to the quality of the care would produce winners and losers. However, the British Medical Association was unwilling to accept a change in remuneration that would result in the loss of income for large numbers of its members. In 2000 the government of the United Kingdom decided to provide a substantial increase in health care funding. This change came as a result of recognition of the serious underfunding of health care in the United Kingdom as compared with that of similar countries, and the publication of data suggesting poor outcomes for health care in the country.

Negotiations took place between the British Medical Association and the NHS, with the assistance of a small group of academic advisers. The academic advisers drew on published national guidelines for sources of evidence, such as those from specialist societies. The details of the new contract were voted on by general practitioners in June 2003. There was a 70 percent turnout, and 79 percent voted in favour.

Although this description of the implementation and communication process is relevant to the interpretation of the included UK studies, most authors do not specify the context of implementation and communication.

4.2.5.2 United States of America

Similar to all other characteristics and the P4Q intervention design, the process of implementation and communication is also very heterogeneous in the USA. Like in the UK, many USA studies do not report on this subject. The history of implementation, especially the level of involvement and support by healthcare providers, is rarely described. The studies that do report on this can be grouped into low and high implementation and communication support for P4Q programmes. Support is provided in one or more of the following ways:

- Involvement of healthcare providers/peers throughout the P4Q programme design process by regular meetings [189, 197, 205, 206].
- Involvement of healthcare providers/peers to review and approve evidence based guidelines as a basis for the P4Q programme [184], use of a multidisciplinary task force to identify evidence based practice and usable indicators [171, 189, 193].
- Involvement of healthcare providers/peers in the continuous updating process of the programme [205, 206].
- Involvement of patients, office staff, educators, etc. [190].
- Clear leadership support [175, 197], coordination by central leadership [190], directors and administrators invited to a project launch meeting [196].
- Support of similar involvement in the care improvement activities [190].
- Use of a small core work group and later refinement by the larger development team [190].
- The contact strategy:
    - Feedback by mail or in person by a medical director [188], the use of multiple individual mailings [183], the use of letters and a brochure [184], notification in monthly bulletin [197, 201, 206, 220].
    - The use of formal presentations at standing meetings [197].
    - Face to face feedback [210], personal communication [201], visit by a project representative offered to each intervention site [196].
    - Annual manual provided with an executive summary, full description of the programme, answers to commonly asked questions [197, 206].
    - Provision of technical specifications for each measure [197], transparency of the methods used to determine performance rates and translation into a financial incentive amount via a public website [206].

- The use of interviews or surveys to check awareness of the programme and to assess acceptability [178, 180, 184, 195], to assess satisfaction [206].
- The use of interviews or surveys to check measurement activities and how the data are used to maintain or improve the quality of care [210], reverse feedback about local guideline implementation efforts [197].
- The use of interviews or surveys to identify barriers impeding target attainment [182].
- The use of focus groups for similar purposes [222].
- The use of constant reminders of measures and targets, directly integrated into daily practice [222].
- The presentation of further improvement recommendations towards the participants [210], the use of suggestions on how to spend earned incentive payments [201].
- Tools and resources provided to support quality improvement, directed at directors and quality improvement teams [175].

Whereas the level of awareness of the programme was acceptable in some studies [180], a low level was found in other studies [183, 184]. One study reports a general misunderstanding by participants of contra indications which could be used to exclude patients [181]. Another study identified an initial reaction of suspicion and distrust, disbelieve, being surprised by low baseline scores, a confrontation with poor documentation, which stimulated protocol development, outreach calls, enrolment in disease management programmes, new monitoring and follow-up assignments to staff, and the beginning of utilization of other not incentivized checklists [222].

A number of studies report that P4Q programme participation was of a voluntary nature [44, 69, 174, 175, 176, 195, 205, 206, 211]. Other studies do not specify the mandatory versus voluntary design option. On the basis of the complete programme reporting it can sometimes be assumed that participation was obligatory. However, this assumption should be treated with caution when reviewing results.

4.2.5.3 Australia, Germany, Italy, and Spain

One Australian study states that the attitudes among participants have generally been positive, with high degrees of compliance and cooperation [213]. The second Australian study makes use of a survey to identify barriers and reasons for not accomplishing targets [214].

No information on implementation and communication was provided in the German study [216].

In the Italian study, work plans and targets were set by regional health authorities [217], without further specifications on implementation and communication.

The Spanish study surveyed the level of job satisfaction of participants and their quality of professional life, using standardized instruments [218]. The results are included in the effects reporting section 4.3.1 (see page 54).

Key points on implementation and communication

United Kingdom

- **A first attempt to implement a P4Q programme in the mid-1980's, was rejected by the professionals, partly because physicians rejected the assumption that quality could be measured. With the introduction of evidence based medicine, the health culture changed.**
- **In a second attempt to implement a P4Q programme, which leaded to the implementation of QOF in 2004, the government decided to provide a substantial increase in health care funding and negotiations took place between government and professionals with the assistance of a small group of academic drivers. Participation in the QOF programme is voluntary.**

USA

- **Implementation and communication support is provided in one or more ways, depending on the programme. They include mainly involvement of providers/peers, involvement of patients, office staff and educators, leadership support or the use of a small core work group. In addition there has been made use of interviews or surveys, focus groups, reminders and presentations to communicate and implement the programme. The level of awareness of the programme varied between programmes as well as the acceptance. Participation was generally voluntary, although it can be assumed that some programmes are obligatory.**

Australia, Germany, Italy and Spain

- **Communication and implementation differed according to the P4Q programme. In some countries a multidisciplinary committee developed the programme, in other countries the targets and work plans were set by the (local) government.**

4.2.6 Existing and concurrent quality improvement initiatives

Existing and concurrent quality improvement initiatives, both on a local and/or system's level, have an influence on the primary outcome measures as indicators of quality of care. Therefore, when reviewing the effects of P4Q these trends and co interventions should be taken into account.

4.2.6.1 United Kingdom

Generally, a trend of quality improvement can be identified in primary care in the UK. This trend was already set into motion before the introduction of the QOF incentive programme in 2004. A number of system wide quality improvement initiatives introduced before and during the UK P4Q interventions have contributed to positive effects in quality performance. The following main categories can be distinguished:

- Changes in standard medical practice in terms of clinical diagnosis and treatment. Examples are the acceleration of the immunization schedule (at a younger age) [132], the introduction of child health surveillance [125], a trend of increasing use of effective blood pressure and cholesterol lowering therapy [151], the introduction of nicotine addiction treatment on prescription [145], and the introduction of a more active screening for cardiovascular risks in diabetes patients [141]. In more general terms, national guidelines and service frameworks have been implemented for hypertension, angina, cardiac rehabilitation, stroke, etc. [126, 143, 145, 167].
- Structural changes in the system of healthcare provision:
    - Changes in setting/infrastructure: the introduction of nurse led primary care diabetes clinics [128], the introduction of a network of stop smoking services [145], service redesign such as managed clinical networks for stroke and coronary heart disease (CHD) [126].

- Changes in staffing policy: the introduction of 'health visitors' and local immunization coordinators [132], a higher recruitment of nursing and administrative staff to support primary care delivery, and a higher number of general practitioners [148].
- Changes in IT support: the introduction of an immunization recall system to the provider [132], the use of clinical prompts [163]. A system of uniform electronic patient records with automatic data extraction was introduced as a relevant factor in all QOF evaluation studies [148].
- Together with the QOF came the introduction of two billion pounds of additional funding [148]. This budget enlargement has to be taken into account.
- A final category consists of care management processes used as tools to support quality improvement. Here we identify educational strategies [132] and feedback and public reporting [148 , 163]. In the UK, feedback and peer comparison of performance is considered part of the P4Q intervention. Public reporting is also applicable for all QOF evaluation studies.

The interrelationship between P4Q as intervention, these other concurrent factors and the reported quality performance effects is not further reported upon in most UK studies. Some trends and co interventions were already in effect before P4Q implementation, others are introduced simultaneously. This obscures the direction of effects. Another related issue concerns the improvement in documentation [128, 165]. It might be that P4Q has led to better documentation of care instead of real quality improvement. It might also be that both better documentation and quality improvement occurred, with mutual reinforcement. This distinction is often unclear, because quality measurement makes use of available documentation in patient records.

4.2.6.2 United States of America

Because of the diversity in the USA, trends and co interventions are more difficult to generalize towards the USA as a whole. In the American P4Q studies, general trends in standard medical practice and the adoption of new scientific evidence into practice is less identified. One example, similar to the UK, is the introduction of an extended schedule of public influenza immunization clinics [178]. In addition, some broader quality improvement initiatives are described within the context of P4Q studies, such as the vaccines for children programmes [181 , 184]. Authors also point out the influence of extensive media campaigns [178] and of general publicity attending P4Q [172].

Structural changes in terms of settings/infrastructure and staffing policy are less described than in UK studies. However, changes in IT support are often reported. An example is the use of patient registries [144 , 191 , 222], which can be part of a fully electronic health record [193] and be centralized [201]. In another study, implementable best practices are integrated in IT [189]. New supportive IT can be applied for coordination and follow up [223].

The use of reminders is a related co intervention [178 , 190 , 193 , 195]. In one USA study a significant inverse relationship was found with the use of reminder postcards: lower achievers on performance measurement sent more cards [177]. Another study found no significant relationship with reminder system use (phone, postcard, flowchart in record) [178].

Finally, the growth in the use of electronic patient records is identified as a relevant factor [172].

One study reported the introduction of gate keeping with prior authorization as a relevant change on the market/payer level [173].

On the level of the provider a very diverse set of care management processes are reported as tools co supporting quality improvement in many P4Q studies.

Firstly, there are similar processes as in the UK: educational strategies, feedback and public reporting. The use of new educational approaches is often reported as a co intervention [171 , 174 , 183 , 190 , 195 , 222 , 223].

Feedback and peer review are often present [174, 181, 183, 184, 188, 190, 191, 193, 205, 206, 210, 222, 223]. However, not all USA studies include the use of feedback. Some provide only a financial incentive as such. Feedback is in most cases provided to the target of the P4Q incentive, but can also be limited to, for example, the clinical administrator level [196].

Public reporting is also a common co intervention of P4Q in the USA [44, 69, 74, 175, 176, 203, 210, 211]. In some studies, public reporting was positively associated with P4Q target performance with odds ratios varying between 1.3 and 1.9 [199, 202]. One author reports the likely effect of anticipation of public reporting [172]. Positive recognition, with an awards dinner or the use of a clinical champion model is also sometimes used concurrently with P4Q [197, 220, 223].

Secondly, a long list of additional tools can be formulated: guideline distribution [201, 222], the introduction of treatment algorithms [174], care pathway development [171], guideline integration into the workflow [222], the use of exception profiling with actionable recommendations [171] and the introduction of a chart review and assessment tool [222]. Some co interventions are more imbedded into clinical practice such as the use of a risk stratification tool [174], the use of a tobacco use patient assessment form [220], introduction of smoking cessation brief counselling [220], the use of a proactive telephone support system for smoking cessation [201]. Another group of co interventions is more focused on patient engagement. These include the introduction of a mechanism for patient engagement in commitment to best practices, patient incentive use, tracking of patient compliance, and direct patient outreach [190]. A final group of co interventions focuses on sharing of information and learning from each other: the promotion of information exchange between clinics [210], and active transmission of best practices at monthly meetings of medical and executive directors [191].

The same documentation improvement effect as reported in some UK studies, is described more extensively in USA studies [180, 181, 182, 197].

4.2.6.3 Australia, Germany, Italy, and Spain

In the Australian studies, one author reports on the likely influence of a period of industrial action (strike) of nursing and ancillary staff as a temporary trend phenomenon [213]. The other study has investigated the use of parent incentives as co intervention [214].

The German study has made use of the introduction of two hour physician group training in smoking cessation methods and the direct patient reimbursement for pharmacy costs associated with nicotine replacement therapy or bupropion therapy [216].

In the Italian study, the use of a broad stimulation programme and a surveillance system with feedback is reported [217].

Finally, in the Spanish study a voluntary professional development scheme was introduced as co intervention [218].

Key points on existing and concurrent quality improvement initiatives

United Kingdom

- **Several trends and co interventions were already in effect before or are introduced together with P4Q implementation, namely changes in medical practices in terms of diagnosis and treatment, structural changes in the healthcare system, additional funding, the use of care management processes to support the programme and the improvement of documentation.**

USA

- **Because of the diversity in the USA, trends and co interventions are difficult to generalize. The most important ones seem to be changes in IT support, the use of reminders, educational strategies, feedback, public reporting and sharing information.**

Australia, Germany, Italy and Spain

- **A small number of trends and co-interventions have been identified, namely the influence of a period of industrial action, the use of parental incentives, physicians' training, a stimulation programme and surveillance system and finally a voluntary professional development scheme.**

4.3 GENERAL EVIDENCE

4.3.1 Reported effect of Pay for Quality programmes

In the following paragraph we summarize the results of the literature study for preventive care, acute care and chronic care. According to study design (see description on page 35) and effect, the following labels are used to grade evidence:

- Strong evidence: a strong design (randomized studies; concurrent + historical comparison studies) with a clear effect;
- Weak evidence: a weak design (Concurrent comparison studies; Historical comparison studies, multiple time points; Historical comparison studies, before-after time point; cross-sectional studies) with a clear effect;
- Conflicting evidence: a significant effect and no significant effect within one design or within a group of weak or strong designs.

Below, a summery of the most important evidence is been given. A more detailed description of the result of the literature study, together with the matching evidence table is provided in Appendix 16 A and 16 B. As will become clear most evidence applies for primary care. The following labels will be used to indicate whether the evidence applies to primary care or hospital care:

- H= hospital care
- P= primary care
- H & P= both hospital care and primary care

Within this section it must be noted that a lot of the information comes from QOF studies. These studies are not randomized. The introduction of the QOF scheme which was implemented for all GP's in the whole country at the same time, didn't allow setting up randomized studies. Nevertheless these studies are of great value. In addition, one could question if randomization is really needed in studies were a universal P4Q scheme is implemented and hence selection problems do not occur.

4.3.1.1 Preventive care

There is strong evidence for a positive P4Q effect on influenza immunization, with an effect size of 6 to 8% [177, 178] (H&P)and weak evidence for a positive P4Q effect on cholesterol screening in adults (P), with an effect size of 3% [188].

For well child visits[c] there is strong evidence for an effect ranging from a negative effect to a 5% effect, depending on the age-group [172, 184] (P). For Cancer preventive screening a 4% improvement was found [69, 74, 172, 183, 195] (H&P). A similar but wider range was found for children immunization (0 - 25%) [181, 182, 184, 206] (P) and for children preventive screening (0 – 29%) [180, 184] (P). It should be noted that the 29% effect was found in a study with a weaker design.

In the prevention of sexually transmitted diseases there was no effect found based on weak evidence and even a negative difference of 11% compared to a matched comparison group based on strong evidence [172, 185] (P). The latter finding concerns Chlamydia screening.

There is almost no evidence on P4Q effects in preventive care on other quality domains than clinical effectiveness.

---

[c] For your information: the equivalent of well child visits in Belgium are the preventive paediatric consultations organized by Kind & Gezin and ONE.

4.3.1.2 Acute care

Concerning emergency care there is weak evidence of a positive effect on timeliness of care [213] (H). The effect on timeliness had a range of no effect to a 10% effect depending on the urgency of care. There was no effect on emergency department – hospital admission patient flow [213] (H). Finally, there is weak evidence for a positive relationship with the smoking cessation referral rate in the emergency department [220] (H).

For myocardial infarction/acute cardiac events there is strong evidence of a lack of effect on the timely administration of thrombolytic agents and of a 5.4% positive effect on timely percutaneous coronary intervention [175] (H). For most other targets P4Q effects ranged between zero and 3% [44, 174, 175, 223] (H). The upper limit of this range was higher for prescribing aspirin at discharge (8.5%) and ACE inhibitor use (9.9%). No effect was found on not incentivized control measures [174, 223] (H).

For coronary artery bypass grafting (CABG) patients there is weak evidence on the absence of effect on various long term outcomes [189] (H). Only a 10% positive effect on the discharge to home ratio was found [189] (H).

There is strong evidence of a 25.5% positive effect on the provision of discharge instructions to heart failure patients [175] (H). There is no effect on ACE inhibitor use and smoking cessation advice [44, 175] (H). For left ventricular failure (LVF) assessment the effect ranges from minus 2.4 to plus 5.1% [44, 175, 223] (H).

With regard to community acquired pneumonia there is strong evidence of a positive effect on pneumococcal screening and/or vaccination (9.5-44.7%) and for blood culture testing (3.5%) [44, 175, 223] (H). For oxygenation assessment and timely antibiotics administration the effect ranged from minus 1.9 and minus 3.2% to zero and plus 4.3% [44, 175, 223] (H).

There is weak conflicting evidence concerning three targets focusing on replacing inappropriate care by evidence based alternatives in the treatment of acute sinusitis (-29%- 14%) [171] (P&H).

With regard to breastfeeding rate there is weak evidence of a positive effect of P4Q with a range between 6 and 12% [217] (H). This was mainly due to a shift from partial to full breastfeeding and not by a higher patient rate beginning with breastfeeding.

4.3.1.3 Chronic care results

Firstly, concerning diabetes for a limited number of targets there is strong evidence of a positive effect, but with a below 5% effect size: weight recording, smoking status recording, and peripheral pulses testing [143] (P) Secondly, another group of targets showed no effect: hypoglycaemic symptoms recording and glycaemia control rate [143, 219] (P). It should be noted that the last target finding is based on weak evidence only**.** In addition, for blood pressure recording and influenza immunization there is strong evidence of no effect, however some studies with a weaker design came to positive effects [127, 143, 152, 187, 219] (P). Thirdly, a wider effect range, between 0 and 25% is supported by strong evidence for HbA1c testing rate, lipid and cholesterol testing rate, nephropathy testing rate and retinal exam rate [69, 143, 172, 187, 190, 191, 222] (H&P). Fourthly, for some targets there is strong evidence of a positive effect with a smaller range: HbA1c intermediate outcomes (0-14%), cholesterol outcome (positive up to 23.5%), blood pressure outcome (1.6-6.3%) and foot exam rate (2.7-45%) [74, 143, 222] (P). For a number of other targets only weak evidence was available. Finally, there is weak evidence of a positive effect of P4Q on the referral rate for poor glycaemia control (23% effect size) as a care coordination measure [161](P).

There is weak evidence for a positive effect on heart failure treatment in primary care: a 23.4% increase in ACE inhibitor use [205](P).

With regard to coronary heart disease there is strong evidence of an absent effect on angina attack recording, exercise capacity recording, weight advice, blood pressure outcome and cholesterol outcome [143](P). In addition, the following targets showed below 5% improvements: exercise stress testing referral, dietary advice, aspirin prescription [143](P). For three targets the range of effect size was larger, however with weaker design studies reporting the maximal effects: blood pressure recording (0.7-21.5%), cholesterol recording (-10.8-41.7%), and smoking status recording (2.39-26.2%) [127, 143] (P). For smoking advice, antiplatelet therapy, ACE inhibitor use and influenza vaccination there is only weak evidence available for positive effects [127] (P). The evidence on the effects on the QOF's additional service domain is mixed [149] (P).

For stroke care only weak evidence is available. The reported effects are large for eight targets, ranging from 17 to 52.1%, and absent for two targets (blood pressure outcome, cholesterol outcome) [134] (P).

With regard to asthma care there is limited evidence available. On some targets there is strong evidence of the absence of an effect (asthma controller use, recording of peak expiratory flow) [191](H&P). On others the effect remains below 5% (smoker status recording, inhaler technique recording, recording of daily, nocturnal or activity-limited symptoms) [143](P). There is weak evidence of a positive relationship with care management processes use and strong evidence of no relationship with not incentivized measures [143, 186] (P).

On hypertension care there is only weak evidence that P4Q led to an increase of 12% of targets achieved [163](P).

There is strong evidence that P4Q has no effect on smoking abstinence and on the smoking cessation advice rate, although a weaker design study found a 21% positive effect on the latter target [197, 201, 216](P). However, there is a positive effect on smoker status recording (7.9-24%) and on the referral rate (6.2%) [196, 197, 201](P).

There is a lack of evidence on the P4Q influence on depression/mental illness primary care. One weak design study found no effect on the percentage of indicators achieved [163](P).

Concerning chronic child care there is only weak evidence available of a positive relationship with receiving outpatient specialty care [173](P).

There is a lack of evidence on chronic obstructive pulmonary disease (COPD) care. Weak conflicting evidence is found concerning the relationship between emergency admission rate and performance on not incentivized measures [149] (P).

With regard to epilepsy there is weak evidence of a small positive relationship between the seizure free rate and the proportion of emergency hospitalisations, which is not further explained [159](P). The impact of P4Q is however unclear.

With regard to chronic kidney disease one weak design study showed a positive effect on the number of visits, Kt/V rate, ultra filtration volume, albumin rate, haemoglobin rate, phosphorus rate, calcium rate, catheter use, and the number of skipped treatments. No effect was found on Kt/V threshold achievement, shortening of treatments and hospital admission rate [179](H).

There is a lack of evidence on osteoarthritis care. One weak design study found an absence of effect on the percentage of targets achieved [162](P).

4.3.1.4 Generic findings

With regard to chronic disease management for multiple patient groups strong evidence is available on a positive relationship with preventive and chronic care targets and on the absent or positive relationship with not incentivized targets [153, 200, 206](P). Weak evidence indicates a positive relationship with the quality improvement initiative rate [200](P).

On the overall QOF achievement there is weak evidence on a 4.2% P4Q effect, with a marginal increase in drug prescription (0.69-1.09%) [126, 148](P). There is also weak evidence that the percentage of targets achieved was 16.9% higher for incentivized targets as compared to not incentivized targets [162](P).

With regard to care management processes (CMP) use, there is weak evidence of a positive effect on IT use (9-27%), and on clinical guideline use [194, 203] (H&P). Furthermore the evidence on CMP use is mixed.

Regarding patient and provider satisfaction there is a lack of evidence available. Weak evidence indicates no relationship with patient satisfaction, providers' quality of life and nurses' perceived demands [218](P). For physicians, a higher workload was reported. In addition, both physicians and nurses reported a management structure support. Intrinsic motivation showed no difference or a positive effect [218](P). There is weak evidence that P4Q based on patient satisfaction targets is positively related to perceived access to care, patient knowledge and use of preventive counselling.

There was no relationship with perceived continuity, integration of care, clinical interaction, interpersonal treatment and trust [204](P).

Next to the above reported effect of P4Q programmes, it is also important to take the variability-reduction or variability-increase into consideration. After all, a positive effect can be attended with a greater variability which is not desirable. However, few articles mention the variability.

Concerning preventive care, variability data were found for immunisation achievement and screening rate. For most of these indicators, there is a considerable decrease in variability after implementing P4Q programmes, only cervical cancer screening shows an increase in variability. For the immunisation rate there are conflicting results, although in most cases a variability reduction can be found [69, 178, 188, 214](H&P).

Concerning the indicators belonging to acute care, no article mentions the variability. Regarding chronic care, some articles about diabetes, asthma, hypertension en smoking cessation mention the variability. Most diabetes process indicators showed a huge decrease in variability. However some process indicators, like retinal screening, lipid testing, nephropathy testing, retinal testing, ACE inhibitor or all blocker use, and microalbumine recording showed an increase in variability or conflicting results. Variability of Intermediate outcome parameters also shows a mixed result. Achieving the blood pressure and HbA1c target showed a large decrease in variability. Achieving the cholesterol target shows a large increase in variability. The amount of asthma and hypertension indicators achieved, showed less variability when introducing P4Q programmes and all smoking cessation indicators showed no difference in variability of a decrease in variability after implementation of P4Q programmes.

However these decreases were rather small. Concerning generic findings, some QOF indicators and several patient and providers satisfaction indicators showed a decrease in variability, however the degree of decrease for the latter was very small.

Key points on the reported effect of Pay for Quality programmes

Preventive care results

- **There is evidence of a positive P4Q effect on influenza immunization and cholesterol screening. An effect ranging from no effect to a positive effect had been reported for cancer preventive screening, children immunization and children preventive screening. For well child visits an effect ranging from a negative effect to a positive effect was reported. Concerning the prevention of sexual transmitted diseased there is evidence of no effect or even a negative effect. For most of the indicators were variability data were reported, a decrease in variability was found.**

Acute care results

- **Concerning emergency care there is an effect ranging from no effect to a positive effect on timeliness of care and no effect on hospital admission patient flow. Finally a positive relationship with the smoking cessation referral rate was found.**
- **Regarding myocardial infarction/acute cardiac events, no effect or a positive effect was found on medication use and smoking cessation advice. On the not incentivized control measures, no effect was found.**
- **Concerning coronary artery bypass grafting, an absence of effect was found on almost all long term outcomes. A positive effect was found on the provision of discharge instructions to heart failure patients, no effect or conflicting results were for the other indicators.**
- **Regarding community acquired pneumonia, a positive effect was found for screening, vaccination and the use of blood cultures. For the other indicators conflicting results were found.**
- **There is a mixed effect on targets focusing on replacing inappropriate sinusitis care by evidence based alternatives.**
- **Concerning breastfeeding a shift was found from partial to full breastfeeding.**
- **No data on variability were found for the acute care indicators.**

Chronic care results

- **Concerning diabetes four groups can be distinguished. One group of targets showed no effect, another group showed a below 5% effect size, a third group showed a wider range effect, between 0 and 25% and a final group showed a positive effect with a smaller range. Wherever reported most diabetes indicators showed a decrease in variability.**
- **A positive effect was found on heart failure treatment in primary care.**
- **With regard to coronary heart diseases, several targets showed an absence of effect, others showed a below 5% improvement, for still others a larger range of effect size was found, with conflicting evidence for cholesterol recording.**
- **An absence of effect or a positive effect was found for the indicators related to stroke.**
- **Concerning asthma indicators, the effect size ranged between an absence of effect to a below 5% effect, a positive relationship was found with CMP use and no relationship with not incentivized measures. The variability diminished after implementing P4Q programmes.**
- **A positive effect was found on the hypertension care targets. Less variability was reporting after introducing P4Q programmes**
- **With regard to smoking cessation indicators the effect size ranged between no effect and a positive effect. After introducing P4Q programmes, variability showed no difference or a small decrease.**

- **Concerning depression/mental illness primary care indicators, osteoarthritis care indicators. Chronic child care indicators, epilepsy indicators and chronic kidney disease, a very limited amount of articles were found, with an effect ranging from an absence of effect to a positive effect.**

Generic findings

- **With regard to chronic disease management there is a positive relationship with several targets and an absent or positive relationship with not incentivized targets.**
- **On the overall QOF achievement there is a positive effect. Several QOF targets showed a decrease in variability after implementation of P4Q**
- **There effects with regard to CMP use are mixed.**
- **Regarding patient and provider satisfaction the effect ranged between no effect and a positive effect. Concerning variability a very small decrease was found after implementing P4Q.**

4.4 REPORTED COST EFFECTIVENESS AND MODELLING EFFECTS OF P4Q PROGRAMMES

4.4.1 Cost-effectiveness of Pay for Quality

Few reports have been identified concerning cost-effectiveness evaluations of Pay for Quality programmes. Only three studies focus on cost-effectiveness of pay for quality programmes, one in the UK [224] and two in the USA [107, 225].

Mason et al. (2008) [224] focus on the cost-effectiveness of the Quality Outcome Framework implemented in primary care and addressed at general practices. Twelve clinical QOF indicators have been considered that have a direct therapeutic impact (see Table 1). These included preventive, acute and chronic care targets concerning hypertension, heart diseases, heart failure, stroke, diabetes, kidney disease, cervical screening and smoking cessation. The QOF payments are based on point achievement, adjusted for practice size and disease prevalence relative to national average values. The annual per patient payment that is paid by the government, ranged from £0.13 to £87.79 in 2004-2005 and from £0.22 to £73.04 in 2006-2007. For the economic evaluation of this programme a threshold of £20 000 per QALY was assumed. According to this analysis, the most cost-effective indicators are the use of ACE inhibitors or Angiotensin Receptor Blocking (ARB) for chronic kidney disease, anticoagulant therapy for atrial fibrillation, and beta blockers for coronary heart disease. Note that the cost-effectiveness varies strongly by baseline uptake. In general, one can state that, when baseline implementation rates are high, larger absolute changes in utilisation are required for indicators to be cost-effective. High baseline utilisation rates imply that many patients already receive the treatment, hence general practitioners receive payments for patients already being treated in a correct way. In 2006, only one indicator was not cost-effective, namely diabetic retinal screening.

Table 1 : Cost-effectiveness evaluation of 12 QOF indicators by Mason et al. (2008)

Quality indicator	Incremental QALYs per patient treated (2006)	Incremental costs per patient treated (2006)	Net monetary benefit per patient treated (2006)*
AF3: treatment with anti-coagulant drug therapy or an anti-platelet drug therapy	1.465 to 2.2	-£1 162 to -£16 922	£45 162 to £46 222
BP5: hypertension, BP 150/90 in past 5 months	0.7	£751	£13 249
CHD9: aspirin	0.0066	-£30	£162
CHD 10: beta blocker	1.89	£234	£37 566
CHD 11: ACE inhibitor/ Angiotensin Receptor Blocker	0.08	£488	£1 112
ChKD4: ACE inhibitor/ Angiotensin Receptor Blocker	0.8076 to 1.5308	-£31 811 to -£32 906	£49 058 to £62 427
CS1: Cervical screening	0.137	£68	£2 672
DM15: proteinuria / microalbuminuria on ACE	0.7210	-£9 662	£24 081
DM21: diabetic retinal screening	0.4865	£9 750	-£21
LVD/HF3: ACE inhibitor/ Angiotensin Receptor Blocker	0.21	£25	£4175
Smoking 2: Smoking advice/referral	0.0157 to 0.0451	£11 to £90	£303 to £812
Stroke9/stroke12: antiplatelet/anticoagulant	0.17	£371	£3 029

* The net monetary benefit is calculated by assigning a value of £20,000 for each QALY gained.

Curtin et al. (2006) [225] focus on the cost-effectiveness of a pay for quality programme in primary care in the USA, directed at physician organizations, and focusing specifically on diabetic care. Between 2000 and 2004 a partnership was established between a Health Plan and a physician organization. The physician organization withholds 10 percent of claims payments to practitioners of which typically 70 percent to 100 percent is available for distribution at the end of the year. Additional dollars derived from gain-sharing programmes were added to the pool. Payment in the Pay for Quality programme varied from 50 to 150 percent, as a result from which an average full-time primary care physician could earn between $6 000 and $18 000 extra in a given year, based on performance. The return on investment calculation, included in this study makes use of a cost trend, to project costs and compare the projected cost with the actual cost. As a result the actual costs seem to be lower than the projected cost. As shown in Table 2, this resulted in a plan saving of $1 894 470 in year 1 (2003) and $2 923 760 in year 2 (2004). The annual system development costs were estimated at $1 150 000. As a result the return on investment estimate would be 1.6 to 2.5.

Table 2 : Return on investment evaluation of a diabetes quality improvement programme by Curtin et al. (2006)

	Plan savings diabetes care against two-year rolling trend	Total cost per year of the programme	Return on investment
Year 1	$1 894 470	$1 150 000	1.6:1
Year 2	$2 923 760	$1 150 000	2.5:1

Nahra et al. (2006) [107] evaluated the cost-effectiveness of hospital Pay for Quality incentives in the USA, focusing on heart care related hospital inpatient care. The pay for quality programme has been implemented in 85 hospitals and provides incentives to increase adherence to heart-care related clinical guidelines. Three process measures for heart care discharge have been considered, namely the percentage of eligible AMI patients receiving aspirin orders at discharge, the percentage of eligible AMI patients

receiving beta blocker prescription at discharge and the percentage of eligible congestive heart failure (CHF) patients receiving ACE inhibitor prescriptions at discharge. The Pay for Quality programme was implemented during a 4 year period. The incentives consisted of add-ons to the DRG reimbursement, with a maximum of 1.2% in 2000-2002, and 2% in 2003. For the interpretation of the economic evaluation a threshold of $50 000 per QALY was considered. Applying this threshold, the results seem to suggest cost-effectiveness, even in a worst case scenario. See Table 3.

Table 3 : Cost-effectiveness evaluation of a heart care related programme by Nahra et al. (2006)

Quality indicator	Discounted QALYs (lower bound-upper bound)	Incentive programme costs	Cost-effectiveness ratio
AMI patients receiving aspirin discharge orders	53.0 - 67.6	-	-
AMI patients receiving beta-blocker discharge orders	141.6 - 261.3	-	-
CHF patients receiving ACE inhibitor discharge orders	538.7 - 1372.4	-	-
Sum of indicators	733.3 - 1701.2	$22 059 383	$30 081/QALY - $12 967/QALY

4.4.2 Modelling costs or effectiveness

Five modelling articles have been identified, of which three studies are conducted in the UK [226, 227, 228], and two in the USA [229]. All these models aim at predicting either the short term financial consequences or the long term health consequences of P4Q programmes.

Kahn et al. (2006) [229] examine the hospital quality and financial performance under two Pay for Quality programmes in the US, namely the Premier Hospital Quality Demonstration programme with a duration of 3 years and the Medicare Payment Advisory Commission (MedPAC) pay for quality programme. Seventeen clinical quality measures concerning heart attack, heart failure and pneumonia are been taken into account. Under the premier hospital Quality incentive demonstration, hospitals can receive annually bonuses up to 2% (top ten percent performing hospitals). Penalties with a maximum of 2% (hospitals below the 10th percentile of performance) are only given in the third year and the penalty threshold is established in the first year of the programme. Within the MedPAC P4Q programme hospitals lose 1-2 % of there payment, to create a pool of funds that can be used to pay bonuses. In both the programmes only the top 20 percent of best performing hospitals receive a bonus. The MedPAC approach would redistribute $140 million in payments, the Premier approach would almost pay $10 million more than it collects through penalties. See Table 4.

Table 4 : Winners and losers by Hospital type and by Pay for Quality Program (millions of dollars) by Kahn et al. (2006)

Type of hospital	Premier pay for quality Scenario		Medicare Payment Advisory Commission (MedPAC) pay for quality scenario	
	Total bonus	Total penalty	Total bonus	Total penalty
All hospitals	$39.4	-$30.5	$139.8	-$139.8
Urban	$34.0	-$25.2	$117.2	-$119.8
Rural	$5.3	-$5.1	$21.6	-$18.8
Major teaching	$8.3	-$4.3	$26.3	-$21.9
Other teaching	$14.3	-$9.4	$50.1	-$51.6
Non teaching	$16.6	-$16.8	$62.4	-$65.1
Tax-exempt	$32.6	-$20.0	$114.8	-$105.6
Investor-owned	$3.0	-$6.3	$10.7	-$17.5
Public	$2.7	-$4.1	$14.2	-$16.6

Fleetcroft et al. (2006) [227] explore the link between financial incentives and the likely population health gains within the QOF system. In this study only 38 of the 91 clinical indicators are taken into account. As explained on page 43, the incentives for each indicator rely on a point system, whereby each quality indicator reflects a number of points. Points are then allocated to a GP for a certain indicator, with a related payment that starts above a minimum threshold level of achievement on the indicator. The clinical domain is 550 points (in 2004) out of 1050. The monetary value of one point is estimated at £120 for a general practice of average size.

The potential health gain on the 38 QOF indicators was estimated and expressed in number of lives saved, based on a study by McColl et al. (1998). These authors subdivided these 38 indicators in 8 interventions. As shown in Table 5, the potential health gain ranged from 2.8 lives saved per 100 000 people per year to 308 lives saved per 100 000 people per year. In addition, the potential payments in connection with the pay for quality programme ranges from no payment to £17 280 per year. In conclusion there seems to be no obvious relationship between payment and health gain for these 8 interventions. Some indicators could generate a large amount of health gain against a low payment within the QOF system, others only generate a minimal improvement in health gain against large payments within the QOF system.

Table 5: Relationship between potential health gains and potential payments within the QOF system by Fleetcroft et al. (2006)

Quality indicator	Maximum lives saved per 100000 people per year (% of total)	Maximum payment for a typical practice per year (% of total)
ACE in heart failure	308.0 (41%)	2 400 (6%)
Influenza immunization in over 65s	146.0 (20%)	3 600 (10%)
Smoking cessation advice and nicotine replacement	120.0 (16%)	10 440 (28%)
Screening and treatment of hypertension	71.0 (10%)	17 280 (46%)
Aspirin in ischemic heart disease	48.0 (6%)	1 320 (4%)
Warfarin in atrial fibrillation	33.0 (4%)	0 (0%)
Statins in ischemic heart disease	13.8 (2%)	2 760 (7%)
Statins in primary prevention	2.8 (0%)	0 (0%)

McElduff et al. (2004) [228] estimated the health gain within cardiovascular patients if a number of QOF quality measures were to be met. Five interventions were taken into account, namely the use of aspirin, a cholesterol lowering treatment, a hypertension management, a treatment with ACE or angiotensin 2 (A2) inhibitors and influenza immunization. The modelling method used incorporates data on clinical effectiveness and the baseline rate (current rate) of performance concerning these interventions. Consequently, the comparator in this study was current treatment. The health gain in cardiovascular patients is expressed in number of cardiovascular events prevented per 10 000 patients, among which is understood, angina pectoris, myocardial infarction, death from coronary heart diseases (CHD), stroke, congestive heart failure, peripheral vascular disease and death from cardiovascular disease. As shown in Table 6, reaching the cholesterol target (lowering the total cholesterol in patients with values above 5.0mmol/l) will result into a reduction of cardiovascular events among patients with CHD, stroke and diabetes with respectively 15.5, 7.2 and 6.5 cases per 10 000 over a 5 year period. Reaching the targets concerning hypertension management, will prevent cardiovascular events with respectively 3.6, 2.9 and 2.9 cases per 10 000 over a 5 year period for patients with CHD, stroke and diabetes. In addition 15.5 events will be prevented by meeting these targets in other patients (no stroke, no CHD, no diabetes). With regard to the targets relating to aspirin, ACE inhibitors/A2 antagonists and influenza immunization, achieving these targets will only prevent a small number of events, either due to an already widely spread use of these guidelines (use of aspirin, ACE inhibitor, A2 antagonist), or due to a low baseline risk of death because of a currently already high compliance with the indicator (influenza).

Table 6: Potential health gain per indicator within the QOF system by McElduff et al. (2004)

Quality intervention	Clinical domain (disease)	Number of CVD events prevented over 5 year period
Cholesterol lowering treatment	CHD	15.5
	Stroke	7.2
	Diabetes	6.5
Blood pressure lowering treatment	CHD	3.6
	Stroke	2.9
	Diabetes	2.9
	Hypertension	15.5
Aspirin	CHD	1.1
	Stroke	0.4
ACE inhibitor /A2 antagonist	CHD and HF	1.2
Influenza vaccination	CVD	0.003

Fleetcroft et al. (2008) [226], estimate the potential population health gain of the full implementation of the 8 clinical interventions in both the original and the revised QOF contract. The population health gain is represented by number of lives saved per 100 000 people per year. This research identified evidence for lives saved on 22 indicators in the original contract and 19 indicators in the revised one. The potential of lives saved in the original contract was 415.77 lives per 100 000 in one year. For the revised contract this number raised with 35.73 lives to a potential number of 451.5. In addition, it is important to point out that the comparator in this study was "doing nothing", whereas there was already a significant baseline activity in primary care before the implementation of the QOF contract. Hence, it should be emphasized that the resulting figures represent a maximum potential. As shown in the table below, influenza immunization (in contrast with the study by McElduff et al (2004) [228]) and primary prevention for hypertension carry the greatest potential for lives saved. It must be noted that influenza immunization was already incentivised before the QOF was introduced, therefore the room for improvement is rather small.

Table 7 : Potential health gain per QOF indicator by Fleetcroft et al. (2008)

Clinical domain (disease)	Quality indicator	Potential lives saved per 100000 population per year (2003)	Potential lives saved per 100000 populations per year by clinical domain	Potential lives saved per 100000 population per year (2006)	Potential lives saved per 100000 populations per year by clinical domain
Atrial fibrillation	AF3: anticoagulant	-	-	21.4	21.4
Asthma	Asthma5: smoking cessation advice/referral	8.8	8.8	See smoking2	-
Hypertension	BP3: smoking cessation advice/referral	5.4	53.6	See smoking2	48.2
	BP5: hypertension, BP, 150/90 in past 9 months	48.2		48.2	
Coronary heart disease	CHD4: smoking cessation advice/referral	2.4	163.2	See smoking2	160.9
	CHD6: BP<150/90	11.3		11.3	
	CHD8: cholesterol < 5mmol	15.8		15.8	
	CHD9: Aspirin	24.8		24.8	
	CHD10: beta blocker	45.9		45.9	
	CHD11: ACE/ARB	1.5		1.5	
	CHD12: influenza immunisation	61.6		61.6	
Chronic	ChKD3: BP<140/85	-	-	26.2	26.2

Kidney disease					
Chronic obstructive airways disease	COPD5: Smoking cessation/referral	2.6	27.6	See smoking2	25.0
	COPD8: influenza immunization	25.0		25.0	
Diabetes	DM4: smoking cessation advice/referral	2.4	109.5	See smoking2	107.1
	DM6/DM20: HbA1c<7.4	26.5		26.5	
	DM7: HbA1c<10	7.4		7.4	
	DM12: BP<145/85	13.5		13.5	
	DM15: Proteinuria/microalbuminuria on ACE	3.4		3.4	
	DM18: influenza immunization	63.7		63.7	
Heart failure	LVD/HF3: ACE/ARB	11.6	11.6	11.6	11.6
Smoking	Smoking2: smoking cessation advice/referral	In domains	Included in other domains	10.9	10.9
Stroke	Stroke4: smoking cessation advice referral	1.1	44.9	See smoking2	43.8
	Stroke9/Stroke12: antiplatelet/anticoagulant	15.8		15.8	
	Stroke10: Influenza immunization	28.1		28.1	

Averill et al. (2006) [95] estimate the potential decrease in Medicare payments due to the reduction of post-admission complications in hospital care in a new pay for quality hospital programme in the USA. The Medicare Payment Advisory Commission (MedPAC) recommended an adjustment of the current Medicare Diagnosis-Related Group (DRG) based Inpatient Prospective Payment System (IPPS) under which complications after admission of a patient are being remunerated. Hence when a complication occurs, payment is being increased and thus poor quality of care is being rewarded. The redesign of this DRG system is proposed to reduce those payments due to post-admission complications and requires that present on admission indicators (which specify whether the diagnosis was present at the time of admission) are available for all diagnoses. No explicit financial incentive or penalty is being given, but the system could result in a lower income for hospitals with a large amount of complications in comparison to the current way of remuneration. Currently, hospitals only report discharge diagnoses. Only hospitals in California and New York are required to provide a present on admission indicator for each diagnosis. Data from California are extrapolated to all of Medicare. This results in Medicare payment reductions to hospitals between 0% to 3.29%, with 34.35% of hospitals having a payment reduction below 0.5%, 49.21% having a reduction between 0.5% and 1.5% and 16.43% having a payment reduction between 1.5% and 3.29%. Nationally, the overall reduction amounts to 1.01% of Medicare DRG hospital payments, which corresponds with a reduction in payments of $1.005 billion.

Key points on reported cost effectiveness and modelling effects of P4Q programmes

Cost effectiveness

- **One study in the UK and two in the USA focus on cost-effectiveness of P4Q programmes.**
- **From the twelve QOF-indicators, investigated in the UK study, only one, diabetes retinal screening seemed to be not cost-effective.**
- **The USA study focussing on P4Q programmes in primary care (diabetic care) showed a positive return on investment ranging between 1.6 and 2.5 per invested US dollar.**
- **The third study evaluated the cost-effectiveness of hospital P4Q programmes focussing on heart care. Applying a $50 000 per QALY threshold, the programme seems to be cost-effective even in a worst case scenario.**

Modelling costs

- **Three UK studies, and two USA studies predicted short term financial consequences or long term health consequences of P4Q programmes.**
- **Concerning the long term financial consequences, one USA study has investigated two P4Q programmes. One programme is break-even, it collects money in a pool of funds that is used to pay bonuses afterwards. The other programme redistributes almost $10 million more than it collects.**
- **Another US programme estimated the potential decrease in payments due to the reduction of post-admission complications in hospital care. The implementation of such a P4Q programme could result in a payment reduction to hospitals between 0% and 3.29%. The overall reduction equals to 1.01% of hospital payments, with corresponds with a reduction in payments of $1.005 billion.**
- **One UK study explored the link between financial incentives and likely population health gain within the QOF system. As a result there seems to be no obvious relationship between payment and health gain for these 8 interventions.**
- **Finally two UK studies estimated the health gain if a number of QOF quality measures were to be met. The first study uses the baseline activity as comparator, the second study uses no activity as comparator. In the first study reaching the cholesterol target and the targets concerning hypertension management were the most effective in preventing CHD events in cardiovascular patients. In the second study, influenza immunization and primary prevention for hypertension carry the greatest potential for lives saved, although is must be noted that influenza immunization already reaches high achievements. Hence, taking the baseline activities into account, the results of this study are an overestimation of potential life gain.**

4.5 EVIDENCE RELATED TO EQUITY

The initial aim of this paragraph was to assess the impact of P4Q initiatives on equity in access, treatment and outcomes. An important first step was to assess how equity is conceptualized in the selected studies (to what extent did the selected papers address the three domains of equity and did they conceptualize equity as vertical or horizontal or both?). Since the selected publications do not include information on access, the following paragraph will mainly focus on equity in treatment and treatment outcomes.

Out of the 32 selected publications, 27 address the relatively new Pay for Quality initiative in the UK. Sixteen of them use the QOF database to do this. The other ten studies make use of alternative databases. This relatively large number of studies on the UK initiative should make it possible to get a broad and deep insight on its impact on equity. Since the 5 other publications each cover another P4Q initiative, we do not get enough in-depth insight in these initiatives to make conclusions on their impact on equity. Therefore the focus of this paragraph will be on QOF.

Inequity in health care occurs when systematically differences are found in access, treatment or outcomes for people from certain subgroups in society. The assessment of equity for socioeconomic groups receives by far-out most attention in the selected studies (in 22 of the 26 selected studies on pay-for-performance in the UK). Socioeconomic status is hereby defined using the area-level Index of Multiple Deprivation or the DEPCAT score (supra). A smaller number of studies (11) also address equity in health care for patients with a different ethnic background. Finally, a few studies also look at gender differences (6 studies) or differences in age groups (7 studies).

4.5.1 Reported evidence on equity

In this part of the chapter we summarize the available evidence on the effect of the introduction of the QOF in the UK on equity in treatment and (intermediate) treatment outcomes. We first summarize the available evidence on the total QOF score, followed by the available evidence per condition (cardiovascular diseases and diabetes), the evidence on the measuring of blood pressure independent from a specific condition and finally the evidence on exception reporting. According to the labels used on page 54, all equity related publications were graded as "weak evidence".

4.5.1.1 Total QOF-score

Two cross-sectional studies [142, 148] provide information on the evolution of the total QOF score after the introduction of the QOF. In their analyses Doran et al (2008) adjusted for area characteristics (deprivation, population density), patient characteristics (% >or= 15 years of age, >or= 65 years of age, % of women, % of ethnic minority), practice characteristics (size of practice population, number of GPs per 10 000 patients, primary medical services contract), GP characteristics (age, gender, % medically educated in the UK) and exclusion rate. [148] Ashworth et al. (2007) analyzed the raw QOF data and where possible adjusted for exception reporting. [142]

Both studies showed that the existing (but small) gap in overall quality of care between deprived areas (the more deprived, the lower the overall quality) after the introduction of the QOF, narrowed during the years following the introduction of the QOF. [142, 148]

According to Doran et al. (2008) the gap in median achievement between year 1 after the introduction of QOF and year 3 narrowed from 4.0% to 0.8%. Practices in areas with initial low quality achievement (i.e. the more deprived areas) report the greatest increases during the following years. This more rapid improvement in achievement of overall quality in practices located in more deprived area quintiles was therefore attributable to poorer initial performance and not to being located in a deprived area per se. [148]

4.5.1.2 Per condition

Coronary heart disease

Two historical before-after studies assessed quality of care as defined by QOF for coronary heart disease (CHD).

Mc Govern (2008) selected 11 CDH related indicators and compared the achievement rates at pre- and post contract time points, adjusting for age, sex, number of co-morbidities, deprivation and practice size. He found a dramatic rise in the recording of CHD related quality indicators. Post-contract, disparities between patient subgroups, continued for some components of care. Women were less likely to be recorded than men in 9 of the 11 components of care, older patients (75+) in 7 of the 11 components of care and the most deprived in 4 of the 11 components of care. A secondary analysis of one of the included tables in the paper identifies an increase of the inequity gap for women (inequity for 7 indicators pre-contract, inequity for 9 indicators post-contract) and for deprived patients (inequity for 1 indicator pre-contract, inequity for 4 indicators post-contract) and a reduction in the inequity gap for older patients (inequity for 9 indicators pre-contract, inequity for 7 indicators post-contract). The authors did not describe this in the findings. [127]

Millet et al (2008) selected 10 CHD related QOF indicators and compared the achievement rates at pre- and post-contract time points (2003 versus 2005) according to the ethnicity of the patients and adjusting for age, gender, deprivation and practice clustering.

Pre-contract Blacks scored significantly worse than whites on 2 of the 10 indicators (BP control and statins prescribed). Post-contract the blood pressure control and the statin prescribing increased for both Whites and Blacks but the improvements were greater in the Blacks compared to the Whites, attenuating the disparities evident pre-contract (no significant difference any more post-contract).

Pre-contract South Asians scored significantly worse than Whites for the measurement of BP but this gap disappeared post-contract thanks to a larger increase in the achievement scores in South Asians than in Whites. Additionally, the increase in achievement scores for the measurement of BMI, the measurement of cholesterol and control of BP was greater for South Asians than for Whites resulting in significant better results for South Asians. [157]

Cerebrovascular disease (stroke/TIA)

Simpson et al (2006) found that the recording of stroke related QOF indicators increased after the introduction of the contract. Larger increases in the recording of risk factors in the oldest patients attenuated the pre-contract age differences. Women had larger increases in recording of quality indicators over time than men, however pro-men gender differences persisted in some components of care. More affluent patients tended to have larger increases in recording of quality indicators than did the most deprived. This resulted in increasing deprivation differences over time in certain aspects: the recording of a magnetic resonance imaging/computed tomography scan, smoking, cholesterol, antiplatelet or anticoagulant therapy, and influenza vaccination. A significant difference between the most and least deprived patients emerged after the contract, with the most deprived stroke patients being less likely to have a record of smoking status and blood pressure. In this study Simpson et al. adjusted for sex, age, number of stroke related co morbidities, deprivation and practice. [135]

Diabetes

Mc Govern et al (2008) selected 8 diabetes related QOF indicators (especially intermediate outcome parameters) before and after introduction of the new GMS contract in 2004. They adjusted for sex, age, number of co morbidities and deprivation. After the introduction of QOF a general rise in recording of quality indicators was observed. Few statistically significant differences were found between deprivation groups (only for recording BP). Differences between the oldest and youngest age groups (+75 versus -65) in the pre-contract dataset disappeared in the post-contract dataset.

The change was statistically significant for all indicators except for achieving cholesterol target. Women have not benefitted equally from the introduction of the new contract. Pre-contract women were as likely as men to have recording of HbA1c, blood pressure, serum creatinine and cholesterol. Post contract women were less likely to have HbA1c, serum creatinine and cholesterol recorded. [128]

In one historical comparison study, before-after time point and one historical comparison study with multiple time points, the authors looked at ethnic differences in the quality of certain indicators of diabetes care, including the achievement of HbA1c targets.

In a first study was found that Hba1c levels reduced over time for all ethnic groups but the magnitude of the improvement appeared to differ between ethnic groups even after adjusting for the effects of age, gender, years since the diagnosis, practice size and deprivation of the area where the patient lives and the area where the practice is located. A significant reduction between the pre- and post-contract measurement of HbA1c was found for the Whites but not for the Blacks and South Asians, resulting in a widening of the existing ethnic disparities in care for Blacks and South Asians. The introduction of the pay-for-performance was associated with a significantly greater improvement in women than in men. The impact on HbA1c was not found to vary significantly with the neighbourhood deprivation. [157]

In a second study with a similar methodology but on a smaller database (adjusting for the effects of age, gender, deprivation level and the clustering of patients in practices), Millet et al. (2007) found that the proportion of patients reaching HbA1c targets increased significantly after introduction of the contract. These increases were broadly uniform across all seven ethnic groups (no significant differences between the groups), except for the Black Caribbean group, which had HbA1c improvements that were significantly lower than in the White British group. [155]

In the same studies described in the above paragraph, Millet et al also investigated ethnic differences in BP targets in diabetic patients.

Blood pressure reduced over time for all ethnic groups. However the magnitude of the improvement appeared to differ between ethnic groups even after adjusting for the effects of age, gender, years since the diagnosis, practice size and deprivation of the area where the patient lives and the area where the practice is located. The average reductions in blood pressure where lower in the black patients than in the white patients, resulting in a widening of the existing ethnic disparities for Blacks. No significant difference between South Asian patients and White patients was found. [157]

In their second study, Millet et al (2007), found that the proportion of patients reaching BP targets increased significantly after introduction of the contract. These increases were broadly uniform across all seven ethnic groups (no significant differences between the groups), except for the Black Caribbean group, which had BP improvements that were significantly lower than in the White British group and this in spite of the fact that this patient group is most in need because of their hereditary predisposition for cardiovascular diseases. [155]

One historical comparison study, with multiple time points looked at QOF indicators for smokers with diabetes. This study showed that significantly more patients had their smoking status ever recorded in 2005 than in 2003. They found a larger increase in recording for women and the non-white ethnic groups (except Bangladeshi) even after adjustment for age, ethnic background, deprivation status and practice clustering.

Also the proportion of patients with documented smoking cessation advice increased significantly. No difference was found according to age, sex or ethnic group.

Concerning the prevalence of smoking an overall reduction in smoking was observed (20% to 16.2%). Both women and men benefited from this reduction but women benefited less. The reduction was not significantly different in the most and least affluent groups and the existing gap remained. Black African and Bangladeshi patients benefited less from this reduction compared to Whites. [156]

4.5.1.3 Blood Pressure

Ashworth et al (2008) conducted a historical comparison study with multiple time points, after the introduction of the QOF, adjusting for practice size and other not specified possible confounding factors. They reported on BP recording and management of BP control in the general patient population (i.e. not in a patient group with a specific condition). Blood pressure recording systematically rises over the years. Initially, there was a 1.7% gap between mean blood pressure recording levels in practices located in the least deprived fifth of communities compared with the most deprived fifth, but, three years later, this gap had narrowed to 0.2%. Improvements in achievement have resulted in the near disappearance of the achievement gap between least and most deprived areas. [169]

4.5.1.4 Generic findings

Exception reporting has been introduced within QOF to allow practices to pursue the quality improvement agenda and not be penalized, where, for example, patients do not attend for review, or where a medication cannot be prescribed due to a contraindication or side-effects. However, there is some concern that exception reporting will worsen health care disparities because financial incentives encourage providers to 'cherry pick' healthier patients or exclude those not achieving targets from public reporting mechanisms, mostly socially deprived patients or patients with a different ethnic background. Additionally, such incentives may widen health care disparities if they increase the resource gap between high and low performing health care providers. Despite this potential for harm, information on the impact of P4P incentives on health care disparities remains limited. [158]

We identified five studies looking at the phenomenon of exception reporting and the possible impact of it on inequity in health care. [84, 134, 147, 160, 230]

In one cross-sectional study immediately after the introduction of the new GMS, Doran et al (2008) analyzed determining factors in the rate of exception reporting by English physicians. They found that characteristics of patients and practices explained only 2.7% in the variance of exception reporting. Living in income deprived households and being a member of racial or ethnic minority were small but statistically significant determinants. [84]

Doran et al (2006) assessed the effect of exception reporting on reported achievement in a cross sectional study in 2005. They found that an increase of 1% in the estimated proportion of patients excluded was associated with an increase of 0.31% in achieved quality for every 1000 patients on the practice list. [147]

Sigfrid et al (2006) explored whether exception reporting is linked to socioeconomic deprivation. They found that patients with diabetes living in deprived areas are more likely to be 'exception reported' from QOF clinical indicators. [160]

Simpson et al (2006) did an analysis on a cross sectional database to study exception reporting for TIA and stroke. They found no significant association between the practice's exception reporting and the practice having proportionately more female, older or deprived stroke/TIA patients. Stroke/TIA patients with the 'top level' exclusion code 'patient unsuitable for inclusion' were more likely to be female, older, and have a diagnosis with dementia when compared to those patients without such a code. The youngest and patients from more deprived parts of Scotland were more likely to have the exception codes: 'informed dissent' or 'no response to letters'. Females were more likely to be excluded from the specific quality indicators of achieving blood pressure or cholesterol control. More deprived patients were not likely to be excluded from these quality indicators. Younger and more deprived patients were more likely to be recorded as having refused to attend for review or not replying to letters asking for attendance at primary care clinics. [134]

Finally, McLean et al (2006) examined the relation between quality of care and socioeconomic deprivation. Hereto they compared quality indicators based on the care delivered to all patients (delivered quality) and the quality indicators used for payment that allow exclusion of patients (payment quality). This made it possible to assess the quality delivered to patients with an exception report code. They found that little systematic association existed between payment quality and deprivation but for 17/33 indicators examined, delivered quality falls with increasing deprivation. An important finding of this study is that absolute differences in delivered quality are small for most simpler process measures, such as recording of smoking status or blood pressure. Greater inequalities are seen for more complex process measures such as diagnostic procedures, some intermediate outcome measures such as glycaemia control in diabetes and measures of treatment such as influenza vaccination. [230]

In Appendix 11 we summarize the findings of the above studies as a tool to answer the first research question (What is the immediate effect of the implementation of the QOF on the existing inequity in treatment and (intermediate) outcomes?). It also clearly shows that there is little information available to answer the research question on the long-term effects of the implementation of the QOF.

In addition it must be noted that currently not much evidence can be found on the underlying mechanisms that could explain these results. Exception reporting, the size of the studied group, etc could possibly have had an influence on the reported results

Key points on reported impact of P4Q programmes on equity

- **The existing gap in overall quality of care between deprived areas narrowed during the years following the introduction of the QOF. Practices in the more deprived areas report the greatest increases in quality achievement, this was due to a poorer initial performance and not to being located in a deprived area per se.**
- **Before the implementation of QOF a clear gap in health care for older patients was documented for cerebrovascular related health care, for coronary heart disease care and for diabetes care. After the introduction of QOF for all observed diseases the total number of indicators in which inequity appears diminished. For the existing inequities in health care for women, deprived patients and patients from other than white ethnic backgrounds, the results are not as clear as for the elderly patients. For women the total number of indicators in which inequity appears decreased for health care related to cerebrovascular disease. And increased for CHD and diabetes after the introduction of QOF. Considering socio-economic groups the relatively small gap for cerebrovascular disease care and CHD care increased after the implementation of QOF, whether for diabetes care the gap got smaller. Finally the small existing gap in CHD care for ethnic minorities disappeared after the introduction of QOF.**
- **Concerning exception reporting some mixed results were found with regard to the relationship between exception reporting and deprivation. However the most recent study reports that patients' characteristics explain only a very limited percentage of the variance in exception reporting.**

5 REVISING THE CONCEPTUAL FRAMEWORK BASED ON EVIDENCE

5.1 INTRODUCTION

After presenting the results focusing on P4Q effects on quality domains, this section will clarify the evidence on the mediating effects of the components and contextual characteristics which were identified in the conceptual framework (see Chapter 3). The end result will be **an updated conceptual framework with identified evidence for the role of each item, to the extent that evidence is available.** To accomplish this we make use of the systematic review results in two ways.

Firstly, the relationships between framework items and P4Q effect, identified as significant versus not significant in primary P4Q evaluation studies, are presented and their direction of effect clarified.

Secondly, by assessing each item's influence through all included studies in a descriptive way i.e. detecting whether there is a prominent difference between studies with positive effects versus studies with absent or negative effects. The results of both approaches are presented within the framework structure. The following labels are used to grade evidence:

- Strong evidence (S): a strong design (randomized studies; concurrent + historical comparison studies) with a clear effect;
- Weak evidence (W): a weak design (Concurrent comparison studies; Historical comparison studies, multiple time points; Historical comparison studies, before-after time point; cross-sectional studies) with a clear effect;
- Conflicting evidence (C): a significant effect and no significant effect within one design or within a group of weak or strong designs.
- No evidence (N): an absence of evidence

5.1.1 Quality goals and targets

As became clear in the previous section, only two quality domains are mainly focused upon in P4Q evaluation studies: clinical effectiveness (S) and equity of care (W). The results for effectiveness can be summarized as showing a positive effect or showing no effect. The studies focussing on equity rather positive results and will be discussed extensively in section 5.2.3. The few studies that focused on care continuity and integration [161, 220] showed positive effects within their study design restrictions (W).

In addition to the domains identified in the conceptual framework, two other main goals can be reported. The first goal is the use of care management processes, which can have one or more of the quality domains as a target (e.g. reminder use to improve effectiveness, equity and coordination). With regard to use of these tools a similar positive effect or absence of effect was reported (W). The second goal is to assess the effect of P4Q on patient experience and on provider work experience. This dimension is not identified in the framework. Here again a positive effect or an absence of effect were found in the included quantitative studies, but the low number of studies should be taken into account (W).

Current P4Q studies make use of process and intermediate outcome indicators. Structural indicators are used to a lesser extent (e.g. the organizational domain in the QOF). Long term outcome indicators are used very rarely (see for example Casale et al (2007) [189] and Downing et al (2007) [149]) and often only as a not incentivized control measure. The results illustrate that this target choice has a substantial influence on the effect of P4Q for such indicators, as predicted by the framework.

Whereas structural and process targets show in general a more positive effect of P4Q, this is more difficult to reach for intermediate outcome targets such as HbA1c < 7.4% in diabetes patients (S).

However, on the various types of intermediate outcomes often also positive effects were found (S). This contrasts with long term outcome targets for which no significant effects were detected, whenever included (W).

It is striking that most studies are focused on the correction of underuse of appropriate care, with varying P4Q results, whereas only two studies also focus on the overuse of inappropriate care (lab testing prescription, medical imaging prescription, drug prescription, etc.) [171, 215]. The first study found positive results, the second reported mixed effects (some positive, some showed no significant difference and the effect seemed to decrease with time) (C). For the second study the behavioural independent nature of the P4Q scheme should be taken into account. GPs agreed beforehand, by means of a democratic majority decision at a regional meeting, to improve prescribing according to a one page evidence based formulary and to be more critical and efficient in their prescribing on antibiotics and gastric medicine in general. However, the bonus was performance independent, and given to all the GPs.

Most P4Q studies are in compliance with SMART goals (specific, measurable, achievable, relevant, and timely), although it was often unclear why and by whom specific patient groups and targets were selected. If we look at P4Q results the reporting of a high involvement of stakeholders led more often to positive results, as compared to studies that did not report such an involvement (S). In addition, with regard to targets being relevant and timely, almost no study indicated that a previously detected quality problem (high variability or low performance) on a specific target was the reason to include it in a P4Q programme. Most studies referred to general lack of quality findings without assessing them in a local context as a first step, and to base indicator selection on those findings. Studies with a baseline measurement only used those results for comparison reasons, not for selection purposes. Especially in acute care conditions a number of baseline reports with already a high performance level are presented (e.g. 80 to 90% achievement on certain included measures). Looking at P4Q results this presence or absence for room for improvement seems to have been of influence on finding positive or no effects (S). A related lack in the reporting of studies concerns the dynamical aspects of quality enhancement. No studies report on a cyclical selection and refinement of indicators through time as a function of room for improvement. For the QOF a number of small scale changes to the indicator set have been made and others are planned, but their effects have not been studied yet.

There is an evolution in the number of targets and indicators which are included in P4Q programmes. Programs during the nineties included often only one or a few targets. Later on, this number expanded gradually through the years with the start up of new programmes. The QOF with almost 150 indicators is one example. In terms of P4Q results a very limited target selection seems to correspond with lesser P4Q effects, especially in the domain of preventive care (S).

5.1.2 Quality measurement

Data validity and its acceptance are reported as sufficient in most studies. With the aid of the Healthcare Effectiveness Data and Information Set (HEDIS), which is standardized over USA health plans to foster comparability, and based on strict data validity criteria many studies make use of existing databases originally intended for other purposes (although mostly also financing related). Differences in data collection method (automatic extraction, secondary data use, additional registration; sampling versus continuous) have not led to substantial differences in clinical P4Q results (S). However, the effect of the choice of the data source on cost effectiveness of the programme and on provider satisfaction (due to the level of paperwork and workload effects) remains unclear after reviewing empirical results.

A few studies tested for data coding quality and the level of gaming the system. Their results show that coding quality was generally good, but that there was also some gaming present, however to a limited extent (W). How detected gaming is dealt with is often not described.

The majority of P4Q evaluation studies include case mix adjustment in their analysis methods. It is however often unclear whether risk adjustment was only used for study purposes or whether it was used as an integral part of the P4Q quality measurement, e.g. to adjust intermediate outcome observations for patient characteristics to calculate performance and award incentives. Exception reporting was only used in UK studies. The level remained stable and has shown to fulfil its purpose in terms of guarding care equity (W). In terms of unintended consequences there is at present no evidence of a diversion of attention effect from other healthcare quality priorities. Studies that have included not incentivized control measures found an absence of effect or an improvement (W).

Figure 3 : P4Q concepts: Quality

Quality
Different possible Quality dimensions (7 dimensions): *Effectiveness (S)* *Equity of care (W)* *Integration and coordination (W)* *Provider experience (W)* *Generic applications (W)* *Other domains (N)*
Quality targets
Structure, process, and/or outcome indicators: *Structure (S)* *Process (S)* *Intermediate outcome (S)* *Long term outcome (W)* *Number of targets and indicators:* *Not too few (S)* *A lack of attention for inappropriate care* *SMART:* *A lack of attention for relevant and timely, based on room for improvement within a dynamical approach*
Quality measurement
Data source and validity: *No difference in clinical results (S)* *Other domains (N)* *Case-mix:* *A lack of distinction between study and programme risk adjustment utilization* *Exception reporting (W)* *Unintended consequences:* *At present not identified (W)*

5.1.3 P4Q incentives

Incentives of a purely positive nature (financial rewards) seem to have generated more positive effects than incentive schemes using a competitive approach (in which there are winners and losers), although this relationship is not for all indicators straightforward (C). This may have impacted the acute hospital care results found in the USA based Premier demonstration project, which showed an absence of result in the majority of targets in multiple included studies for myocardial infarction, pneumonia and heart failure.

The use of a fixed threshold versus a more continuous scale to capture best performers versus best improvers are both options that gave way to positive results in some studies and absence of any effect in others (C). What is clear however, and reported in various studies, is that the positive effect was much higher for initially low performers as compared to initially already high performers (S). This was the case using each of both options. This corresponds with the above mentioned room for improvement.

At present the included studies do not enable to make a further distinction on the effects of different incentive structures (bonus, fee schedule, withhold, regular payment increases, and quality grants). The algorithm presented by Custers et al (2008) [75] remains untested.

There is no clear relationship between incentive size and the reported P4Q results (C). Some results comply with logical expectations: QOF results are in general more positive (size of 20 to 30% of income) and the reported Premier Demonstration project results are in general of a lower or absent effect size (incentive size of 1 to 2% of income). But other studies show the opposite direction of relationship, such as both studies of Hillman and colleagues finding no significant effects (1998, 1999) [183, 184] with an incentive size of 20% and the study by Coleman et al (2007a) [144] finding large effects with a fixed amount of $5 as incentive. Several authors of evaluation studies in the USA indicate that a diluting effect of the incentive size, due to payer fragmentation, is likely to have impacted the P4Q results.

A similar incongruence can be found for determining the appropriate frequency of measurement and payment, which is the next item of the conceptual framework (C). Both studies with a very short time interval (weekly) [177, 178] and with a long time interval (yearly) [170, 222] show positive results. Studies finding an absence of effect made more use of quarterly and semi annual rates [74, 183, 184, 195], but many studies finding positive effects also used these rates.

There is an absence of evidence with regard to the choice between direct income stimuli and quality improvement investment stimuli, due to a lack of programmes and studies including the second option. The QOF, which led to mainly positive effects, is based on a combination of both. Practices receive a bonus as part of their operational revenues and can use it to reinforce the practice resources, tools and infrastructure and/or to allocate additional income to individual physicians.

The effect of simplicity versus complexity of the P4Q programme as such is difficult to assess, based on the included studies. Having a sufficient number of targets seems to be of importance, as already mentioned above. Contrary to theoretical expectations the complexity of the QOF in the UK shows no apparent negative effect on P4Q results (S). Both the approaches of using composite measures and/or all-or-none measures show mixed evidence (C), but the absence of effect in some of the related studies is likely to be related to other framework items such as incentive size, competitive nature of the programme and the use of long term outcome measures [174, 175, 189, 206].

Figure 4 : P4Q concepts: Incentives

Incentives
Incentive structure: *Lack of evidence on diverse options, best use of theoretical guidance (N)*
Threshold value and/or improvement: *In both a larger effect size for initially low performers (S)*
Weight of different quality targets: *Weighting according to target specific workload and according to sets of target types (S),* *Conflicting evidence on composite or all or none measures (C)*
Size (net additional income achievable): *Conflicting evidence, best use of theoretical guidance (C)*
Frequency: *Conflicting evidence, best use of theoretical guidance (C)*
Relative or absolute (competitive or not): *Conflicting evidence, best use of theoretical guidance (C)*
Stable and long enough: *Lack of evidence due to current P4Q initiation phase (N)*
Simplicity and directness: *No apparent negative effect of back office complexity, when combined with front office simplicity (S)*

5.1.4 Implementing and communicating the programme

Making new money available to fund a P4Q programme shows positive P4Q effects (for example in the UK), while other options show in general more mixed effects (S). However, as the cost effectiveness results indicate, in long term always adding additional funding is no option. Again, a dynamical approach is possible: adding already planned new funding with the use of an innovative P4Q design as a first step and keeping this percentage of the budget as a stable P4Q resource in the following budget periods (see the QOF in the UK). However, as the QOF also illustrates, accurate simulation and planning of the necessary P4Q resources in advance may prevent unpleasant surprises in terms of budget equilibrium. The competitive option, mainly used in the US, shows mixed effects (C). Balancing the resources with estimated cost savings is studied only in one study [69], with positive results (S).

In the UK, there was no stepwise introduction of P4Q. This has led to the need to make a number of corrections afterwards on a national scale. In the other countries demonstration projects have been used (or are ongoing) before considering national implementation. As suggested by the framework this phasing was based on geographical area, provider type, using pre-existing measure sets, etc. Some programmes made use of pay for reporting as a first step. For the USA and other countries it is at present too early to tell whether the lessons learned in such a phased approach leads to a higher positive impact of P4Q as a result.

Studies in the USA finding an absence of effect of P4Q on many measures focused relatively more on programmes using voluntary P4Q schemes [174, 175, 195, 206], although this cannot be generalized (the QOF in the UK is also a voluntary scheme) (C). One study specifically studied the hypothesis that voluntary schemes will lead to a selection bias of overly representation of already high performers, which might reinforce a lesser room for improvement [206]. These authors found that in each of the tested periods there was such a selection bias for only one or two of the tested eleven indicators (S).

Communication and participant awareness of the programme is identified in this review as an important factor of influence affecting P4Q results (S). A number of studies finding no P4Q effects clearly relate this to an absent or insufficient awareness of the existence and the elements of a P4Q programme [183, 184]. Especially in preventive care this might have limited P4Q effects. For preventive targets it also was not always clear who was responsible for providing care (e.g. vaccinating by GPs versus by other public preventive services involved).

When looking at different communication strategies we observe that the studies that made use of more extensive and more direct communication to the involved providers in general found more positive P4Q effects (S). A similar observation can be made in terms of the importance of the involvement of all stakeholders, in the first place the providers themselves, when developing the P4Q programme. However, here again it has to be noted that findings for studies with high involvement remain mixed (C).

As illustrated both in the UK and the USA, P4Q programmes are often a part of a larger quality improvement initiative and therefore combined with other interventions such as feedback, education, public reporting, etc. In the UK this seems to have reinforced the P4Q effect (S). In the USA the combination effects are more mixed (C). Reiter et al (2006) [221] found a significant influence of interventions to support structure and process change (W). A few supportive elements are almost indispensable when implementing P4Q: the presence of a quality measurement system and the use of feedback. These are already interventions as such. A few USA programmes treat these aspects as the responsibility of the providers themselves, but in the majority of programmes these elements are bundled into one package of interventions. At the moment the added value of adding staffing support, education, public reporting, guideline distribution, patient engagement, etc. remains unclear. Many studies with positive results make use of one or more of these supportive means in combination with P4Q, but the contribution of each element as such to the combined effect cannot be isolated. Concerning IT support multiple studies report a positive significant relationship (W) [138, 186, 194, 198, 199, 202, 207].

5.1.5 Evaluation of the programme

With regard to sustainability of change there is strong evidence that target performance does not regress while being incentivized (S). The evolution of target performance after it has been dropped from a P4Q indicator set is at present unknown (N). In addition, while being incentivized, there seems to be an upper limit in terms of how far any target can improve (S). This corresponds with the concept of 'room for improvement'. After a target has reached a plateau of performance the goal could shift from improving towards sustaining the level of quality of care.

Evaluation is used to assess P4Q programme performance (S). However, this review can only report this for programmes reported in peer reviewed literature. This is likely distorted by publication bias.

In terms of reviewing and revising the process currently there is no evidence of a widespread iterative and cyclical approach of quality improvement (N). This may be due to the still early stage of P4Q dissemination.

Figure 5: P4Q concepts: Implementation, communication and evaluation of the programme

Implementing and communicating the programme
Involvement of providers in setting goals: *Lacking and conflicting evidence, best use of theoretical guidance (C, N)* *Communication to whom (providers, patients, ...):* *High importance of provider communication and awareness (S)* *Mandatory or voluntary participation:* *Conflicting evidence, best use of theoretical guidance (C), No evidence of selection bias in terms of performance history due to voluntary participation (W)* *Staged approach of implementation:* *Modelling and piloting can prevent unexpected budgetary effects (S)* *Detail and terminology of the communication* *Evidence Based communication* *Targeted or widespread communication:* *High importance of direct and intensive provider communication (S)* *Stand alone P4Q programme or embedded in a broader quality project:* *A bundled approach reinforces the P4Q effects (S) and serves as a recognition of the full spectrum of non financial quality improvement initiatives*
Evaluation of the programme
Quality Measurement: see paragraph on quality *Sustainability of change:* *Target performance does not regress while being incentivized (S),* *There is an upper limit on target specific quality improvement (S),* *Lack of evidence on post P4Q target performance (N)* *Validation of the programme:* *Evaluation is confirmed in peer reviewed literature (S),* *Lack of evidence on the use of evaluation in programmes with absent or elsewhere reporting (N)* *Review and revising the process:* *Too early stage and/or insufficient use of continuous iterative quality improvement cycles* *Financial impact and return on investment: see cost effectiveness results*

5.1.6 Health care system characteristics

Congruence with the health system values was present in most P4Q programmes although the ethical question whether breastfeeding should be incentivized using negative incentives for providers not increasing this rate, as was the case in the Italian study, serves as one example of a potential exception. In the studied schemes, P4Q as such is focused specifically on clinical effectiveness (as mentioned above). Only a few USA studies combined this with explicit efficiency related incentives, without a remarkable difference in results with regard to the other quality domains (W).

A national level of P4Q decision making leads to more uniform P4Q results, as illustrated by the UK example in contrast with USA initiatives, which show more diverse forms of experimenting and innovation (S). This affects many of the previously described central elements such as the level of incentive dilution, the level of incentive awareness and communication, etc.

These worries were described in the USA situation and to a much lesser extent in the UK situation. Only one programme in the USA, the Integrated Healthcare Association (IHA) directed programme in California, is known for its alignment of different payers in how P4Q is developed and implemented.

The conceptual framework predicted that the dominant payment system would be of influence on P4Q results. One study found a small significant positive relationship of capitation with P4Q results [198]. Another study came to similar findings [173]. Since both studies concern early generation P4Q schemes with a large focus on cost containment and utilization management, these findings should be treated with caution (see also Pourat et al, 2005). Sutton & McLean (2006) [136] found in a UK study a negative association between the level of capitation payment and P4Q performance. Rittenhouse and Robinson (2006) [186] reported that a lower percentage of capitation was significantly related to more care management processes use for preventive care.

The first results do not conform to the conceptual framework expectations, the latter do (C). It should be noted that UK results for P4Q are in general positive, with the use of capitation as a general payment scheme combined, whereas USA results, often combined with Fee For Service, are more variable. Again we emphasize that many other variables act as concurrent mediators, which might cloud theoretically expected relationships.

With regard to the influence of the degree of competition between providers one study reports that higher perceived competition for attracting patients is positively related to both incentives to increase services as to incentives to decrease services (W) [212]. This seems to correspond with the distinction between underuse and overuse corrective targets. Because this study originates from the USA with an environment with high patient volatility and high levels of public reporting (consumer driven healthcare), these results might not be readily transferrable to other country healthcare systems.

Figure 6: P4Q context: Health care system characteristics

Health care system characteristics
Values of the system: *Lack of reporting on the level of congruence (N)* *No negative effect of combining clinical quality with efficiency goals (W)*
Type of system (e.g. insurance or NHS): *Independent of the identification of P4Q initiators and coordinators, uniformity of P4Q design is important (S)*
Level of Competition: *Lack of evidence on how the general level of competition in the health care system influences P4Q (N). A lower patient volatility is assumed to support P4Q on the one hand. P4Q and consumer driven care might reinforce each other on the other hand.*
Decentralisation of decision making: *Centralized decision making supports uniformity, the avoidance of incentive size dilution, transparency and awareness (S). However, this should be combined with local priority setting based on room for improvement (S).*
Dominant payment system (FFS, salary, capitation, ...): *Conflicting evidence, best use of theoretical guidance (C)*

5.1.7 Payer characteristics

The relationship of the typology of the payer (e.g. private/public) to P4Q results is not reported to be significant in the current P4Q evaluation studies (W), but this has received little attention.

One major element, which is called in the original conceptual framework 'other incentive programmes running', confirms the role of the dilution effect stated above. One study reports the significant positive relationship with the proportion of revenue generated by the P4Q payment system [194]. Other authors found a similar relationship for the influence of other health plan or Health Maintenance Organization (HMO) reimbursement mechanisms [185], the difference in belonging to one or more managed care plans [180] and the percentage of HMO penetration [207, 208]. All these results confirm the importance of the dilution effect in the American healthcare system (W).

In the USA, where various payer and insurance arrangements are present, delegation of utilization management for hospital admission by an HMO to the physician organization showed a small association with P4Q performance [198, 199, 207]. However, having hospital admitting privileges or being a preferred Medicaid provider showed no relationship in another study [180] (W).

There is no further evidence on the effect of other payer characteristics.

Figure 7: P4Q context: Payer characteristics

Payer characteristics
Use of clinical guidelines in current policy: *Lack of reporting, although this seems one of the mediators of the general quality improvement trend co existing with the introduction of P4Q (see UK example) (N)* *Variable patient contribution in function of provider and/or technology performance:* *Lack of reporting in the USA (N). P4Q in the other countries is designed without a variable patient contribution.* *Other incentive programmes running:* *Important influence of a dilution effect by other incentive programmes when not aligned (W)* *Availability of management information systems:* *The different approaches do not translate into differences in clinical effect (S). Lack of reporting in the included studies of payer and provider work experience effects (N).* *Number of payers: see importance of dilution effects (W)* *Accuracy of the data information system and underlying databases:* *Use of sufficient validity safeguards in most studies. Evidence of gaming by providers to a very limited degree (W).* *Vision of the payer with regard to health objectives:* *Lack of reporting (N)* *Typology (private/public; local, regional, national…):* *Scarce reporting of an absence of effect of these differences (W)*

5.1.8 Provider characteristics

Although mentioned by some authors [207], in the included studies there has been a lack of attention for the effects of (dis)congruence with professional culture and with internal motivation. Through high involvement and democratically decision making on the implementation of P4Q it seems that these issues can be addressed, as the UK example shows. But it remains unclear what the impact is in terms of P4Q results, as compared to some programmes in other countries where P4Q sometimes was imposed on care providers by the state, employers and/or health plans. The same is true for the influence of leadership support [207]. Roski et al (2003) [201] question the effect of turnover in senior management without studying this further. One study found no significant association with the nature of the organizational culture, but did find an association with having a patient centred culture [192] (W). Ashworth et al (2005) [140] mention the potential influence of existing professional standards and pride.

As mentioned before, some authors suggest that the level of knowledge and awareness of the existence and design of the P4Q programme might influence results [193, 201]. Except for the finding of a relationship with the awareness of a clinical guideline intended as a supportive tool (W), this has not been specifically studied quantitatively [209].

Next to the already discussed room for improvement, the history of engagement with quality improvement activities [196] has a significant relationship with P4Q results (S).

With regard to the target unit of the P4Q incentive, programmes aimed at the individual provider level report in general positive results (S). The study by Young et al (2007) [187] is one exception. An absence of effect is found more in programmes targeted at the organization level [74, 172, 174, 183, 184, 188]. Again, there are also programmes showing positive results at this level, but this seems to require additional efforts (C). Incentives at a team level showed positive results in all three studies (S) [171, 205, 206]. One study with incentives at the administrator/leadership level found mixed results (S) [223]. A combination of incentives aimed at different target units is rarely used [191].

Finally, on each of these levels a number of characteristics have been further investigated. The first group concerns individual provider demographics. Provider age was in one study positively related to performance and acceptance (W) [195]. In this study younger physicians made more use of feedback data, while older physicians made more use of cues and stickers. According to the results of Doran et al (2006) [147] provider age effects are moderate (W). Wang et al (2006) [137] found that older providers were less likely to participate in voluntary schemes and performed less due to differences in the organizational domain (structural support), not in the clinical domain (W).

According to one study male providers were more likely to perform better on P4Q programmes [177]. This contrast with the findings by Wang and colleagues (2006) [137] who found that male providers were less likely to participate and to perform well. Doran et al (2006) [147] also found that female physicians performed slightly better (C).

With regard to provider ethnicity one study found a strong relationship indicating that non white physicians were more likely to perform better on P4Q (W) [195].

The level of provider experience showed no significant effect on P4Q results (W) [170].

There is mixed evidence on the effect of the specialty and/or general practitioner background of the provider (C). Some studies found no significant relationship [183]. In the study by Rosenthal et al (2008) [69] providers meeting P4Q targets were more likely to be specialists than general practitioners. The same is reported in other studies [144, 170, 173, 177]. To the contrary other studies found a positive relationship with the percentage of general practitioners [202, 209]. These differences seem to depend on the nature of the targets being studied. According to their content some fall better within a general practitioner's scope of work and expertise, while others fall within specialists' areas. Grady et al (1997) [195] found a positive relationship with having a second specialty (W).

Grady et al (1997) [195] reports that not residence trained physicians are more likely to perform better on P4Q (W).

A medical education in the UK is weakly related to better P4Q performance on the QOF [147]. One study in the USA reports that physicians trained abroad perform better [144] (W).

On the level of provider organizations most research has been done concerning general practices, medical groups and independent practice associations. Little results are known on a hospital level.

One collection of studies has reported on the effect of those distinctions: the difference between medical group, Independent Practice Association (IPA), hospital and community based P4Q performance. Medical groups are likely to perform better than IPAs (W) [199, 202, 209]. Rittenhouse & Robinson (2006) [186] reported that hospital and community based care performed better than IPA based care (W). The relatively good performance of medical groups or networks of medical groups is also confirmed by Mehrotra et al (2007) [200] (W).

A positive relationship is reported for the age of the group or the age of the organization [202, 208] (W).

In the USA, ownership of the organization by a hospital or health plan is positively related to P4Q performance, as compared to individual provider ownership (W) [198, 199, 202, 207, 208, 209]. Practice ownership by a provider is associated with incentives to increase services (W) [212]. Full ownership of groups is associated both with incentives to increase some and to reduce other services. One study found only for preventive care a positive association with an organization being profitable and P4Q results (W) [186]. Bhattacharyya et al (2008) [211] found no relationship with the size of revenues (W). Some of these authors point to the difference in available resources for investment purposes as part of a possible explanation.

The teaching status of a hospital is positively related to P4Q performance [211] (W). Geographical location is sometimes also positively related to P4Q results (W). This is illustrated by two studies finding an association with a location in the Midwestern USA and in California as compared to other USA regions [199, 211]. There is no difference in performance between rural, urban and mixed areas [132] (W).

There is conflicting evidence concerning the influence of the size of the organization in terms of number of providers and number of patients (C). Some studies report a positive relationship between the number of patients and P4Q performance (W) [95, 131, 140, 153, 208, 209, 211]. Others report no relationship [183] or a negative relationship [142, 177] (W). There is a small negative relationship with the practice population size (W) [147]. Practices with a large patient population are also more likely to exception report more patients (W) [84]. The size of a hospital is not related to P4Q results (W) [211].

Group practices perform better on P4Q than single handed practices according to some studies [132, 177, 183] and the other way around according to other studies [195] (C). Mehrotra et al (2007) [200] found a positive relationship with having more than the median number of physicians available. Other studies came to similar results [198, 199, 202, 207] (W). Sutton & McLean (2006) [136] found a similar positive relationship for the size of the team and the non principal proportion of the team (recently trained) (W). Ashworth et al (2005) [140] found this kind of relationship for the size of staff budgets (W).

A smaller practice size is also related to other factors such as having patients with poorer health, being located in a deprived area, having more patients from minority ethnic groups, etc. [137] (W). These interrelationships have to be taken into account when assessing the practice size characteristic and its P4Q effects. The complexity of such relationships is also illustrated by McLean et al (2007) [129] who mapped factors related to remoteness of the practice area.

Tahrani et al (2008) [166] report that the performance gap between large versus small practices which existed before QOF implementation has disappeared afterwards (W).

Figure 8: P4Q context: Provider characteristics

Provider characteristics
Awareness, perception, familiarity, agreement, self-efficacy: *High importance of provider awareness (S)* *Lack of reporting on the other experience dimensions, except some evidence on the importance of involvement (W)* *Other motivational drivers (intrinsic, professional culture, altruism…):* *Lack of reporting (N)* *Medical leadership, role of peers, role of industry:* *Lack of reporting (N)* *Existence and implementation of clinical guidelines:* *Lack of reporting, although this seems one of the mediators of the general quality improvement trend co existing with the introduction of P4Q (see UK example) (N)* *Level of own control on changes:* *Lack of reporting (N), but in almost all studies (except smoking cessation studies) controllable measures were targeted. In addition most studies use intermediate instead of long term outcome targets.* *Target unit (individual, group/organisation, hospital, nursing home, department):* *Evidence of positive effects on the individual and/or team level (S).* *Conflicting evidence on the level of an organization (medical group, hospital) and on the level of leadership (C).* *In case of not-individual, size of unit (# providers):* *Conflicting evidence on solo vs. group practice performance (C)* *Positive relationship with the number of providers within a practice (W)* *No relationship with hospital size (W)* *Role of the meso level (principal or agent):* *Lack of reporting (N)* *Demographics (age, gender, specialty…):* *Significant effect of provider age, gender, training background, geographical location, and having a second specialty (W)* *No significant effect of provider experience and rural vs. urban location (W)* *Organisational resources available:* *Weak evidence on the influence on P4Q effects, as measured through hospital/medical group/IPA status, age of the group or organization, organization vs. individual ownership, and teaching status of an organization (W)* *Organisational system change and extra cost/time required:* *Lack of reporting (N)* *Number of patients and services per patient:* *Conflicting evidence (C)* *Room for improvement:* *Strong evidence on the influence on P4Q effects (S)*

5.1.9 Patient characteristics

The evidence on relationships of patient demographics has been described extensively in the previous section. With regard to age, ethnicity and socio-economical deprivation level positive as well as negative results were found, depending on the disease (W). This will be discussed more extensively in section 5.2.3.

Currently there is a lack of research and evidence on the effects of patient educational status and insurance status.

There are also no negative findings reported with regard to the use of exception reporting as a P4Q supportive tool in the UK (W).

Almost no empirical research has focused upon the patient experience and patient satisfaction with regard to P4Q. The Spanish study as one exception found no significant differences (W). There is no further evidence on how P4Q and patient awareness of P4Q affects the patient provider relationship.

Finally, patient behaviour in terms of lifestyle, cooperation and therapeutic compliance might affect P4Q results, as described by the conceptual framework. Again, there is a lack of evidence on this specific topic. However, as part of P4Q target setting and measurement the selection of indicators in the included studies indicates that this issue is taken into account in almost all studies. The structure, process and intermediate outcome measures used have a high degree of provider controllability. There are a few exceptions like the use of long term smoking cessation outcomes, which is in general less controllable and more patient lifestyle related.

Figure 9: P4Q context: Patient characteristics

Patient characteristics
Demographics, Co-morbidities: *Closing performance gap with regard to patient age and unclear result with regard to gender, and ethnicity (W)* *Socio-economics, Insurance status:* *Unclear results with regard to socio economical deprivation level (W)* *Lack of reporting on the influence of insurance status (N)* *Information about price and/or quality:* *Conflicting evidence on the interaction of P4Q with public reporting (C)* *Patient behavioural patterns (cultural and consumer patterns, compliance):* *Lack of reporting (N)*

5.2 DISCUSSION AND CONCLUSIONS

5.2.1 Reported effect of Pay for Quality programmes

Previous systematic reviews found that the evidence on P4Q effectiveness is mixed, with initial studies finding a lack of impact or incongruent effects which are target specific. They also reported a lack of evidence on the incidence of unintended consequences of P4Q. A high need for further research on these issues was expressed [15, 16, 37, 38, 40, 46, 75, 92, 231, 232, 233, 234]. The systematic review that was performed in this chapter builds further on a wider search strategy to identify more primary P4Q evaluation studies. In addition, it provides an update of the findings till 2009 as year of publication. Because a lot of studies have been published in 2007 and 2008, this review allowed assessing whether already more evidence is available on P4Q development, implementation and evaluation, how this relates to our conceptual framework and which recommendations can be formulated for research and practice, both within an international and a Belgian scope.

The current results are mainly based on a rapid increasing amount of experience in the USA and the UK. Although a number of start up countries were also included, they still are in such a premature phase that the lessons to be learned from their experience are still limited. However, where appropriate we will refer to their choices of P4Q design and implementation.

Before discussing the results the following methodological limitations should be taken into account:

Although the systematic review was performed using multiple databases and several languages, there exist still other databases and languages which we were unable to check.

The tool used for the quality appraisal to eliminate low quality primary studies was based on several existing validated tools, which were combined into a generic tool applicable to P4Q studies. The use of this combined tool is not validated as such.

Because observational studies are the main source of information on P4Q, it was decided to report P4Q results comprehensively, without a restriction to randomized designs. A distinction between strong and weak evidence was made. Selection bias can therefore not be excluded.

Not all relevant contextual information is available for each of the included studies. For example, leadership as a factor might impact P4Q results, but is rarely reported upon. As has become clear in the previous study description section, for the contextual factors, the P4Q intervention description and the process of implementation and communication, the completeness of reporting is often lacking in P4Q evaluation studies.

The limitations in terms of review selection criteria have to be taken into account. A strict P4Q definition was used (e.g. public reporting is not considered to be a form of P4Q). The population does not include studies outside a primary care or general hospital setting (e.g. psychiatric care and nursing homes).

The limitations in terms of studies' selection criteria have to be taken into account. With regard to context, studies are more performed in urban areas than in rural areas. Since a number of studies use a minimal patient panel size per provider or per target for each provider, it is likely that smaller practices are underrepresented. Furthermore, because many studies focus upon a stabile and regular type of patients, results may not be applicable for one time only and quickly physician switching patients. The same is true for patients who are exception reported in the UK system and are excluded from most UK study samples, because their performance data are rarely available.

As described in chapter 4 (section 4.2.6), there is a general trend of quality improvement present in some of the countries where P4Q studies took place, like in the UK. Only results based on strong evidence take this trend into account to isolate the P4Q effect from the time trend effect. In addition, P4Q is an intervention that's typically combined with other interventions. Because of the mix throughout all P4Q studies, it is impossible to separate the P4Q impact from the other concurrent quality improvement initiatives' impact. The studies discussed here are always based on a 'bundled' approach.

P4Q is still a very recent phenomenon. For many medical conditions there is no P4Q evidence (urology, intensive care, geriatrics, most of surgical care, etc). Secondly, there are medical conditions for which some first evaluation attempts have been performed, but for which the evidence is still scarce and no conclusions about P4Q effects can be drawn (urinary tract infections, skin infections, gastric infections, depression/mental illness, chronic child care, COPD, epilepsy, hyperthyroidism, chronic cancer care, osteoarthritis). In more generic terms the same can be said for the introduction of new drugs and for P4Q effects on the use of most care management processes when assessed separately.

However, this still leaves a large group of target conditions for which evidence is available. The effects on the clinical effectiveness domain are reported firstly.

What first comes to mind when reviewing the results is that P4Q is no magic bullet. Throughout studies and targets non significant or small size effects are regularly encountered. On the contrary, large effect sizes are also sometimes reported. To come to a comprehensive overview the results are categorized in a systematic way, starting from a negative effect up to large positive effects:

For a few targets a small number of studies have identified a lower degree of quality improvement when using P4Q as compared to the degree of improvement when not using P4Q. This was the case for Chlamydia screening with an 11% negative difference in one weak design study, for left ventricular failure (LVF) assessment in heart failure patients with a 2.4% difference in another strong design study and for oxygenation assessment and timely antibiotics administration in pneumonia patients in a third strong design study (1.9 and 3.2% respectively). Finally, for cholesterol recording in coronary heart disease (CHD) patients one weak design study found a 10.8% negative difference. It should be noted that the same LVF assessment, oxygenation assessment and timely antibiotics administration showed a positive P4Q effect up to 5.1%, no significance and a positive 4.3% effect respectively in other strong design studies. For cholesterol recording in CHD patients this even went up to a positive effect of 41.7% in other weak design studies.

There are also a number of targets for which only an absence of effect is found. This was the case for most of the long term outcome targets in the one weak design study focusing on coronary artery bypass grafting (CABG). In heart failure patients in hospital care no effect was found on ACE inhibitor use and smoking cessation advice, based on strong evidence. The same was true for one myocardial infarction (MI) indicator, four diabetes care indicators, five CHD indicators, two stroke targets, two asthma indicators, two smoking cessation indicators and three chronic kidney disease indicators.

On the following set of indicators P4Q effects show mixed evidence, ranging from no significant effect to a small below five percent effect. This set includes cholesterol screening in adults (3%), well child visits (0-5%), cancer preventive screening (0-4%), most MI indicators (0-4.4%), one pneumonia indicator (blood culture testing, 3.5%), three diabetes indicators, three CHD indicators, and four asthma indicators.

There are also indicators showing a much larger range of P4Q effect, according to the specific studies, but still also with no effect in some. Here we identify children immunization (0-25%), children preventive screening (0-29%), the timeliness of emergency care (0-10%), two MI indicators (0-8.5% for prescribing aspirin at discharge and 0-9.9% for ACE inhibitor use). In diabetes indicators, the effects on four testing rate targets varied between zero and 25%. For HbA1c as an intermediate outcome the range was 0-14%. These figures are based on strong evidence.

In the included studies, there are a number of targets which show only a positive effect. This includes influenza immunization (6-8%), the provision of discharge instructions to heart failure patients (35.5%), pneumococcal screening and/or vaccination for pneumonia (9.5-44.7%), and three diabetes indicators (cholesterol outcome: up to 23.5%, blood pressure outcome: 1.6-6.3%, foot exam rate: 2.7-45%), based on strong evidence. The same is true for smoking status recording (7.9-24%) and referral rate (6.2%) aimed at smoking cessation patients. Weaker evidence supports a similar finding for reducing inappropriate treatment in acute sinusitis (14-29%), breastfeeding rate (6-12%), ACE inhibitor use for heart failure in primary care (23.4%), blood pressure recording (0.7-21.5%) and smoking status recording (2.39-26.2%) for CHD patients, eight targets (17-52.1%) for stroke patients, and finally, for hypertension care targets (12%).

The overview above shows that a P4Q programme can lead to the full spectrum of clinical effectiveness results, ranging from no effect, a negligible effect, a substantial effect to a high effect, depending on the local programme and the targets selected. The evidence base has become large enough to make these distinctions. The few negative differences that were found in only four studies out of more than hundred are bound to appear in any review of so many targets, when including observational study designs. Further scrutiny of these specific results may provide possible explanations.

There are some important points to consider when interpreting the results. These have to do with the purpose of P4Q (i.e. the basic philosophy supporting the concept), how to use it as a quality supportive tool and the expectations in terms of effect size.

It seems that there are two ways of looking at P4Q.

One philosophy sees P4Q as a broad target based rule set, mainly focusing on minimally required quality standards which are the same for all providers, and which remain relatively stable in terms of target selection. The target selection in this philosophy is mainly driven by the level of evidence supporting the target, being able to measure the target and providers having sufficient control over the targets. In this philosophy no adaptations are made according to local needs. National or regional needs guide the process. From this point of view, are the results above adequate to support widespread P4Q dissemination? The answer will be no if one expects P4Q to cause a 'leapfrog' jump in quality improvement on the whole set of targets, while keeping the programme fixed and static as it is. However, because in this philosophy local needs are not accounted for in terms of the potential room for improvement for specific targets, is such a 'leapfrog' expectation in fact not unrealistic? In many of the programmes on a number of the included targets an already pre existing high baseline level of target achievement was reached before P4Q programme implementation. For example one Australian study showed already a 100% performance on its most important target included. One cannot do better than 100% and for many targets it can be questioned how much improvement is still possible once thresholds of above 80 and 90% have been reached. If these targets are to be included in P4Q programmes, as a kind of long term control measure, then their result of showing no significant effect or a modest sized effect is likely the best that can be expected. This sheds already a completely different light over the wide range of results reported above. The above philosophy will lead to more and more of such no significant and small sized P4Q effects, as everybody will reach the thresholds during multiple feedback cycles. Then afterwards the programme may be refocused at other national or regional priorities. The results indicate that P4Q can fit such a purpose, however without getting the maximal P4Q effects possible out of the system. In the long run this leads to providers receiving a bonus for things they are already doing without changing behaviour. One can question whether this is the best option for society, both in terms of health gain as in terms of cost effectiveness as will be discussed in section 5.2.2 (page 87).

There is however a second P4Q philosophy conceivable to amend the shortcomings mentioned above. As the results show, high effects can also be reached by implementing P4Q, but different ones according to local programmes and targets. This corresponds to the fact that the highest gains are found for providers performing the least on a specific target, as shown in several studies. So the purpose of a P4Q programme, instead of or combined with reaching top performance on all minimal measures, may also be to correct locally existing quality gaps. And the results indicate that P4Q may also fit this purpose when aligned with the targets with the highest local room for improvement. It is remarkable to note that almost no study really focused on local target relevance and timeliness as additional selection criteria. With relevance we mean that local baseline data are used to further prioritize final target selection, based on the identification of quality of care gaps.

Timeliness points to the fact that P4Q can be used in a much more dynamical way than currently used by seeing it as part of the basic concept of quality improvement management: cycles of continuous improvement, also referred to as PDCA (plan do check act) cycles. These two criteria, which are fundamental to quality of healthcare theories and practice, are underused in current P4Q practice, as presented in evaluation studies. By simply making it possible for local providers and organizations to prioritize targets based on their biggest quality gaps a P4Q programme would become focused on those targets where the highest quality gains can be expected and the chance to reach high gains as compared to absent or small gains will increase. This modification will also raise local responsibility, involvement and awareness, which is sometimes lacking in current P4Q programmes. Finally, by such prioritizing the whole P4Q system automatically becomes much more dynamical, beyond a periodic review of target selection, towards a more continuous process.

The current evaluation studies have shown that P4Q is a feasible concept to implement and that it shows the potential to fulfil its purpose, being a stable minimal standard approach and/or a more dynamical and focused approach. In the end we hypothesize that both options would reinforce each other, giving P4Q both a national/regional and local dimension.

Next to clinical effectiveness there is also a quickly growing body of evidence on the equity of care effects, which will be extensively discussed in section 5.2.3.

There is a lack of evidence of the effect of P4Q on other domains such as care continuity, care coordination, patient/provider experience and efficiency. Isolated studies show that directing P4Q at coordination of care also can result in substantial positive effects (58% on patient acceptance in emergency care, 23% on the referral rate for poor glycaemia control). However, as shown in emergency care, for P4Q to be effective on coordination, a one sided incentive system is unfit, and the incentives of the coordinating parties should be aligned. One Spanish study found absent or only positive effects of P4Q on patient and provider experience. The only negative aspect was a higher workload as perceived by physicians. It is clear that a lot more empirical research is necessary to clarify these effects.

Another remarkable finding is the lack of attention for P4Q effects on the inappropriate use of resources. Although an acute sinusitis study found substantial decreases in the use of inappropriate drugs, medical imaging, etc. most P4Q programmes neglect such effects to avoid a connotation with a hidden cost containment goal as the primary P4Q purpose. However, correcting underuse in the use of drugs, testing and medical examinations will often coincide with a reduction of the previously used inappropriately used alternatives (e.g. one type of antibiotics instead of the other). These evolutions merit further investigation.

In terms of unintended consequences the current results are consistent. Firstly, there seems to be no effect of P4Q implementation on other not incentivized quality targets. This means that there is neither a negative effect, nor a positive spill over effect. Some authors criticize the fact that there is often no direct relationship with patient outcomes such as readmissions, complication rate and mortality rate. Because P4Q is a new phenomenon it can however be questioned whether such a long term relationship can develop in such a short time frame. Measuring P4Q effects and long term outcome effects at the same time, right after P4Q implementation does not take into account that potential P4Q effects on those measures will take possibly years to develop, e.g. in the prevention of diabetes complications. Therefore, at present no sound conclusions concerning those effects may be formulated yet.

5.2.2 Reported cost effectiveness and modelling effects of P4Q programmes

As stated by the three studies concerning cost-effectiveness, P4Q programmes can be cost-effective. From the 12 QOF indicators considered by Mason et al. (2008) [224] 11 indicators seem to be cost-effective. Only one indicator, namely diabetes retinal screening was not cost-effective according to their economic framework. The USA hospital incentive programme for heart-related care, studied by Nahra et al. (2006) [107], also seems to be cost-effective, both in worst as in best case scenario. And finally the study by Curtin et al. (2006) [107, 225] concerning a pay for quality programme focussing on diabetes patients in primary care, shows a positive return on investment.

It must be noted that these studies are mostly based on a few assumptions in cost and health gain, generally based on literature data. Consequently the results of these studies are estimates, and have to be treated with caution.

In order for an indicator or a programme to be cost-effective, an absolute change between baseline utilisation and the end-utilisation is required. It is important to notice that the cost-effectiveness of a certain indicator depends on the baseline uptake. A first reason concerns the room for improvement. If there is already a high baseline achievement for a certain indicator, it is more difficult to reach an even higher score.

Secondly, if baseline uptake is high, the mean absolute change in utilisation needed for an indicator to be cost-effective needs to be larger than when the baseline uptake is low, because physicians also receive an incentive for those patients for whom the target was already reached and no additional health benefits have been achieved. As a result a high absolute change (and thus a high additional health gain) is required to compensate for the payments without extra health gain. It must also be noted that cost-effectiveness of an indicator will be partly determined by the duration of the P4Q programme. To be cost-effective it is desirable to adjust the payment according to the observed utilisation levels throughout time.

We can conclude that P4Q programmes seem to be cost-effective. It can be stated that implementing indicators with a low baseline achievement, have the highest potential to be cost-effective. Since these conclusions are based on only three studies, it is recommended to augment the number of cost-effectiveness evaluations of P4Q programmes in the future.

Kahn et al. (2006) [229] modelled the financial gains and losses for hospitals, using two P4Q programmes. In both programmes, urban hospitals are more likely to attract bonuses as well as to get penalties resulting in a financial gain in one programme and a small financial loss in the other programme. Rural hospitals, receive less bonuses, but also less penalties, resulting in a small financial gain in both programmes. Teaching hospitals are more likely to have a financial gain unlike non teaching hospitals, which receive a larger amount of bonuses but also get more penalties, resulting in a small financial loss in both programmes. Finally government and investor-owned hospitals are more likely to suffer a financial loss, in contrast to tax-exempt hospitals that are more likely to experience a financial gain.

The payment reduction, as a result of a new inpatient prospective payment system was estimated by Averill et al. (2006) [95]. The new system provides hospitals with a financial incentive to reduce complications and to improve the quality of care. This resulted in a reduction of the Medicare DRG hospital payment with approximately 1%. These savings have the potential to increase payments to high performing hospitals.

According to Fleetcroft et al. (2006) [227], QOF payments don't reflect likely health gain. The use of ACE inhibitors in heart failure relates to a maximum of 308 lives saved per 100 000 people per year and is linked to a maximum payment of £2 400 per practice per year. In contrast the screening and treatment of hypertension only saves a potential maximum of 71 lives per 100 000 people per year but costs on average £17 280 per practice per year.

As a result the GPs might focus on the highly rewarded indicators with sometimes relatively low population health gain.

McElduff et al. (2004) [228] estimated the health gain among cardiovascular disease (CVD) patients within QOF. Reaching the target concerning cholesterol-level and hypertension would result in important reduction in number of CVD events. Fleetcroft et al. (2008) [226 , 228] confirm the potential for significant health gain regarding clinical indicators. The difference between actual health gain and potential health gain can be explained by several determinants among which the room for improvement (baseline activity) and the prevalence of conditions or associated event risk. Additionally, the possibility of exception reporting within QOF reduces the potential maximum health gain. Furthermore, better recording rather than actual improved performance could increase the achievement. Gaming among practitioners could also influence the health gain. The health gain in patients with multiple conditions may be less or more than the sum of the beneficial effect of each intervention. And, finally, some indicators have potential side effects, which could reduce health gain (e.g. adverse events of some drugs).

5.2.3 Reported impact of P4Q on equity

(In)equity in health care is a central point of attention of many health care systems and tackling this inequity has been an important objective in the development and reorganization of health services. [63] There is widespread concern that the focus on quality improving systems driven by financial incentives may lead to a widening of the existing inequity in health care. Within this report the impact of the introduction of the Quality and Outcome Framework in the UK on equity in treatment and (intermediate) outcomes was investigated. More specifically with this study we want to target the following three sub domains: the immediate effect of the implementation of QOF on the existing inequity in treatment and (intermediate) outcomes, the effects on long term and the contribution of exception reporting in treatment and (intermediate) outcomes.

Several limitations in the selected studies complicate the formulation of the evidence, prompting utmost prudence in interpreting and generalizing the results of this study.

In the assessment of *equal access* to care it is essential to look for differences in social or ethnic background, gender … between the users of health care and the non-users of health care, both with the same need for care. None of the studies addresses this issue: they do not include information on the ratio users/non-users (both in equal need for care) and on the variation in the characteristics of the users and the non-users. This makes it impossible in this study to pronounce upon overall equity in access.

The majority of the studies make no judgments about the appropriateness of the indicators or the treatment targets for both groups. As a result, similar screening or treatment rates can actually mean under treatment of certain groups, hence inequity [63].In none of the selected studies normative need, felt need or expressed need is taken into consideration when observing differences in treatment and/or (intermediate) treatment outcomes. In the majority of the studies the authors (inexplicitly) adopt a comparative approach to need: when variations are found between the treatment rates and outcomes of two groups of patients with the same condition (e.g. low-income versus high-income diabetic patients), inequity is presumed. Characteristic of a comparative approach of need is that it makes no judgments about the appropriateness of the indicators or the treatment targets for both groups. E.g. when no differences are found in the cervical screening rates between population A and B most of the selected studies would presume equity. However knowing some groups have a higher risk on cervical cancer related to number of sexual partners, similar screening rates actually mean under treatment of this second group and so inequity.

As a result utmost prudence is necessary when interpreting the results of the studies: the absence of social, gender or age differences should not automatically be interpreted as absence of inequities. [63]

Questions can be asked about the relevance and the completeness of the indicators that are used to measure quality. Although initiatives such as the QOF cover many important aspects of quality of care, the inherent strength and complexity of the doctor-patient relationship supports quality at a much deeper level which is not captured by the QOF indicators. The same reasoning applies to the fact that the selected publications mainly focus on intermediate outcomes and less on final outcome measures. To what extent equity in intermediate outcomes or process indicators predict final outcomes, not to speak to what extent the found (in)equities in health care predict (in)equities in health?

The selected studies have weak study designs according to the labels presented in chapter 4. Of the 27 studies studying the pay-for-performance initiative in the UK

- 17 have a cross-sectional design with only one point of measurement;
- 7 studies have a serial cross-sectional design with several points of measurement in time (of which only 3 with both measurements before and after the introduction of the new initiative)
- 3 studies have a longitudinal design with several points of measurement in time and linking of the data from the same study subject (e.g. patient) over time

This means that only 6 of the 27 studies have a study design that is appropriate to describe the effects of the implementation of the initiative. None of these 6 studies report on the effects more than two year after the implementation of the initiative. The most recent data analyzed in the 32 reported studies are from 2007 [148, 169]

An important number of studies use the practice outcomes and/or use area level scores of deprivation as a proxy for the socioeconomic status of the patient. These studies assume that the eventually associations observed at the practice or area level reflect the same association at the individual level. This may not be true, a problem known as the ecological fallacy [133 , 160 , 164].

Notwithstanding these limitations this study comes to some interesting findings that can certainly contribute to the knowledge base of the equity debate.

As discussed in section 5.2.1 the quality of care in the UK generally improved with the introduction of QOF and for the majority of the observed indicators all citizens benefit from this improvement. However, the extent to which different patient groups benefit tends to vary and to be highly dependent on the type and complexity of the indicator(s) under study, the observed patient groups (age groups, males versus females, socioeconomic groups or ethnic groups), the characteristics of the study (design, level of analysis, covariates, …) and the level of detail of the studied indicators.

Before the implementation of the QOF a clear gap in health care for older patients was documented for stroke care, for CHD care and for diabetes care. After the introduction of the QOF, for all observed diseases the net gap[d] becomes smaller. For the existing inequities in health care for women, deprived patients and patients from other than white ethnic backgrounds, the results are not as clear as for the elderly patients. Pre-contract, for women a net gap in health care was documented for stroke, for CHD and for diabetes care. For health care related to stroke the net gender gap got smaller after the implementation of the QOF. For CHD and diabetes care the net gender gap increased.

Considering socioeconomic groups, the relatively small gap for stroke care and CHD care increased after the implementation of the QOF, whether for diabetes care the gap got smaller. Finally, the small existing gap in CHD care for ethnic minorities disappeared after the introduction of the QOF.

When looking at inequity at a more detailed level, the level of individual indicators, the findings become even more complex, scattered and sometimes contradictory.

Hereby we summarize the most marked results:

- Post-contract improvements in blood pressure control and statin prescribing increased for both white CHD patients and black CHD patients but to a larger extent for blacks, completely attenuating the disparities evident pre-contract. The same results were found for the measurement of blood pressure in the South Asian patients.
- Regarding stroke related QOF indicators, for the recording of a magnetic resonance imaging/computed tomography scan, smoking, cholesterol, antiplatelet or anticoagulant therapy, and influenza vaccination, a significant difference between the most and least deprived patients was found.
- Recording of HbA1c and the achievement of BP goals increased more in white diabetes patients than in black diabetic patients and in South Asian diabetic patients resulting in a widening of the existing ethnic disparities in care for blacks and Asians.
- Similar increases in HbA1c measuring and BP measuring were found across all ethnic groups except for the Black Caribbean group who had lower achievement in BP goals and in HbA1c targets.

[d] Net result or net gap: If the total number of indicators in which inequity appears pre-contract > the total number of indicators in which inequity appears post-contract, the net result is a decrease of the gap. If the total number of indicators in which inequity appears pre-contract < the total number of indicators in which inequity appears post-contract, the net result is an increase of the gap.

In general we see that all citizens benefit from the improvements in quality of care and the extent to which they benefit determine whether the existing gap narrows (when the least off have a larger growth than the best off) or increases (if the least off have a smaller growth than the best off). However, for some indicators a new gap arises there where there was no gap pre-contract. For example a significant difference between the most and least deprived patients emerged after the contract for the recording of blood pressure, the recording of smoking status and giving smoking advice. Also pre-contract diabetic women were as likely as men to have their HbA1c, blood pressure, serum creatinine and cholesterol recorded where post-contract inequities in these indicators appeared.

For some indicators, the increase in quality of care for the initially deprived groups was even larger than for the other patients, resulting in an inversion of the gap or a 'positive discrimination': for the measurement of BMI, the measurement of cholesterol and the control of BP a positive discrimination of South Asian patients with CHD compared to white British patients with CHD was described.[157]. Also for one indicator in diabetes care (serum creatinine recorded) the inequity inversed towards a pro-elderly distribution of the indicator [128].

In 2000, Victoria et al formulated the inverse equity hypothesis. This hypothesis proposes that affluent groups in society preferentially benefit from new interventions, leading to an initial increase in inequalities. Deprived groups only begin to benefit once affluent groups have extracted maximum benefit. Health inequalities ultimately diminish because deprived groups start with a lower baseline level of health and health care uptake and have higher potential gains [148, 235]. The above results do not unanimously confirm the first part of the hypothesis (i.e. just after the introduction of a new intervention the more affluent areas or groups in society benefit most).

With regard to the persistency of these changes over time only two studies were found [84, 142]. In the first year after the introduction a clear socioeconomic gradient was recorded, with progressively lower achievement and greater variation in achievement, with increasing area deprivation. However this gradient was not steep.

Both Doran et al. (2008) and Ashworth et al (2007) showed that after 3 years this existing (but small) gradient between deprived areas had almost disappeared. [84, 142] Moreover, using regression models including area, practice, patient and GP characteristics, Doran was able to prove that the increase in achievement over time was not significantly associated with area deprivation but was very strongly associated with previous practice performance: "the lower the achievement in the previous year, the greater the increase in achievement." [84]

This is a very important finding because it might indicate that the QOF indeed is a truly equitable public-health intervention since the improvements in quality achievement by practices are inversely related to previous performance and not to the level of deprivation of the area where the practice is located. However, alternative explanations for the described phenomenon could also exist: it is possible that the increase in quality already started before the introduction of the QOF (there are some indications for this) and that the better off groups already nearly reached their full growing potential by the time the QOF was introduced. [84] This might explain the reduction of some of the pre-contract health care gaps as described in the previous paragraph[84].

With regard to exception reporting there is some concern that this might be used as an excuse for substandard care of patients or to exclude patients for whom the targets had been missed, mostly socially deprived patients or patients with a different ethnic background, rather than because of a genuine clinical reason (also known as 'gaming'). The most recent and most comprehensive study that addresses this topic is the study of Doran et al (2008); they report that the characteristics of the patients (e.g. gender, socioeconomic position) explain only 2.7% of the variance in exception reporting. This does not confirm earlier studies with more limited study designs reporting that practices in financially deprived areas are more likely to exclude patients (McLean 2006). Doran et al (2008) conclude that "Exception reporting brings substantial benefits to pay-for-performance programmes, providing that the process has been used appropriately.

In England, rates of exception reporting have generally been low, with little evidence of widespread gaming" [84]. However, it can be argued that nevertheless the exclusion system succeeds in not being socially selective, it does not succeed in rewarding the additional work required in deprived areas [230].

In general, hopeful results were found. It can be states that after the introduction of the QOF at least some of the existing inequities became smaller and the positive effects seem to continue over the years. Still it is important to keep in mind that equity in health care is just a small piece of the larger jigsaw of determinants explaining inequity in health.

5.2.4 Revising the conceptual framework based on evidence

In this section not all of the findings as presented before in section 5.1 (page 71) will be repeated. Here only a summarized overview of the most important do's and don'ts is discussed. The same notation of levels of evidence is used ('S' = strong evidence, 'W' = weak evidence, 'C' = conflicting evidence, 'N' = no evidence). In the case of conflicting evidence or an absence of evidence, the recommendations are based on theoretical grounds.

5.2.4.1 Quality goals and targets

1. Take all SMART aspects into consideration when selecting targets (including relevance and timeliness). (S)
2. Measure potential unintended consequences (especially in care equity, patient experience and provider experience). (S)
3. Consider both appropriate and inappropriate care, as both cannot be separated. (C)
4. In short term, make use of structure, process, and intermediate outcome indicators. Each of these indicator types has their own value (e.g. IT adoption enhancement as a structural goal). (S)
5. Keep the number of targets feasible and transparent, but also sizeable within the full scope of delivering healthcare. (S)
6. Make use of a cyclical and dynamical quality improvement approach. (S)

5.2.4.2 Quality measurement

1. Make use of validated data already available as much as possible. (N)
2. Provide an audit system to prevent and detect gaming. (N)
3. Apply case mix adjustment on intermediate outcome measures. (S)
4. Apply exception reporting to guard individualized care. (W)

5.2.4.3 P4Q incentives

1. Make use of a non competitive approach (C). Budget equilibrium can be guarded alternatively by applying a corrective factor on all P4Q incentive payments, equal in size for all participants (N).
2. Make use of rewards. Punishments can be reserved for gross negligence (N).
3. Reward both best performers and best improvers (N).
4. Follow theoretical indications about a sufficient incentive size (about 10% of total payment), since evidence is still inconsistent (C).
5. Provide free choice to providers to use the incentive to invest in quality or to increase income (see QOF example) (S).
6. Align the complexity of the system with the complexity of healthcare delivery. Use more transparent and clear means to communicate the incentive drivers to providers. (S)
7. Weight targets in short term in function of related workload and according to target type (structure, process, and outcome). Add in long term the related cost savings to this equation. (S)

5.2.4.4 Implementing and communicating the programme

1. Base the first P4Q programme on new money. To keep budget equilibrium couple already planned budget increases to the P4Q condition. (N)
2. Implement P4Q using a phased approach. Make initial use of demonstration projects to avoid unpleasant surprises. Include both baseline and comparison group measurements. (S)
3. Make initial use of a voluntary programme. Ensure by sufficient involvement (democratic decision making) and a sufficient incentive size that the majority of providers participates. (C)
4. Do everything possible to support communication and awareness of the programme, especially in a direct and intensive way towards the participating providers. (S)
5. Provide P4Q as a package together with other quality supporting tools. (S)

5.2.4.5 Evaluation of the programme

1. Incentivize a specific target for a sufficiently long time period (based on a learning curve and clinical criteria concerning the effect interval). (S)
2. When a target performance plateau has been reached, focus on maintenance of the level of quality of care. Include other priority targets to redirect quality improvement resources. (N)
3. Sample regularly the performance on targets removed from the incentivized set. (N)
4. Evaluate the P4Q programme as a whole on a regular basis, using scientifically valid methods. (N)

5.2.4.6 Health care system and payer characteristics

1. Include only targets congruent with the health system and provider values. Ensure consensus. Make sure that the system fits with internal motivation and the non financial drive to provide healthcare. (N)
2. Provide one uniform P4Q system (in which local priorities may vary as targets) from all payers to all participating providers to support transparency, awareness and a sufficient incentive size. (S)
3. Integrate P4Q as one part of the healthcare payment system, with other incentive types. directed at complimentary goals (income security, patient and intervention volumes). (C)

5.2.4.7 Provider characteristics

1. Take into account the level of congruence with professional culture, but realize that P4Q may also support a cultural shift. (N)
2. Both when implementing and evaluating P4Q include the level of leadership support. The same is true for the history of engagement with quality improvement activities. (N)
3. Target incentives at least at the individual provider level, when he or she works in a larger organization. Combine individual incentives with team based incentives when appropriate (hospital setting). (S)
4. Be aware and take into account that provider age, gender, ethnicity, and training background will influence P4Q acceptance and performance. The same is true for the organization's purpose and structure (see medical groups versus IPAs in the USA), the age of the organization, the ownership (degree of resources available), the (non)teaching status, its geographical location and the number of providers within a practice or organization. (W)

5.2.4.8 Patient characteristics

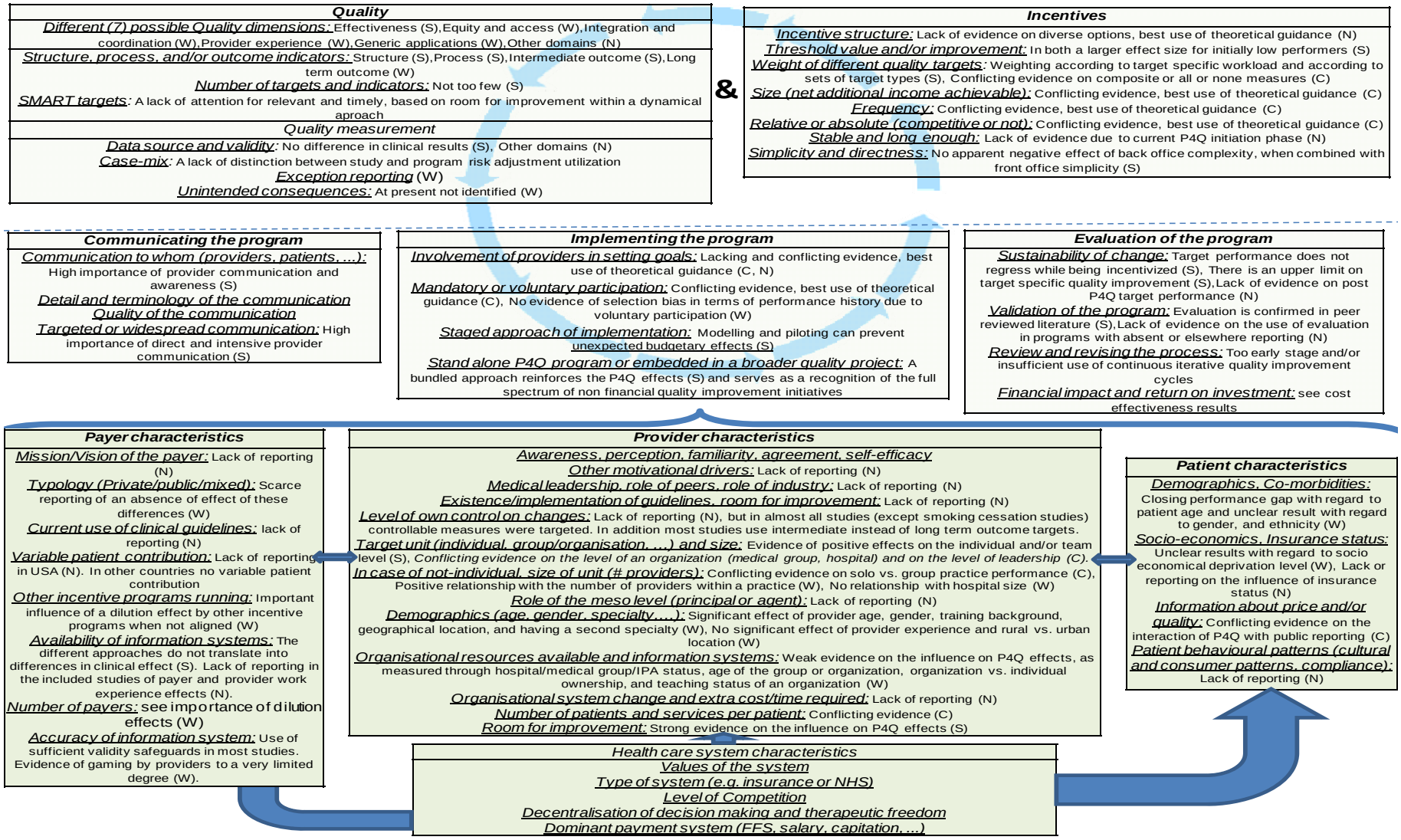
1. Monitor the effects of patient age, gender, socio-economical status, ethnicity and number of co morbidities on P4Q results. (W)
2. Take into account the experience of the patient as part of the targets, and during programme development, implementation and evaluation. (N)
3. Take into account patient influence when selecting targets and defining exception reporting criteria. (S)

5.2.4.9 Revision of the MIMIQ model

Figure 10 resumes the conceptual model that was presented in chapter 3. However in this version the results from the literature study are incorporated. For each item from the model, the strength of the evidence and the direction of the evidence are indicated.

However, it must be noted that for those items, where no evidence can be found yet, there might be still good theoretical reasons to take them into account when implementing a P4Q programme: no evidence does not necessarily mean no desirable effect. In Figure 10, this is formulated as ”best use of theoretical guidance”.

Figure 10: Revision of the MIMIQ model



6 INTERNATIONAL COMPARISON

6.1 INTRODUCTION

This chapter aims at answering the following research questions:

How is P4Q applied and how is it influenced by market, payer, provider and other healthcare system characteristics.

Four countries are taken into account, the USA, the UK, the Netherlands and Australia. For each country, semi structured interviews were conducted with key experts. (See chapter 2 for further details on the methodological approach)

From chapter 4 it appears that the majority of P4Q schemes are conducted in the USA and the UK. The Netherlands and Australia are two countries which are still in a starting phase of implementing P4Q, with only a few P4Q schemes operational. For the UK and the USA, two key experts per country have been interviewed. For the Netherlands and Australia, one key expert per country was interviewed. The results of these interviews will be reported in accordance with the topics of the conceptual framework as mentioned in chapter 3.

6.2 DESCRIPTION

6.2.1 Health care system characteristics

The health care system characteristics are essential determinants in the implementation of P4Q. An extensive overview of the health care system characteristics has been provided in chapter 4.

6.2.2 Existing P4Q interventions

6.2.2.1 The United Kingdom

Since 2004, quality has been introduced as a major part of the general practitioners remuneration in primary care in the UK. This occurred by means of the implementation of the Quality and Outcome Framework (QOF) in the UK health care system.[236] This framework has been described extensively in chapter 4. Participation in QOF is voluntary, however because of the high incentive, participation is high. Previous schemes, like 'the good practices allowance', which was launched in 1986, failed in their objective. At the time the medical culture was characterized by the idea that quality could not be measured and that there is no such thing as 'a bad doctor'. From that time on, the medical culture has changed in a sense that physicians and government began to recognize that the quality is not as high as we like it to be and that there is some variation between and within countries. High quality programme failures like 'the Bristol case', in which the death rate for congenital heart surgery in the Bristol Royal Infirmary was much higher than in other hospitals, was a kind of trigger to change medical culture. In the UK, P4Q programmes came together with already existing quality initiatives, hence physicians in the UK had already become familiar with quality assessments (for example: audit programmes).

In 1990, there has also been a small P4Q scheme related to the immunization of children and to cervical cytology, targeted at achieving 90% of children and 80% of eligible women respectively. Initially this scheme wasn't very popular but after a while it was accepted and a larger coverage for these 2 indicators was obtained.

Recently, a pilot P4Q hospital scheme in the north West of England, called 'advancing quality'[e], has been introduced, which is basically an attempt to replicate the 'Premier Demonstration project', a hospital P4Q programme which was implemented in the US (see next paragraph). This scheme is an answer to 'the next stage review' report, in which the surgeon lord Darzi, reasserted quality care being the key dimension in terms of where we want health care to go to.

---

[e] for more information see the following link http://www.advancingqualitynw.nhs.uk/

6.2.2.2 *The United States*

For decades, the US is characterized by many different P4Q schemes, using many different initiators (a more extensive description can be found in chapter 4) Consequently P4Q schemes are largely adopted by one payer at a time, although there are several regional multi-payer initiatives. The P4Q concept became popular around the time that the Institute of Medicine released its report 'Crossing the quality chasm' that indicated much underuse of Evidence Based Care and in which it was stated that payment systems should recognize quality. Large employers, which are the most important purchasers of private health insurance in the US, found the concept op P4Q appealing. Companies like General Electric and IBM have followed high-reliability manufacturing principles and supply chain principles in their own businesses and they are frequently trying to apply these principles to health care purchasing. Around 2000-2001 some big employers were organizing public reporting initiatives (the leapfrog Group) and others began to get interested in Pay for Quality programmes. Although some health insurance plans picked it up on their own, P4Q implementation in the US would not have happened without the support of the employers.

Many of the P4Q experiments in the US were performed in the private sector. Nowadays, more than half of managed care insurers, HMOs and PPOs are currently using P4Q schemes. Most state Medicaid programmes are also using P4Q schemes. Medicare, which is the most important payer in the US, has had a P4Q programme for hospitals, named the Premier demonstration project, based on which experience Medicare has the intention to launch a permanent P4Q hospital programme. According to the US experts there is an extremely weak P4Q programme for physicians, involving voluntary submission of quality data within Medicare.

According to US experts there is a link between P4Q and managed care in the US. In some places where managed care was never important, there is much less of an effort by payers to have any determination over what kind of health care services are delivered. Consequently P4Q schemes haven't been implemented that much in these places.

Currently there are more than 100 physician targeted programmes and about 40 hospital targeted programmes, most of them run by private health plans. Providers have multiple payers, each with their own P4Q schemes. This fragmentation makes it fuzzy for physicians to distinguish the quality indicators for each P4Q scheme and to modify behaviour accordingly.

In some metropolitan areas that have been dominated by managed care, new initiatives tried to align different P4Q schemes by different payers in one scheme, as for example in the Integrated Healthcare Association[f] (IHA) in California. The IHA has taken a lead in coordinating a P4Q programme with the attempt to line up all care, measuring the same thing and rewarding physicians and hospitals more or less in the same way. This has led to a uniform P4Q scheme with only one set of indicators despite all the different payers.[237] The programme is starting to expand the set of measures, not only including effectiveness measures but also efficiency measures.

It must be noted that P4Q is only a part of all the activities that are being used in the US to 'buy on the basis of quality'. One of the other activities is 'tiering', in which patients have to pay less for healthcare delivered by high quality and efficient providers and they have to pay more for healthcare delivered by low quality and inefficient providers. Likewise some health insurance plans offer 'narrow networks': instead of contracting with every physician, purchasers contract with a smaller set of physicians who have demonstrated better performance.

---

[f] For more information see the following link http://www.iha.org/

6.2.2.3 *Australia*

The last 20 to 25 years there has been a desire amongst Commonwealth Government in Australia to reduce the amount of money paid to general practices based on FFS and increase the amount of money that is paid at general practices on another basis than FFS. The government is trying to move away from simply rewarding production.

Regarding the medical services in primary care, there is a Pay for Quality programme named the Practice Incentive Program (PIP), which is organized at a national level and applicable for all accredited general practices.[238] The incentives in the PIP project are paid by the commonwealth government. The incentives in the Queensland project are paid by the state. The government conducted some modelling exercises to make sure there was enough money to pay for the indicators. In both programmes participation is voluntary.

Beside these P4Q programmes, 15 years ago a hospital accreditation programme was introduced in Victoria in which a small supplementary payment was provided. However this payment was not directly linked with quality.

A number of states have incentive programmes regarding waiting lists and waiting times in hospital services. Currently there is only one state (Queensland) that has a real Pay for Quality programme for hospitals (the Clinical Practice Improvement Program[g]). This P4Q programme was implemented as a result of a significant quality scandal in 2005 and aims at improving quality. About 16 quality of care indicators are included in the programme. Clinical networks, which are run by clinicians (cardiac network, mental network, etc) were involved in the development of these indicators. A number of health insurance funds have additional requirements on hospitals to undertake patient experience or patient satisfaction surveys, but hospital incentives aren't directly linked to these indicators.

6.2.2.4 *The Netherlands*

In the Netherlands P4Q has been initiated by the government. A first P4Q initiative was developed mid 1990's. General practitioners were paid for influenza immunization of their patients, and for carrying out cervical smears amongst their eligible patients.

Currently three Pay for quality initiatives in primary care can be distinguished in the Netherlands.

1°. The organization of care groups regarding certain disorders (for example diabetic care), in which participants are being paid based on the outcome on quality indicators.

2°. A second initiative is related to a practice accreditation programme of general practitioners. Participation in the programme, which consists of supplying information on clinical indicators, information on the organization of the practice and patients surveys, is being rewarded by a limited allowance. In the first year physicians get an accreditation for participation.

After the first year general practitioners have to establish an improvement programme. In the second year accreditation can be obtained if physicians have acted according their improvement programme.

3°. A third initiative consists of a bonus programme developed on top of the accreditation programme. In this experiment there are about 75 voluntary health care providers and 2 voluntary private insurers. To start this experiment a restricted set of clinical indicators has been selected, related to 4 chronic diseases, prevention and medication use. Moreover, some patient experience indicators and management indicators are included. All indicators are being evaluated every year, except the management indicators, which are being evaluated every 3 year. The incentives are being paid by the health plans. Participation is voluntary, but this will change. The moment a new payment system will arise, participation will become mandatory. Currently, it is not clear yet if health plans will work with one coordinated P4Q programme in the future or if each of them will have their own P4Q scheme.

---

g for more information see the following link http://www.health.qld.gov.au/cpic/

Concerning hospital care, several public reporting initiatives are been implemented. Currently no P4Q programmes are yet been developed[h].

Key points on interventions

UK

- **In the past several P4Q schemes have been launched in primary care. These initial P4Q schemes were limited in scope, but likely did improve cervical screening and immunisation rates over time.**
- **In 2004, QOF has been introduced successfully in primary care. This P4Q programme targets clinical indicators, patient experience and practice organization. Recently a pilot P4Q scheme, called 'advancing quality' has been implemented in the North West of England.**

USA

- **The USA is characterized by its many different P4Q schemes, in primary care as well as in hospital care. Large employers put pressure on private insurers to initiate the first P4Q schemes. Currently, Medicare and more than half of managed care insurers, HMOs and PPOs, are using P4Q schemes. In California the IHA has successfully introduced a uniform, multi-payer P4Q scheme. P4Q is only one of the activities used in the US to improve quality.**

Australia

- **Regarding hospital care there is only one P4Q programme, named the Clinical practice improvement programme that is implemented in Queensland. In primary care the Practice Incentive Program is introduced at national level.**

The Netherlands

- **Three P4Q programmes in primary care can be distinguished in the Netherlands. The organization of care groups, the accreditation programme of GPs, and a bonus pilot programme, which is developed on top of the accreditation programme**

6.2.3 P4Q concepts

In this section generic findings across the four countries will be discussed.

6.2.3.1 Quality

Quality Dimension

P4Q programmes are mostly initiated in response to the higher demand from government and insurers to deliver quality.

All the experts agreed that effectiveness should be measured in any P4Q programme. However, there are some other domains that could be of use. According to some experts the inclusion of deprivation measures in P4Q programmes could be interesting (equity). Especially in the US, deprivation measures are not always included in P4Q programmes. However, there has already been a small effort in the US to reward providers who serve low income and racial/ethnic minority populations more for the same level of improvement. Nevertheless, it must be noted that some experts believe that P4Q is probably not the right source for improving health for minority groups. Other kinds of investments e.g. in public health sector may be a better way to improve health in minority populations.

In the US there is also great concern about cost-containment. There exists a belief that not only quality but also efficiency gains have to be rewarded. To the extent that P4Q is being used to try to guide care towards more cost-effectiveness use, the underlying payment system will be determinative.

---

h For more information see the following link: www.kiesbeter.nl

For example in a health care system based on a fee for service system, it will be difficult to diminish for example the MRI use with a small bonus, because physicians can earn much more than the amount of the bonus by carrying out an unnecessary MRI. This is of course directly related to decisions on the size of the reward. Concerning cost-effectiveness one US experts fears that people will find quality not that important as cost. In Australia and also in the Netherlands more recently, there is also a great focus on timeliness. There are financial incentives to reduce waiting time for patients for surgery and for waiting times in the emergency room.

Quality Indicators

In most P4Q programmes mainly process measures are been used, as these indicators can be influenced within small periods of time and within limited time frames. It is also believed that process measures are in the control of physicians. According to one US and one UK expert it is important to have outcome measures because achieving a process measure does not necessarily result in a better health outcome. Intermediate outcome measures would be adequate, because these measures can be linked to certain hard outcomes. Others (UK, NL) believe process measures are sufficient to measure performance, on the assumption that improved process measures will lead to better health care outcome.

Current programmes initially target underuse. However a lot of health care systems are confronted with overuse and misuse, therefore targeting this in P4Q programmes can be useful. Recently the measures for overuse and misuse are rising, and the focus within P4Q programmes is changing from underuse to overuse.

In the UK exception reporting is allowed, this enables providers to exclude individual patients from the calculations for specific targets, because there was a valid reason for not reaching the target in that individual patient, which was not related to quality of healthcare. Most experts agree that it would be useful to include exception reporting in a P4Q programme.

Quality measurement

In the UK, clinical indicators are automatically extracted out of the electronic health records by the government.

In the US, P4Q data are largely based on billing data. Physicians complain that billing data ("claims data") aren't rich enough to capture quality. However, according to experts, billing data are sufficient for process measures. For outcome measure on the contrary, clinical data is necessary.

In Australia the data collection system is very transparent and there seems to be sufficient trust among health care providers. Some of the data are collected manually. Concerning the PIP programme in general practice, data is provided to Medicare Australia. Related to the data collection within this programme there have been complaints concerning the different computer software programmes and about the amount of paperwork.

In the Netherlands, electronic health records are available, but not all indicators included in the P4Q scheme are comprised in the files. Physicians need to complement these data with other data that have to be introduced manually in the data system. Data are collected at the practice level and send to health insurers by GP organisations for additional payments.

According to the experts, collecting data manually is too time-consuming. The experts agree that the data should ideally come out of the work flow through electronic health records. All were convinced that data collection just for the purpose of P4Q is not a desired way of collecting data.

Risk adjustment

Experts believe that risk adjustment, where the provider's case-mix is taken into account, may not be necessary if one focuses on process measures, and that it is only necessary if the focus is on outcome measures.

6.2.3.2 Incentives

Size

Different incentive structures are being used. The UK makes only use of *financial rewards* and approximately 25% of the practitioner's income is generated by the QOF. This large investment has led to improvements in quality. However it remains unclear whether the quality improvement seen to date was worth the investment.

The US makes use of both *financial rewards (bonuses) and withholds*. In the US the incentive size ranges between 2 and 10% of the payments, however only 10 to 20% of the payers are involved, consequently 1 or 2% of provider's income are generated from P4Q schemes, leading to a dilution effect. Another example of an incentive is based on the difference in performance measure between different health plans in counties with more than one health plan, as assessed by Medicaid. Initially, when there was not as much money to pay bonuses, new Medicaid enrolees who did not select a health plan would be assigned to a better performing health plan at a ratio that reflect the quality of that plan in comparison to other plans. This is a financial incentive in the sense that there is financial gain for the organization having people come to their plan. *Withholds* are mostly used when programmes have to be budget neutral. The difference between a bonus and a withhold is somewhat semantic in the US. A withhold mostly consist of giving only a part of the money physicians or hospitals normally get. The remaining part is only being given if physicians or hospitals performing well. It can be interpreted as a withhold in the sense that physicians and hospitals were expecting to get that money. A withhold can also consist of 'shared savings', for example to reduce overuse. When healthcare overuse can be reduced, costs and consequently the physician's income will also diminish, subsequently this saved amount of money is being divided between government and physicians, so physicians lose only half of the original reduction of income. Finally, another kind of withhold is being used in a hospital care P4Q related system by CMS concerning preventable complications. When hospitals submit their bill (consisting of all diagnoses a patient has had at discharge), the potentially avoidable complications can't be included in the list of diagnoses. In this context it should be noted, that it is advised to develop audit systems to select fraud.

In the UK, the P4Q system is not competitive. In the US on the contrary, P4Q systems are often competitive. According to one UK expert there is no justification for competition, rewards should relate to the levels of performance.

In the Australian P4Q programmes discussed above, incentives are presented as *financial rewards*.

In the Netherlands the P4Q pilot programme also only make use of *bonuses*, although as a consequence of the limited budget, insurances have difficulties to pay out these bonuses. As a result of the new insurance law a new payment system based on the delivered quality is in development. It is not yet clear how this new payment system will be organized and if the modalities of the above discussed pilot project will be incorporated.

Some experts (US, UK, NL, AUS) argue that the incentive size in the US is probably too small and the size in the UK is too high. According to the experts the ideal incentive size should range between 5% and 25% of physicians income. An incentive that is too high could provoke gaming effects, an incentive that is too low on the other hand could limit the impact in terms of quality improvement.

Target

In the UK practices are owned by groups of GPs and QOF incentives are targeted at the practice level. The GP team can decide to invest the incentive in the practice and hence trying to improve quality even more or to divide the incentive between GPs. The hospital P4Q demonstration project which has been implemented recently in the North West region of the UK, targets its incentives at the department. In the US, team care almost doesn't exist, therefore payments are often targeted at the individual. However, in the Californian IHA for example, project payments are made to very large multidisciplinary medical groups.

Individual physicians do not receive a bonus, but the group bonus may serve for investing in the enhancement to the electronic health record, in case managers, etc. In that sense, team care can be a better option. In the Queensland hospital P4Q project in Australia, incentives go to the unit. In the Netherlands, incentives are targeted at the practice level like in the UK. Most experts share the opinion that ideally, incentives should be targeted at the provider unit or the practice group, and this gives the group or the unit the possibility to invest in the practice (for example improving the IT system). Incentives that are given on a higher level could create a moral hazard problem because the individual incentive is less important and therefore the stimulus to perform well dilutes.

Key points on P4Q concepts

Quality

- **Experts agree that there are several other domains that could be of some use to measure besides effectiveness, like equity, cost-containment and timeliness. In most countries process measures are used. According to the experts, process outcomes and intermediate outcome measures are adequate measures to control quality.**
- **Concerning the collection of the data, they all agree that, data should be extracted automatically out of the EHR.**

Incentives

- **In the UK, Australia and the Netherlands only bonuses are used in the P4Q schemes. In the US both bonuses as withholds are used. In the UK financial rewards generate approximately 25% of GP's income. In the USA, only 1 to 2% of provider's income is generated from the P4Q schemes, because of the heterogeneity in the use of P4Q schemes.**
- **Experts agree that the incentive size should range between 5 and 25%.**
- **An incentive that is too high could provoke gaming effects; an incentive that is too low on the other hand could limit the impact in terms of quality improvement.**
- **Most experts share the opinion, that ideally, incentives should be targeted at the provider unit or practice group, this gives the group the possibility to improve quality by investing in the practice or organisation. Incentives that are given on a higher level could create a moral hazard problem.**

6.2.4 Implementing and communicating the programme

6.2.4.1 Involvement of providers

The introduction of the QOF system in 2004 was a result of an 18 months lasting negotiating period between the government, the British Medical Association and academics. It was a consensual process that went trough a vote. The targets set in the first QOF version were easily reached. The QOF scheme is large and complex and was never intended to stay static. It is a process of updating and adapting indicators which has to be done by academics in association with the British Medical Association and the government. As a consequence in the most recent revision, indicators that weren't worthwhile were deleted or changed after negotiation. This process of indicator development is to be taken over by the National Institute for Health and Clinical Excellence (NICE).

In the Netherlands, the P4Q scheme is also a result of a negotiation process between all stakeholders, and the indicators were chosen in dialogue with the physicians. In Australia, clinical networks, which are run by commissions (cardiac network, mental network, etc) were involved in the development of these indicators.

In the US, with all the different schemes, the degree of negotiation between insurance companies or government and providers depends on the scheme. In the development of the IHA programme in California the physician organisations were involved from the very beginning. The inclusion of these physicians was very much part of the reason that the initiative got off the ground.

The experts agree that it is important to include providers in the implementing phase of a P4Q programme and to involve them in selecting the indicators. In fact, it is essential that the indicators have clinical credibility, that they have to be developed in consent with relevant clinicians and that they should be based on evidence based criteria.

6.2.4.2 Acceptance by professionals

In the UK, participation in the QOF scheme is quite high, probably because physicians were largely involved in the negotiation process, because the targets are easily reached by the majority of GP's and because they can only gain additional income and can not loose money. According to the UK experts, such a loss of income would never be accepted by the professionals.

In US acceptance was mixed: some physicians were supportive and became interested in making the most of P4Q by helping to develop good measures, while others were less supportive, because US programmes involve frequently a withhold and physicians feel that they are already underpaid.

In Australia there is an interplay of forces between 3 groups. The Australian medical association is generally opposed to P4Q as it supports heavily FFS arrangements. The royal Australian college of GP's, which is a professional organization, is more supportive to these quality payments. The Australian general practice network, which is another GP organization network, is very supportive.

The pilot P4Q programme in the Netherlands included only physicians who were supportive. It is difficult to predict the reaction of the majority of physicians when the scheme becomes nationally implemented.

Experts agree that there remains substantial concern about P4Q schemes among groups of professionals. There is still a very polarized view of physicians, some are very opposed, and others are very supportive.

6.2.4.3 Mandatory or voluntary participation

As indicated above, participation in programmes can be mandatory or voluntary. The UK and Australia make use of voluntary P4Q programmes. In the Netherlands, participation in the pilot project is currently voluntary, but most probably participation will be mandatory in the future, with new arrangements in the health care system. In the US both mandatory as voluntary programmes can be identified.

Some experts believe that participation in P4Q programmes, which make use of bonuses, should be voluntary. If you make these programmes mandatory, some participants will not make the effort to score well. However some experts state that data collection should be mandatory, to allow benchmarking within the group.

Key points on implementing and communicating the programme

- **It is important to include providers in the implementing phase of a P4Q programme because it leads to greater acceptance.**
- **Currently there is still a polarized view of physicians concerning P4Q.**
- **Most experts believe that participation in a P4Q programme should voluntary, although according to some experts it can be important to make data collection and hence participation mandatory to allow benchmarking within the group.**

6.3 EVIDENCE

6.3.1 Likely effect according to the experts

For the UK there is clear evidence that quality has improved somewhat as a result of the introduction of P4Q. However in 2005 further improvement slowed down. In the US there is probably a modest positive effect (1-2%) but not across all programmes. Because most P4Q programmes are natural experiments without a control group (US, UK, Netherlands), it is unclear if the effect can be assigned to P4Q alone. There are a lot of co-interventions like public reporting etc. Furthermore it is mostly unknown in what direction not incentivized quality indicators have evolved, an underlying trend of improving quality is assumed by most of the experts. In the Netherlands the first results show an improvement on clinical indicators (around 10%) and on patient experiences (around 5%), here too it has to be noted that there was no control group and moreover there were some co-interventions like accreditation.

All experts agree that P4Q programmes are no magic bullets. Overall it seems that financial incentives are modestly effective. P4Q programmes could have some value it they are organised in the right way, however, these programmes should be seen as part of a range of quality improvement initiatives. Other quality initiatives can be: investing in quality training of physicians and learning them to be more pro-active, investing in electronic health records and transforming practices to be more efficient (e.g. financing IT investment). In the US, Kaiser Permanente, which is a medical group, has moved down this path quite successfully, with physicians support, and with Electronic Health Records.[239] For instance, when scheduling an appointment the system prompts the physician for conducting recommended tests, and it sends lab results electronically to the patients. There is a high focus on standardizing the delivery of care, and hence reducing variability.

In a sense P4Q is a response to a lack of patient activation and patient recognition about what kind of care they should demand. Consequently, insurers and governments should do well to figure out ways to make patients more engaged in demanding effective and cost-effective care for themselves.

6.3.2 Unintended consequences according to the experts

Until now, not much evidence of unintended consequences was reported. However, there is some evidence that equity is probably improving in the UK, i.e. the gap between deprived areas and not deprived area rose until 4.5 to 5% in the first year of P4Q and has now narrowed to 0.5%. However is it hard to asses whether this is a result of P4Q or caused by a secular trend. There is also evidence that inequalities in chronic disease management between ethnic groups have not been attenuated. Additionally, there seems to be no negative effect on access and patient experiences.

Yet, according to the experts there are some topics which should be monitored closely in any P4Q programme. These topics include:

Quality indicators that weren't incentivized are likely to get less attention and may have gotten worse.

Absence of a control system could provoke gaming of the system, especially when the incentive size is quite high. It must be noted that gaming is a real threat, and even a control system might not always capture those who game the system.

It is assumed by some experts, that difficult patient groups are probably concentrated in some practices. Those practices get paid less overall and end up with fewer resources. This can be a paradoxical effect of P4Q.

Big and powerful physician groups are probably better in negotiating better prices, leaving individuals and small groups to lag behind.

Some experts fear some reduction on continuity and coordination of care. There is a tendency to take care out of routine care into special clinics (e.g. diabetes clinics) which effective method to make sure all checks are done. Consequently, the patient potentially sees more people in practice.

It must be noted, that the QOF system, which consists of one coordinated programme, makes innovation difficult. The included indicators in a P4Q programme make it difficult for providers, to implement for example new treatment methods or new chirurgical methods. In US on the contrary, all the different P4Q schemes make developing and implementing innovations easier.

Key points on evidence

- **According the experts financial incentives are modestly effective in improving quality. They all agree that P4Q is no magic bullet; however P4Q programmes could have value if organised and implemented in a correct way.**
- **The programmes should be seen as part of a range of quality improvement activities.**
- **Concerning the unintended consequences until now not mush evidence was reported. There is some evidence that equity is probably improving, however it is hard to asses whether this is the result of P4Q or caused by a secular trend. According to the experts monitoring of unintended consequences remains important.**

6.4 DISCUSSION AND CONCLUSION

6.4.1 Key recommendations for design and implementation

To successfully implement a P4Q programme, peculiarities of health system need to be taken into account. The context in which P4Q is being introduced is very different in different parts of the world, so this makes it hard to generalize. The experts suggested the following recommendations for a country at the first initial stage of considering the implementation of P4Q:

Quality goals and targets

1. Be clear about what the priorities and objectives are. It is important to consider which domains to include in the P4Q system: it should be feasible, payable, and lead to better quality. For example financial incentives may be not appropriate in the patient safety domain This domain is probably best covered by for instance critical incident analysis, when things went wrong
2. Indicators should be derived from evidence based criteria and the health care providers should be included in setting the indicators. Chose indicators where there is still room for improvement;
3. Include different quality domains in the P4Q programme: effectiveness, deprivation measures, timeliness, cost-effectiveness
4. Do not only focus on underuse but also on overuse in health care;
5. Include a sufficient amount of measures. To few measures can lead to "teaching to the task" (physicians only focus on incentivized indicators and ignore the unincentivized indicators);
6. Strive to include process parameters with a clear and proven link with outcome;
7. Organisational indicators are of limited value although some could be important, e.g. if one is aware of problem areas with major problems, e.g. it is impossible for disabled people to get access to surgeries because they are only to reach by means of the staircase;
8. Measure unintended consequences.

Quality measurement

1. Establish an audit system;
2. No one has ever suggested that clinical guidelines should relate to all patients, so allow for an exception reporting mechanism like in the UK;
3. If possible, make use of available data collection systems to measure quality. An unnecessary increase in administration workload should be avoided.

P4Q incentives

1. The ideal incentive size should range between 5% and 25%, although it seems that some P4Q programmes with a small incentive can also induce a striking effect;
2. Find a balance between rewarding high achievement and rewarding improvement (There is an argument that payment should be related to improvement, that would give more incentives to low scoring practices);

Implementing and communicating the programme

1. It is important that government and clinical leadership recognize that quality is variable and improvement is important;
2. Include government/ insurers as well as the health care providers and academics from the start in the negotiation process to implement a P4Q programme;
3. Invest in IT development and make data collection automatically. This makes participation less time consuming and it makes gaming more difficult. P4Q can be seen as an opportunity to promote the use of IT and electronic health records;
4. Make use of a phased approach. For example start in a certain region, start with a limited set of indicators or implement an adaption year in which participation is being remunerated;
5. Allocate a well defined amount of money to the development and implementation of a P4Q programme;
6. Make sure that health care providers who will be subject of P4Q have adequate information about their own performance and adequate support for quality improvement;

Evaluation of the programme

1. Examine unintended consequences and think about how schemes could be developed to maintain/improve equity;
2. Measure your baseline first (in the UK, the first targets were easily reached because baseline wasn't measured properly. This caused the government financial embarrassment).

Health care system and payer characteristics

1. Try to create a uniform P4Q system which is applicable for all physicians (not like in the US with its diversity of schemes and the payer fragmentation problem, where physicians often don't know what targets should be achieved in which programme);
2. It is important to recognize that P4Q is not a magic bullet. P4Q programmes could have some value it they are organised in the right way, however, these programmes should be seen as part of a range of quality improvement initiatives.

Provider characteristics

1. Incentives should be targeted at the provider unit or the practice group. Incentives that are given on a higher level could create a moral hazard problem.

Patient characteristics

1. Monitor the effect on unintended consequences concerning patient characteristics.

6.4.2 What the future holds

All experts are convinced that P4Q will continue to exist in their country. The Australian and Dutch experts think that P4Q in their country, which is currently still in a starting phase, will gain importance.

They all agree it is important to attach research to the introduction of P4Q schemes. On the whole the UK experience, in terms of a modest improvement in quality, is probably consistent with US experience and other countries' experiences, but ongoing evaluation is necessary on the following items:

- There is a need for research on the optimal size of the incentives and whether a bonus or a withhold or a combination is desired, on the level to which incentives are paid (individual, group, organisation) and on rewarding high achievement or rewarding improvement to improve care and to reduce variation;
- There are still some questions about impact on patient experiences. It is advised by the experts to focus more on patients and less on providers;
- It is important to do more research on unintended consequences, risk adjustment, exception reporting and equity;
- The development of new sets of indicators is an ongoing process (process as well as intermediate outcome measures);
- It is unsure what the effect of P4Q will be in the future. More research on the permanent impact of P4Q on quality should be carried out;
- In most P4Q programmes more quality measures focused on primary care (e.g. vaccination target, cervical smear target, diabetes targets) and less on specialty care. It is a methodological challenge to develop specialty care indicators for specialist and hospital care.

6.4.3 Conclusions

Most recommendations made by the experts are consistent with the ones resulting from the evidence (see 4.4.3 Revising the conceptual framework based on evidence). Only a few additional issues were cited by the experts. They agree that before considering implementing P4Q programmes, the government as well as clinical leadership should recognize the importance of quality and the variability of quality between physicians. Later on, all stakeholders (insurers, government, health care providers, and academics) should be included from the start in the negotiation process of implementing P4Q. Furthermore the expert draw the attention on the fact that indicators should be derived from evidence based criteria and that quality should be targeted on those indicators that show a lack of quality. In addition, the providers have to be involved in setting the indicators. The experts advise to strive to include process parameters that have a clear and proven link with outcome. When setting the indicators, not only underuse but also overuse should be targeted. Finally it is important toe recognize that P4Q should be seen as one of many different quality initiatives, only a combination of initiatives could lead to a quality improvement in the health care system of a certain country or region.

Overall conclusion by the experts on quality improvement is that financial incentives are modestly effective. Experts agree that P4Q programmes are no magic bullets, however they can be of value, when organised in the right way. These programmes should be seen as part of a range of quality improvement initiatives. Before developing a P4Q scheme it is important to stipulate those areas where quality improvement is desirable. Physicians should be involved in the developing process and in setting the indicators. Experts agree that it would be wise to make use of a phased approach.

For collecting the P4Q data it would be advantageous to invest in IT development and to make use of a system in which data are extracted automatically. Despite the fact that until now, there hasn't been any evidence on unintended consequences, experts agree that further investigation into unintended consequences is desirable. Concerning the indicators, experts suggest finding a balance between rewarding high achievement on quality indicators and rewarding improvement. The ideal incentive should range between 5 and 25%. An incentive that is too high could provoke gaming effects; an incentive that is too low on the other hand could limit the impact in terms of quality improvement. Finally it may be advantageous to create a uniform system, in which indicators are the same for all physicians

Key points on discussion and conclusion

- **The following most important recommendations are made by the experts: be clear about what the priorities and objectives are, include all stakeholders in the negotiation process, invest in IT development and make data collection automatic, make use of a phased approach, find a balance between rewarding high achievement and rewarding improvement, examine unintended consequences, develop other quality improvement initiatives to complement P4Q schemes**
- **All experts agree that P4Q programmes can be of value when organised in the right way. They are convinced that P4Q will gain on importance in the future but they all agree that ongoing evaluation is necessary.**

7 P4Q IN BELGIUM

7.1 DESCRIPTION OF (PAY FOR)[i] QUALITY INITIATIVES IN BELGIUM

7.1.1 Introduction

Chapter 7 transfers knowledge from previous chapters towards the context of Belgian healthcare. The overall study objective is to assess the feasibility, advantages and disadvantages of P4Q implementation in Belgium.

Previous chapters already extensively clarified that health system, payer, provider and patient characteristics are of a significant influence on how P4Q might or might not reinforce quality of care within a specific context. In addition, the design of components of the central quality circle, which is the core of any P4Q initiative, also depends on what is already in place in Belgium in terms of quality support.

Based on reports such as provided by the European Observatory on Health Systems and Policies (2007) [240], the health system, payer, provider and patient characteristics in Belgium can be described as followed:

Healthcare jurisdictions in Belgium, as a federal state, are divided over a national and regional[j] level. The federal government is responsible for the regulating and financing of the compulsory health insurance, determining accreditation criteria, financing hospitals, legislation covering professional qualifications, registration of pharmaceuticals and their price control. The three regional governments are responsible for health promotion (prevention), maternity and child health services, different aspects of elderly care, the implementation of hospital accreditation standards, and the financing of hospital investment. In terms of jurisdiction especially the tools supporting cost containment are managed at the federal level.

The Belgian health system is primarily funded through social security contributions and taxation. Public sector funding as a percentage of total expenditure fluctuates around 70%.

The six non-profit non commercial Belgian sickness funds, acting as care purchasers, receive a prospective budget from the National Institute for Sickness and Disability Insurance to finance the healthcare costs of their members.

The dominant payment system of medical providers is Fee For Service, although the amount of fixed payments is increasing (e.g. for the use of medical imaging, clinical biology and certain drugs as part of hospital care). Capitation is only used to reimburse a few primary health care centres. Specialists and the majority of GPs are paid on a FFS basis. Hospital accommodation services, nursing activities and emergency services are financed via a fixed prospective budget system based on diagnosis related groups.

GPs and specialists work mostly as private self employed independent contractors, with the exception of some GPs in primary health care centres and specialists in university hospitals who are salaried (<1%).

---

i In this chapter's title 'Pay for' has been put between brackets, because, as becomes clear in the chapter results, most current quality initiatives in Belgium do not comply with the study's definition of P4Q. Even programmes using financial incentives do not relate these directly to the performance on quality target measures, except one initiative.

j Apart from the federal state, three regions and three communities can be distinguished in Belgium. The regions are based on the name of the geographical territory they represent, the communities are 'language'-based. Belgian has three communities: Flemish community - Vlaamse Gemeenschap - Communauté flamande; the French community - Franse Gemeenschap - Communauté française and the German community - Duitstalige Gemeenschap - Communauté germanophone. However, within this report the term 'regional' will be used whenever we are talking about these three communities.

Belgium is a country with a high access level of patients to healthcare and almost universal compulsory health insurance coverage (99%). Currently the level of gate keeping is minimal. Patients are free to chose which provider (GP, specialist, hospital) they consult, without restrictions. The basic right to health care has been set out in the Belgian Constitution.

Patients participate in healthcare financing via co-payments, for which the patient pays a certain fixed amount of the cost of a service, and via co-insurance, for which the patient pays a certain fixed proportion of the cost of a service (10 to 40%). A maximum of out-of-pocket expenses is safeguarded.

Competition in the Belgian health system operates mainly through the competition for patients between providers and the competition for members between sickness funds.

A typical characteristic of the Belgian health system is the participation of several stakeholders in its management. An important part of the health system is regulated by national collective agreements made between representatives of health care providers and sickness funds.

Since the mid 1990s attention for the quality of healthcare in Belgium has increased. Quality control is applied by the government through accreditation of healthcare institutions (Hospital Act), accreditation of providers, peer review, audit and visitation. Hospitals report to the federal government (colleges of physicians) on their quality status on a yearly basis. The regional government of Flanders requires healthcare institutions to implement a quality manual, plan and evaluation cycle as part of accreditation conditions. In addition, specific campaigns address issues such as hand hygiene, antibiotics use, etc.

Based on the conclusions and recommendations of the previous chapters the Belgian healthcare system, payer, provider and patient characteristics seem to fit with P4Q in the following way:

There is strong evidence (next to international expert support) of the importance of the uniformity of the P4Q design, independent of the identification of P4Q initiators and coordinators. In this respect the concurrent levels of regional and national healthcare jurisdictions may pose a threat. The same risk is present at the level of the six sickness funds as purchasing/payer organizations. They are however united as part of RIZIV representation. With regard to P4Q implementation in Belgium a clear consensus based, all stakeholder inclusive approach is recommended, including complementation of the regional and federal level. This enables a combined centralized and decentralized (e.g. local priority setting based on room for improvement) decision making process. Multiple incentive programmes at the payer level should be aligned.

With regard to the level of competition the low level of patient volatility in Belgium will support P4Q on the one hand. P4Q is not reinforced by consumer driven healthcare on the other hand.

Incentive theory states that P4Q targeting underuse fits with a dominant FFS payment system. It's therefore a reinforcement of the P4Q effect on this type of targets. Vice versa it may pose a threat to the P4Q effect on overuse targets. However, many of these overuse targets (medical imaging, clinical biology, etc.) are increasingly reimbursed as fixed payments, which diminishes this threat.

In terms of provider characteristics specifically for hospital care the current focus of legislation, financing and quality management at the organizational level may pose a threat to P4Q targeting individual providers and/or team members. Such direct targeting in Belgium requires specific attention in P4Q regulation.

Finally, with regard to patient characteristics the use of case mix adjustment at various levels of the Belgian healthcare system can be leveraged. However, more specific attention for the patient's role is needed (equity, knowledge and involvement, experience of care, behavioural patterns, etc).

What kinds of P4Q applications currently are available in Belgian healthcare is not clear. Nor is it currently known how current components of the central quality circle might fit with P4Q development, implementation and evaluation. Knowing this would further answer the question about practical feasibility of P4Q in Belgium.

Therefore chapter 7 consists of two main parts. Section 7.1 focuses on the content of current quality and pay for quality initiatives in Belgium as a starting point. It offers a description of various quality initiatives and the extent to which they fit with a P4Q definition. Section 7.2 studies feasibility by positioning quality initiatives within the empirically revised conceptual P4Q framework. This allows the identification of current strengths and weaknesses of Belgian quality support and how this translates into opportunities and threats for P4Q implementation.

Chapter 7 addresses, together with chapter 8, the following research questions:

What are the conditions to apply international P4Q models or to enlarge Belgian quality improving experiments?

Design, implementation, evaluation. **What are the current initiatives in Belgium in the public and private sector to enhance quality of care by way of financial incentives linked to quality of care measures? Is there any evidence of their impact on quality?**

Conditions. **To what extent are the current financing scheme, databases and other tools (guidelines, quality indicators) appropriate to implement P4Q in the Belgian health care setting? What are the most important facilitating and hindering factors?**

7.1.2 Methods

The output of the systematic review on P4Q (see chapter 3 and 4) nor an additional search strategy using Dutch and French entry terms within the predefined search string identified any Belgian study.

These limitations highlighted the need to collect the information as broadly as possible on a local scale by direct contact of all stakeholders involved. Because interviews of forty persons were planned as part of the chapter 8 data collection, these meetings were used as an opportunity to seek additional information on existing (P4)Q initiatives. They were asked to refer the research team to other persons involved in (P4)Q initiatives to provide additional information. These persons were contacted by phone and email, using a standardized template to collect (P4)Q initiative data.

Numerous initiatives, programmes and applications were mentioned by the stakeholders and additional contact persons. However, the main bulk of information did not concern programmes which complied with the study P4Q definition. A number of programmes did involve a kind of financial incentive. But this incentive wasn't in any programme *directly* related to the measured performance of participants with regard to predefined quality targets. In this section we provide however an overview of Belgian quality initiatives that were mentioned by the experts. These programmes provide possibilities to develop pay for quality initiatives, as will be further analyzed in section 7.2.

7.1.3 Results

A total of 14 programmes are described below. These most often mentioned programmes are:

Primary + hospital care

1. Care itineraries ("trajets de soins", "zorgtrajecten")
2. Clinical pathways
3. Providers' accreditation

Primary care

1. Breast cancer screening prevention bonus
2. Capitation funding revaluation in the primary health care centres (MM, WGC)
3. EPA tool
4. Global medical record (DMG/GMD)
5. Prescription feedback of the National council for Quality Promotion (CNPQ/NRKP)
6. Preventive module in global medical record

Hospital care

1. Centres of reference
2. Hospital accreditation
3. Hospital benchmarking
4. Quality and Patient Safety Framework for hospitals
5. Reference payment

These programmes are presented in a structured tabular format, addressing programme initiation, implementation, target audience, content, primary focus, type of indicators used, purpose of indicators used, type of incentives, results, publications and budget (if publicly available).

Table 7: Care itineraries

I. CARE ITINERARIES	
Initiated by	RIZIV/INAMI. In 2001 RIZIV/INAMI took the decision to develop specific chronic care programmes, combining the contribution of GP, specialists and other disciplines in a scientific and coordinated way. These transmural programmes were called "Care itineraries / Zorgtrajecten / Trajets de Soins". From 2003 till 2007, field tests took place in Leuven and Aalst. The practical organization will start from June and September 2009 with two defined care itineraries: on chronic renal failure and diabetes mellitus type 2. Other chronic care fields are envisaged for the future: COPD, Chronic Heart Failure, Frail Elderly, osteoporosis, depression, cancer …
Overall objective	To reinforce the continuity of care between patient, GP, and specialist and tackle underuse mutually
Date of implementation in Belgium	June and September, 2009
Target audience	Partnership between patient, GP and topic specialist around specified chronic conditions (first itineraries are diabetes and chronic renal insufficiency). Focus is on the patient with chronic conditions. The aim is to optimize the quality of the follow-up, and the outcome parameters, defined for the condition. Individual targets are defined, and an optimum treatment scheme together with an ideal follow up scheme is planned. Supporting disciplines like educators, home nurses and dieticians are brought together in a local primary care team. Local pharmacists are invited to play their role.
Content	Individual contract with patient (involvement in reaching targets), GP (coordination) and specialist (support, continuing education and updating) Installing local multidisciplinary networks to support professional dealing with chronic conditions. 1. At the individual level, the patient, entering the care-itinerary and declaring to do the best of his/her possibilities, is completely reimbursed for all contacts with their GP and the topic-specialist. Some supporting disciplines are reimbursed following a defined

	quality scheme. The GP, as care coordinator, and the specialist for support earn 80€ per patient per year. 2. At the quality content level, the National Council for the Promotion of Quality defined the inclusion and exclusion criteria, the optimal follow up and the average targets in line with international guidelines and options. From 2011 onwards, outcome data will be collected electronically to evaluate the evolution over time, and to judge the impact of the system. 3. At the local multidisciplinary network level regions for implementation of chronic care itineraries will be defined, care managers will be installed, and the local GP's groups, the local integrated homecare services (GDT/SISD) and the specialists covering the area are brought together. Pilots are contracted for 4 years, and granted 93.000€ per area (between 75 000 and 186 000 inhabitants). Data will be aggregated per local area to support feed back and benchmarking.
Primary focus (structure, process outcome)	Reaching individual defined targets (in line with international targets), by a set of processes (minimum obligatory set), supported by rearranging and re-tasking of existing structures adding 2 new functions: the primary care educator and the care itinerary manager
Type of indicators used (if any)	Set of outcome parameters (from 2011). For the different diseases, some of the crucial targets are considered as quality indicators. For diabetes the choice was made to use HbA1c, systolic BP, LDL cholesterol, and BMI as indicator. With these indicators, the message is strengthened that diabetes is a complex metabolic syndrome that include cardiovascular risk management, and that weight control is important. For Chronic Renal Insufficiency, eGFR, blood pressure, Hb, Creatinin, and parathormone are selected.
Purpose of indicators used	Evaluate the global programme and provide regional data to networks for benchmarking.
Type of incentives	Annual supplementary capitation fee (80€) for GP and for SP, no co-payment for patient, defined fee for educator, financial support for local multidisciplinary networks (93 000€/ 125 000 inhabitants). Balanced package of annual care-payments for GP's and specialist, omitting all financial constraints for patients, enhanced payment for upgrading of home nurses and dieticians as educators, and important introduction of an opportunity for a management function to support primary care as an organizational structure.
Results (if available)	From 2011 on
Sources (most relevant ones)	International publications of preparative pilots 2003-2007 in Leuven and Aalst, budget evaluation report end 2009. [241, 242]
Total budget	€30.7 million

Table 8: Clinical pathways

II. CLINICAL PATHWAYS	
Initiated by	Center for Health Services and Nursing Research (CZV-KULeuven), Université Catholique de Louvain in Belgium and the Kwaliteitsinstituut voor de Gezondheidszorg (CBO) in the Netherlands
Overall objective	To improve and maintain quality of care in a balanced way with attention for all relevant quality dimensions
Date of implementation in Belgium	April 2000
Target audience	Hospitals, Rehabilitation Centers, Primary Care Organizations
Content	A care pathway is a complex intervention for the mutual decision making and organization of care processes by hospital providers and policy makers for a well-defined group of patients during a well-defined period (e.g. the use of in hospital arthroplasty pathways, the use of transmural oncology pathways, etc.). Defining characteristics of care pathways include: (i) An explicit statement of the goals and key elements of care based on evidence, best practice, and patients' expectations and their characteristics; (ii) the facilitation of the communication among team members and with patients / families; (iii) the coordination of the care process by coordinating the roles and sequencing the activities of the multidisciplinary care team, patients and their relatives; (iv) the documentation, monitoring, and evaluation of variances and outcomes; and (v) the identification and use of the appropriate resources. The aim of a care pathway is to enhance the quality of care across the continuum by improving risk-adjusted patient outcomes, promoting patient safety, increasing patient satisfaction, and optimizing the use of resources. Currently about 1100 clinical pathways are (being) developed in 106 participating institutions.
Primary focus (structure, process outcome)	Process is the main focus by identifying and optimizing key interventions throughout the patient care flow. However, outcomes are also measured to guide and monitor the process.
Type of indicators used (if any)	Process & Outcome indicators are used within a before after comparison.
Purpose of indicators used	Disease specific outcome indicators and process indicators measuring the compliance to the key interventions.
Type of incentives	Only within organization incentives, no financial incentives from government except for indirect effects (e.g. financial yield of a lower length of stay).
Results (if available)	Improved coordination and follow up of the care process. Clinical results are disease and case study specific based on actual bottlenecks which were found in pre-test (as-is phase).
Sources (most relevant ones)	[243-245]
Total budget	Unknown

Table 9: Providers' accreditation

III. PROVIDERS' ACCREDITATION (Licensure and certification)	
Initiated by	Physician-Insurers National Commission of the National Institute for Health and Disability Insurance (NIHDI) (INAMI/RIZIV)
Overall objective	To support continuous learning with regard to up to date medical knowledge and competences
Date of implementation in Belgium	December 13th, 1993
Target audience	GPs and specialists, dentists

Content	Common conditions for GPs and specialists: A minimum amount of 60 continuing education units within the last 3 years, and among them 9 units in ethics/economy and 6 gained for participation in the peer-review group (GLEM/LOK) A minimal yearly activity rate (e.g. 1 250 contacts for a GP, and a minimum of 5/day) No repeated remarks from the medical profile commission (diagnosis and prescription processes) To participate in quality assessment organized by peers To take part minimum twice a year in a peer-review group (GLEM/LOK) Specific conditions for GPs: To hold a medical record for each patient and to communicate its relevant elements to other providers when necessary To be a "certified" GP To exercise as a GP as main activity and to ensure the continuity of care Specific condition for specialists: To transmit all relevant patient data to the patient's GP Conditions for dentists: A minimum amount of 500 continuing education units per 5 years, and among them 50 units in ethics/economy/organizational domain To take part minimum twice a year in a peer-review group To participate to sporadic data collection when requested A minimal yearly activity rate of 300 contacts
Primary focus (structure, process, outcome)	The focus is on structures and processes.
Type of indicators used (if any)	No direct indicators are measured.
Purpose of indicators used	The providers' accreditation was initiated to promote quality and cost-effectiveness of care. The effectiveness of inter provider linkage, through the communication of patients' data and the complementarities of their medical specificity, and the continuing education of physicians were among the initial goals (1993). Currently, the accreditation is a kind of quality label for those who perform continuous medical education and use the medical technologies in a wise and socially justified way.
Type of incentives	Doctors: Revaluation of the consultation fee (+3.27 € for a GP, from +1.07 € to +4.88 € for a specialist), fully paid by the Social Security. A flat-rate amount per year of 580 € (2009). Dentists: a flat-rate amount of 2 355.74 €/year (2009)
Results (if available)	In 2003, a survey (see reference below) assessed the satisfaction of participants. The doctors experienced an improvement of the quality of and the participation in continuous medical education. They quoted a better awareness of ethical, economical and qualitative aspects of their job. They experienced an improvement of their practice through a better knowledge, intercollegiate assessment and multidisciplinary working. The current conditions for obtaining the individual accreditation were considered sufficient, but the retribution should have been increased.
Sources (most relevant ones)	[246, 247]
Total budget	Flat-rate amounts: doctors €14 159 000, dentists €11 620 000 (budget 2009). Costs within nomenclature (doctors only): €189 700 000 (2008).

Table 10: Breast cancer screening prevention bonus

IV. BREAST CANCER SCREENING PREVENTION BONUS	
Initiated by	Regional Flemish government in cooperation with Logo ZuiderKempen (area) and the University of Antwerp
Overall objective	This geographically sampled demonstration project, with a concurrent control area, investigates whether the use of a prevention bonus leads to an increase in patient participation in the screening of breast cancer.
Date of implementation in Belgium	2009-2010
Target audience	General practitioners and gynaecologists.
Content	The positive financial incentive is as such the main focus of this project. The incentive should reinforce the degree of advice and motivation to participate in screening, as supported by providers, towards women between 50 and 69 years of age.
Primary focus (structure, process, outcome)	Process
Type of indicators used (if any)	Mammography use is expressed in terms of the patient participation rate.
Purpose of indicators used	To evaluate the effect of the use of a financial incentive
Type of incentives	A financial reward is awarded if the area participation rate increases with at least 2.5% (stimulation of cooperative team approach). The bonus size is determined by the participation effect size (2 to 8 euro for an effect of respectively 2.5 to 10% or more). The bonus is awarded to the practice of the provider for each patient that participated in screening. The time interval of incentive provision is two years and a one time event.
Results (if available)	Not available
Sources (most relevant ones)	Not available
Total budget	Depends on project results (level of pay-out), max. €212 000

Table 11: Capitation funding revaluation in the primary health care centres

V. CAPITATION FUNDING REVALUATION IN THE PRIMARY HEALTH CARE CENTRES	
Initiated by	French- and Dutch-speaking federations of "Maisons médicales" and "Wijkgezondheidscentra", who asked the National Institute for Health and Disability Insurance (NIHDI) (INAMI/RIZIV) for the revaluation.
Overall objective	To contain costs due to overuse
Date of implementation in Belgium	1992
Target audience	Care providers (GPs, nurses, physiotherapists) who concluded a capitation fee contract with the legal health insurers.
Content	Money saving in referral in 3 domains (hospitalization, medical imaging and clinical/ medical Biology), compared to the national level, is granted by an increased capitation fee.
Primary focus (structure, process, outcome)	The focus is on processes, and on financial outcomes.
Type of indicators used (if any)	The global saving for the insurance within the 3 domains cited above, considered as a whole, is the only economic indicator. A cost overrun within one domain can be balanced by money saving within another. The indicators were yearly measured during the first years, but no more currently. The latest published assessment was part of the KCE report 85, made on 2002, 2003 and 2004 data.[248]
Purpose of indicators used	The purpose is to support a rational use of medical imaging and clinical/medical biology, and second line care in general. The programme tries to support the role of GPs to reduce the duration of hospital stay of their inpatients.
Type of incentives	A 10% revaluation of the capitation fee.

	The revaluation is given globally to all providers in the capitation fee system, without any consideration for their individual results.
Results (if available)	The costs are lower in the capitation fee system for specialist care in general, institutional care, medical imaging, clinical/medical biology and in vitro nuclear medicine.
Sources (most relevant ones)	[248]
Total budget	10% of the global funding for the Belgian capitation system represents about €4 380 000 (2008).

Table 12: EPA (European Practice Assessment)

VI. EPA (European Practice Assessment) TOOL 2005	
Initiated by	WONCA Europe / EQUIP Quality Assurance board/ TOPAS-Europe
Overall objective	To support the continuous use of quality of care enhancement, to assess and improve structure and processes in general practice
Date of implementation in Belgium	2001: first testing, 2007: starting with KCE project
Target audience	General practices
Content	GPs that enter the EPA system appoint a quality manager (colleague or auxiliary personnel), and engage with all the members of the Team (physicians, administrative and supporting personnel) to start a quality evaluation and improvement process that takes 3 year. Three sources of information are used (199 indicators): Staff questionnaires with quality related items, grouped in 5 areas (infrastructure, people, information, financial accountancy, quality and safety). Patient Questionnaire, taken by 75 randomized patients Practice visits and coaching by a trained EPA coach. Benchmarking is used to compare results of a particular practice using a visual scale VISOTOOL to observe differences regionally or even internationally. A Plan Do Check Act (PDCA) cycle is started to discuss weak points, prepare actions and use opportunities, under guidance of the local quality manager and the EPA coach. At the end of the process, a new evaluation is done to see if targets are met.
Primary focus (structure, process outcome)	Measurable elements of quality issues on 5 areas or the patient questionnaire
Type of indicators used (if any)	199 indicators of quality related issues
Purpose of indicators used	Results can trigger change within a practice. Benchmarking allows comparison between practices.
Type of incentives	Until now, there are no consequences or rewards for the obtainment of the EPA-quality label. In Germany, EPA is one of the suppliers of mandatory practice accreditation, is probably of best quality and has a market share of approximately 20 percent.
Results (if available)	Preparative field tests have run in different countries, including 32 Flemish practices (Domus Medica) and some Walloon practices. Ten Flemish practices finished their 1st year of EPA-process (9 group-practices and 1 duo-practice, on the average with 4 GP's and 3 non medical collaborators), most of them as a teaching practice of the Leuven HASTA network. In a former KCE project the feasibility was studied in Flanders and the French speaking part of Belgium. This feasibility study shows that prior interest in EPA is low in the GP community. A number of logistic and organizational problems were encountered. It proved attractive to participants, but it can be augmented by coaching of participants in more than a one-off project to identify and achieve targets for quality improvement.[249] In the absence of commitment of the government, a network of universities and one scientific organization will offer EPA as a service to training practices in the near future.

Sources (most relevant ones)	See http://www.topaseurope.eu/; [249-253, 254 , 255]
Total budget	The estimated costs are +/- €1 000 per practice per year (so for Belgium €1 per patient per year)

Table 13: Global medical record

VII. GLOBAL MEDICAL RECORD	
Initiated by	NIHDI (INAMI/RIZIV)
Overall objective	To support continuity of care between patients and GPs.
Date of implementation in Belgium	May 1, 1999 for a predefined population (May 1, 2002: for any patient)
Target audience	General practitioner and patient
Content	At the patient's formal request the GP manages the GMR that must include a selection of medical data. The objectives are to centralize the patient's medical data, to devote this central role to the GP, to improve the communication of the patient's data between health providers and to reduce medical shopping by gaining the loyalty of the patient.
Primary focus (structure, process outcome)	Process
Type of indicators used (if any)	Existence of a GMR with at least one contact with the patient per year
Purpose of indicators used	Improve coordination of care.
Type of incentives	Financial. Payment of the GP for this new service: 27.50 €/year (2009) Full reimbursement of this annual fee to the patient and a decrease of the patient's co-payment for other services provided by the GP.
Results (if available)	Continuous increase of the number of GMR (globally in Belgium: 44% of the insured population in 2005)
Sources (most relevant ones)	NIHDI (INAMI/RIZIV), insurers
Total budget	€104 966 317 (€27.50/GMR, based on figures 2007)

Table 14: Prescription feedback

VIII. PRESCRIPTION FEEDBACK (national council for quality promotion)	
Initiated by	National Council for Quality Promotion
Overall objective	To reduce overuse of drug prescription in an evidence based way
Date of implementation in Belgium	2003
Target audience	GP and specialists
Content	The goal is to encourage an EBM-driven prescription attitude. The individual feedbacks include: Analysis of the doctor's prescription of target drugs (antibiotics, antihypertensive drugs) with reference to global Belgian prescription of the same molecules and reference to peer review group members of LOK. The median prescription level is used as reference point. Recommendations or guidelines concerning the use of the target drugs. Invitation to discuss the results in the local group of medical evaluation (GLEM/LOK)
Primary focus (structure, process outcome)	Process.
Type of indicators used (if any)	Volume and choice of prescription of target drugs
Purpose of indicators used	To develop doctor's awareness to be driven by EBM prescription To decrease the volume and cost of prescribed drugs To optimize adequate choice of antibiotic

Type of incentives	For antibiotics: increase of GP's payment in 2004 (condition: change of global trend in prescription). This was a onetime general measure; support to the local groups
Results (if available)	Global tendency to decrease prescription of antibiotics (consequence of multiple approaches)
Sources (most relevant ones)	[256, 257] , NIHDI website
Total budget	€10 million in 2004 to increase the GP's payment + Logistic support

Table 15: Preventive module in global medical record

IX. PREVENTIVE MODULE IN GLOBAL MEDICAL RECORD	
Initiated by	NIHDI
Overall objective	To support evidence based preventive care
Date of implementation in Belgium	2009
Target audience	General practitioner and patients with an age between 45 and 74 years.
Content	Primary and secondary prevention. Examples of actions are influenza vaccination for elderly patients, cancer screening, cardiovascular risk screening, and smoking cessation advice. Implementation feasibility in a non electronic medical record is also secured, while electronic application formats are under preparation.
Primary focus (structure, process outcome)	Process targets, which are patient group specific (preliminary selection): counselling and advice (nutrition, physical exercises, smoking, alcohol), anamnesis and clinical investigation, aspirin use for cardio-vascular indications, screening (colon, uterus and breast cancer), vaccination (diphtheria, tetanus, influenza, pneumococcal), biological testing (glycaemia, creatinine, proteinuria, cholesterol)
Type of indicators used (if any)	Criteria for target selection are (1) the evidence base, (2) the feasibility of performance, based on organisational conditions at the system's level, (3) the clinical impact (burden of disease), (4) the availability of target specific supportive tools and materials
Purpose of indicators used	To support and evaluate preventive action on different levels (GPs, local councils)
Type of incentives	A supplementary fixed fee of 10 euro per GMR, based on the module availability during two years and on reporting of module data during the third year.
Results (if available)	Not available. Evaluation of the initiative based on the volume of GMR coverage, sampling of target performance, and patient experience survey
Sources (most relevant ones)	Not available
Total budget	€38 169 957 (€10/GMR, based on figures of 2007)

Table 16: Centres of reference

X. CENTRES OF REFERENCE	
Initiated by	DKV-Belgium as a private insurer (demand from foreign patients).
Overall objective	To define selection criteria for the selection of centres of reference. To create a network of centres of reference for the treatment of life-threatening or complex diseases. To inform foreign patients about these centres of reference.
Date of implementation in Belgium	In preparation
Target audience	Belgian hospitals.
Content	Selection criteria: objective and measurable.
Primary focus (structure, process,	A combined quality index.

outcome)	
Type of indicators used (if any)	Structure indicators (the following list is not complete): staff size, scientific activity (total citation score of the team, clinical research) quality of room, of food, additional services to patients (psychological and social support, pain clinic, esthetical care, revalidation). Process indicators (the following list is not complete): attractiveness index, safety index, check in- diagnosis delay, diagnosis-treatment delay. Outcome indicators (the following list is not complete): disease-free survival in cancer, quality of life.
Purpose of indicators used	To publicly rank hospital performance
Type of incentives	Selection as a centre of reference if selection criteria demonstrate that a certain threshold is reached.
Results (if available)	Not available
Sources (most relevant ones)	Not available
Total budget	Not available

Table 17: Hospital accreditation

XI. HOSPITAL ACCREDITATION	
Initiated by	Hospitals in cooperation with an accreditation body (Joint Commission International, Nederlands Instituut voor Accreditatie van Ziekenhuizen (NIAZ), International Office for Standardization (ISO))
Overall objective	To publicly recognize the level of quality of care offered by a hospital
Date of implementation in Belgium	Hospital specific (2000, 2006…)
Target audience	Hospitals, including both its managers as professional care providers
Content	Initiatives to externally assess hospital performance against pre-defined explicit published standards in order to encourage continuous improvement of the health care quality. Target standards are applied as intermediate endpoints to achieve. These may cover the entire process model of a hospital.
Primary focus (structure, process, outcome)	Main focus on structure and process. There is a current lack of using (intermediate) outcome measures to evaluate programme results.
Type of indicators used (if any)	Limited clinical indicator use due to a focus on organisational and transversal aspects which are not addressed in current available databases. The choice of indicators is mainly accreditation body specific. Examples are the availability of guideline support (structure) and the actual use of guideline support (process) in medical decision making.
Purpose of indicators used	To improve quality and to strengthen public accountability. Accreditation results in a periodic 'award' of recognition.
Type of incentives	An accreditation initiative often includes charging external services to a hospital by a fee or subscription, next to a cost increase working towards standards. There is no direct positive financial incentive present. Indirectly, a positive public recognition may improve the financial status through higher patient volumes.
Results (if available)	Evidence is lacking.
Sources (most relevant ones)	[258]
Total budget	Undisclosed and programme specific

Table 18: Hospital benchmarking

XII. HOSPITAL BENCHMARKING	
Initiated by	Federal government (PATH project, supported by WHO, multidimensional 360° feedback project, patient safety indicators project), Hospital networks (e.g. Vlaams Ziekenhuisnetwerk), Center for Health Services and Nursing Research-KULeuven + UCL (Navigator, Delta), Institute for Public Health (IQED), sickness funds
Overall objective	To benchmark quality performance between participants, to analyze trends and identify room for improvement
Date of implementation in Belgium	Most projects started during the late nineties/ beginning of this century. The 360° feedback project and the patient safety indicators project started respectively in 2006 and 2008. The latter includes a retrospective analysis of the period 1999-2004.
Target audience	Most projects focus on general hospitals. However, Navigator and Delta also serve nursing homes and psychiatric hospitals as separate target groups. Results from the 360° and the patient safety indicator projects are fed back to all Belgian general hospitals. In 2007 eleven Belgian hospitals participated in PATH. The Initiative for Quality Promotion and Epidemiology in Diabetes Care (IQED) focuses on care provided in all Belgian diabetes convention centres. Eighty three institutions participate in Navigator in 2009, and 37 in Delta. The Vlaams Ziekenhuisnetwerk consists of 18 general hospitals. Data of the sickness funds has the advantage to enable both a primary care and hospital care focus, linked on the individual patient level.
Content	Iterative cycles of indicator development, application in benchmarking and revision. Benchmark results are used to develop target specific quality improvement strategies as a part of hospital policy, moving from information towards actions. The PATH project includes 7 clinical effectiveness and safety core indicators and 7 tailored indicators on both a national and international level. The 360° feedback project includes 7 clinical performance indicators, 10 economical performance indicators, 4 patient centeredness indicators and 12 capacity, staffing and innovation indicators. The patient safety indicator project includes 20 targets. IQED makes use of a broad diabetes care indicator set. Navigator includes 454 general hospital indicators, 357 psychiatric hospital indicators and 123 nursing home indicators. Most projects also include technical support, educational support, etc.
Primary focus (structure, process outcome)	PATH, 360°, IQED, Navigator and Vlaams Ziekenhuisnetwerk include process and outcome measures. The patient safety indicator project is mainly outcome driven. Delta targets the measurement of the patient experience of the clinical structures, processes and outcomes.
Type of indicators used (if any)	Caesarean section rate, mortality rate, readmission rate, prophylactic antibiotic use, day surgery rate, etc. Some are hospital generic, others are medical condition specific. The development and validation process of these indicators varies between initiatives.
Purpose of indicators used	To benchmark through time and in comparison with peer hospitals with a similar case mix.
Type of incentives	There are no financial incentives provided. In most projects

	hospitals pay a periodic fee to support benchmarking operations. An initiative like the hospital benchmarking in the Vlaams Ziekenhuisnetwerk provides indirect positive financial incentives through other shared services like material purchasing.
Results (if available)	No results evaluation available, with the exception of periodic hospital feedback reports.
Sources (most relevant ones)	See KCE report vol. 41 [259] PATH: http://www.euro.who.int/document/e89742.pdf IQED: http://www.iph.fgov.be/EPIDEMIO/epien/PROG33.HTM Navigator: http://www.navigator.czv.be Delta: http://www.czv.kuleuven.be Vlaams Ziekenhuisnetwerk: http://www.vznkul.be/root/index.html 360° and patient safety indicator project: internal documents provided by FOD-VVVL
Total budget	Unknown (not publicly available)

Table 19: Quality and patient safety in hospitals

XIII. QUALITY AND PATIENT SAFETY IN HOSPITALS	
Initiated by	Federal government.
Overall objective	To support the systematic use of quality improvement strategies, with attention for multiple quality domains
Date of implementation in Belgium	Since July 2007, additional financing has been approved within the remuneration of hospitals and this for the co-ordination of the quality and patient safety in Belgian hospitals.
Target audience	Hospitals (general, psychiatric, rehabilitation)
Content	**Contract year 2007-2008: Six parts** The mission, vision, goals and strategy with regard to quality and patient safety The first part of the contract aimed at further stimulating the hospitals into formulating a mission, a vision, goals and strategic objectives that also integrate patient safety. Structures and functions with regard to quality and patient safety Within the framework of the second part of the contract, the hospitals were asked to give an overview of the existing structures with regard to quality and patient safety by means of an organisation chart. These are the committees or functions that are financed by the federal government and/or legally imposed. Hospital survey on patient safety culture The third part of the contract aimed at assessing the patient safety culture. The federal authorities have used the Hospital Survey on Patient Safety Culture of the AHRQ (Agency for Healthcare Research and Quality) to do so. Reporting and analysing (near) incidents The fourth part of the contract related to the reporting and analysing of incidents and near incidents. The description of three quality projects The fifth part of the contract related to the description of three new quality projects: one for the field of 'economic

	performance' / 'capacity and innovation', one for clinical performance and one for patient safety. Multidimensional feedback including further details on a selection of 12 indicators The sixth part of the contract only applied to acute hospitals. The hospitals were asked to select 12 indicators and to develop actions for improvement. They were asked to select three indicators per field and to develop them in detail: 3 indicators for the field of economic performance, 3 for capacity and innovation, 3 for clinical performance and 3 for patient safety. **Contract 2008-2009: three parts** As a first part to the contract hospitals were asked to develop a multi-year patient safety plan, and based on the results of the evaluation of the patient safety culture (as performed in the previous year) two actions needed to be developed for quality improvement in two separate domains. As a second part to the contract the hospitals were asked to analyse a process where they could freely choose between a process as suggested by the government, or a process related to intramural transfers or a process based on results from the multidimensional feedback. As a third part to the contract hospitals were asked to list all of their indicators that were used for the management of quality and safety within the hospital. **Contract 2009-2010: three parts** As a first part to the contract the hospitals will be asked to develop a reporting and learning system for incidents and nearly incidents. In addition the hospitals will be asked to describe five incidents including the method used for the analysis as well as the actions that were taken to achieve the necessary improvements. As a second part to the contract the hospitals will be asked to develop or to (re)analyse an existing or new process within a multidisciplinary context. They could freely choose between a process as suggested by the government, or a process related to intramural transfers or a process based on results from the multidimensional feedback. 3. As a third part to the contract the hospitals will be asked to develop a multidimensional and integrated set of indicators related to quality and patient safety.
Primary focus (structure, process outcome)	The yearly contracts will always be based on the Donabedian's triad and consist of three pillars: the development of a safety management system (structure), the analysis of processes (process) and the development of a multidimensional set of indicators (result). By 2012 all hospitals should have an integrated safety management system, assess both intramural and transmural care processes and use an integrated and multidimensional set of indicators.
Type of indicators used (if any)	Structure, process and outcome indicators
Purpose of indicators used	The purpose is hospital specific. They are mostly used as an input, based on benchmarking, to prioritize quality targets and related quality improvement strategies in a cyclical manner.
Type of incentives	Yearly budget
Results (if available)	The mission, vision, goals and strategy with regard to quality and patient safety

The results from the first contract year show that most of the hospitals have a mission, a vision and strategic/operational goals. However, only half of the hospitals have a "real mission" that meets sound criteria. In addition, the hospitals do not explicitly link their mission, vision and strategic goals to each other. Another important finding is that most hospitals mix the terms "mission", "vision" and "strategy". It is also striking that the number and types of operational goals considerably vary between the different hospitals.
As to the communication of their mission and vision, the hospitals use many channels and they are very creative in spreading their mission and vision among patients and personnel. Other important conclusions to be made for this chapter are that only a third of the hospitals apply general quality frameworks (like the EFQM model). In general, we can conclude that it is really necessary to standardise and harmonise the meaning of "quality and patient safety in hospitals" because those concepts are interpreted in a very different way today. Such a standardisation, however, should take into account the specific features of each hospital as well as its patient population.

Structures and functions with regard to quality and patient safety

The committees that the hospitals have mentioned the most are the quality committee, the executive committee and the patient safety committee. More than half of the hospitals, both in Flanders and in Wallonia and Brussels established this last committee in 2008. The patient safety committee is characterised by a large participation of board members and a multidisciplinary composition. Most of the hospitals also have at least one full-time equivalent quality co-ordinator. Finally, various channels are found to be used in order to spread the organisation chart within the institution.

Hospital survey on patient safety culture

The results of this culture assessment have also been the subject of a benchmark that has been conducted for 132 Belgian hospitals. These results are presented in a separate report. Most of the participating hospitals (96%) have conducted a hospital survey on patient safety culture according to the methodological rules. The level of participation was lower for physicians than for other hospital staff. About half of the hospitals that have conducted the culture assessment have already formulated actions for improvement that are linked to the results of the culture assessment.

Reporting and analysing (near) incidents

The vast majority of the participating hospitals make use of a system to report incidents and near incidents. There is, however, a whole range of issues that can be reported. The anonymity of the reporter, the unit and the patient is mostly respected in rehabilitation centres. In the acute and psychiatric hospitals, reporters can choose between anonymous and non-anonymous reports in more than half of the cases. The reporting system is usually confidential, which means that the patient's identifying data, the reporter and the unit are not mentioned to third parties. A large majority of the hospitals also have a hospital-wide reporting system. Hospital-wide reporting systems are mostly used in rehabilitation centres,

	followed by psychiatric hospitals and acute hospitals. In slightly more than half of the hospitals, the reporting is exclusively done in writing. A minority of the hospitals already has an electronic reporting system and in one third of the hospitals the system allows for both written and electronic reporting. A positive point is that the hospitals really take a lot of initiatives to stimulate reporting, which include both written and oral (formal and informal) forms of communication, campaigns and training. One third of the hospitals, and particularly acute hospitals, use specific methods to analyse incidents and near incidents. This shows that there is still a lot of work to be done in this field. In about half of the hospitals the incidents and near incidents are analysed by the quality co-ordinator. The used analysis methods vary a lot. The actions for improvement, which are based on the analyses, that are mentioned the most relate to medication and falling prevention. The description of three quality projects In all, more than 500 projects have been submitted, mostly by acute hospitals, followed by psychiatric hospitals and rehabilitation centres. Many projects submitted by acute hospitals relate to supporting activities, medication, clinical paths and hospital hygiene. Psychiatric hospitals have mainly developed projects on medication, staff policy, patient flows and aggression. Rehabilitation centres have mainly submitted projects with regard to medication, hospital hygiene, falling prevention and the registration and analysis of (near) incidents. Medication safety is the subject on which the three types of hospitals have submitted the most projects. Multidimensional feedback including further details on a selection of 12 indicators The following indicators have been cited the most for the respective fields: degree of financial independence, clinical paths, amount of caesareans and decubitus. The actions for improvement have been subdivided into the following categories: financial management, human resource management, clinical aspects, informatics and patient safety.
Sources (most relevant ones)	See www.patient-safety.be for the FOD patient safety website and the Hospital survey on Patient Safety Culture [260]
Total budget	For the contract year 2007-2008 the budget amounted to € 6.8 million. The related contract on the co-ordination of quality and patient safety was signed by 80 % (n=164) of the acute, psychiatric and rehabilitation centres in 2007 and 90% of the hospitals in 2008.

Table 20: Reference payment hospitals

XIV. REFERENCE PAYMENT –HOSPITALS-	
Initiated by	RIZIV/INAMI
Overall objective	To reduce the level of overuse of specific care interventions
Date of implementation in Belgium	Start of development in September 2002, revised during 2002-2008. Latest version implemented in December 2008 (applicable for hospital admissions since 2006).
Target audience	Medical care in acute and chronic general hospitals, with an exclusion of one day stays. Twenty surgical and 12 medical commonly present DRGs are included for SOI 1 and 2 (low severity of illness patient admissions). This equals 18% of all hospital admissions.
Content	Comparison of hospital expenses for medical imaging, clinical biology and technical services with the national median per diagnostic group (APR-DRG) and SOI level + 10%[k]. Since 2009 the time frame of included services can be broadened to include the identified services during 30 days before admission to prevent cost shifting. High LOS outliers are excluded from the national median calculation. Services with existing standard expense payments are also excluded. Since 2009 hospitals are prospectively informed about the reference norms as to enable the prevention of excesses.
Primary focus (structure, process, outcome)	The targeted expenses are related to the processes of medical imaging, clinical biology and technical services, although not specifically measured in non financial terms.
Type of indicators used (if any)	Financial data/outcome. No structure, process or clinical outcome measures.
Purpose of indicators used	To standardize expenses and reduce variability as justified by the national average comparison.
Type of incentives	When the total selected hospital expenses exceed the reference norm, a hospital is obliged to refund the total sum of exceeding (higher) differences above 1000 euro. There are no incentives related to an expense position below the reference norm. Since 2009 minimal thresholds can be established to prevent overly pressure on the national average.
Results (if available)	No results evaluation available, with the exception of periodic hospital feedback reports.
Sources (most relevant ones)	See www.riziv.be and KCE report vol. 17 [261]
Total budget	About 16.7% of total national expenses for the three types of included services.

7.1.4 Discussion

Although there is a lack of peer reviewed publications on Belgian quality initiatives, a lot of programmes support quality. Here we were only able to present fourteen of them, but even these fourteen illustrate the wealth of initiative and effort to address quality, and the diversity of approaches followed.

Evidence on programme effects is however often lacking. The different approaches are very programme specific. Currently it is unclear how these initiatives all fit together in their mutually shared mission of improving quality of care. There is no encompassing national, regional or local quality improvement strategy directed at healthcare. Although diversity leads to creativeness, experimentation and innovation, the lack of a common platform (harmonization of priority setting, target selection, supporting tools, etc.) induces a risk of creating quality gaps between initiatives and between decision making levels.

[k] Each diagnostic group (APR-DRG) is subdivided into four categories of severity of illness (SOI), ranging from 1= 'low severity' up to 4= 'high severity'. The median expense at this sublevel is increased with ten percent to calculate the threshold of reference payment.

It's like different pieces that are not brought together in a full quality improving puzzle. This lack of quality improvement tuning also leads to cost effectiveness questions concerning budget allocation. Why is and remains a large proportion of quality supporting resources reserved for an incentivized provider accreditation system, while other programmes, often with a higher EB supported effect size, are supported with lower resources? A general P4Q philosophy implies that payment becomes a tool directed at those priority areas and initiatives yielding the highest results in health gain. This requires however that the resources consumed through historical merits (such as provider accreditation) may be reallocated based on evidence and stakeholder consensus. This implies not a reallocation in terms of who receives financial resources and their amount, but a reallocation in terms of the criterion used, the basis of distribution.

It can be confirmed that none of the described programmes currently fit the definition of P4Q, which is based on the actual measurement of quality and the linkage of those measurement results to the allocation of a financial incentive.

This section has provided a general overview of what is available in Belgium to support components of a quality circle. However, these components need to be specifically related to the revised P4Q framework to thoroughly analyze programme designs as compared to P4Q recommendations, and the feasibility of modification. This subsequent step is described in the following section.

7.2 FEASIBILITY STUDY OF IMPLEMENTATION OF P4Q IN BELGIUM

7.2.1 Introduction

When addressing the feasibility of P4Q within a Belgian context there are two main questions to be answered, building further on the theoretical and empirical guidance presented before:

Can P4Q be implemented in Belgium building further on existing quality improvement initiatives to design, implement and evaluate P4Q framework components?

Can P4Q be implemented in Belgium starting from scratch, i.e. designing, implementing and evaluating all P4Q framework components independently of existing Belgian quality improvement initiatives?

Next to feasibility, this chapter will explore the advantages and disadvantages of both options.

7.2.2 Methods

This chapter makes use of the empirically revised conceptual framework (Chapter 5) to analyze strengths and weaknesses in current quality circle components to define P4Q threats and opportunities. The level of correspondence for both the quality initiative independent approach and the quality initiative dependent approach with the set of 'to do's' is used as the central parameter. The feasibility of modifications is assessed.

7.2.3 Results

Appendix 12 provides an overview table of both options as placed within the P4Q framework.

7.2.3.1 Starting from existing quality improvement initiatives

Quality dimensions

A strength and P4Q opportunity is the fact that twelve of the fourteen initiatives do not address efficiency exclusively. Only two initiatives (capitation revaluation of primary health care centres and reference payments of hospitals) are only focused on cost containment within the efficiency domain of quality. Since P4Q looks beyond efficiency only, these two initiatives are difficult to relate. It is remarkable that such efficiency focused initiatives show some distinct characteristics as compared to the others: They rely on mandatory participation.

Both use a relative incentive in which allocation is dependent on the performance of others. None of the target selection criteria are used in these initiatives (e.g. level of evidence). Both focus exclusively on process measures (medical imaging, clinical biology, etc.) and financial outcomes. Both are static in terms of goal specification and make use of secondary data analysis based on administrative data. Whereas the capitation initiative includes a financial reward, the reference payment initiative is based on a negative financial incentive. This drives it further from P4Q recommendations. Characteristics like the mandatory nature, and the relative incentive allocation are mostly absent in the other twelve initiatives more focused on other quality domains. The rationale for this difference is at present unknown.

A P4Q programme should, as a first priority, be focused on clinical effectiveness. Furthermore, if other domains are also addressed, the P4Q opportunity is reinforced. One initiative, providers' accreditation, has no very clear quality domain focus. However, one can assume that continuing education as one of the intervention items supports effectiveness, cost-effectiveness and coordination of care.

The global medical record initiative in primary care is primarily focused on continuity and coordination of care. This makes it a less good candidate to fit with P4Q in its initial stage, unless additional payments are introduced as an incentive to reach specific targets, as is the case with the preventive module (initiative IX).

The other ten initiatives all focus upon clinical effectiveness in combination with one or more other domains. This supports the P4Q opportunity. However, some of these initiatives give participants more leverage to prioritize quality domains locally. Examples are clinical pathways and the quality and patient safety in hospitals initiative. When such initiatives are incentivized financially, the balance between national, regional and local quality domain priorities deserves specific attention. Based on different points of view (standardized minimal EB approach vs. local adaptable approach) this can be seen as an additional opportunity or threat to P4Q implementation.

Target population

Both in primary care as in hospital care there are sufficient quality improvement initiatives as a P4Q opportunity. Some (provider accreditation, care itineraries, clinical pathways) focus on the two settings. This might reinforce patient centeredness throughout a patient's trajectory, and support coordination, continuity and communication to minimize gaps of care. However, implementing P4Q in these settings simultaneously will likely also be a greater challenge. Therefore both a cautious as a more ambitious approach can be followed. Stakeholder consensus can assist in making the appropriate (stepwise) selection.

The different types of care (preventive, acute, chronic) are all addressed in initiatives to support P4Q implementation. Until recently preventive care received lesser attention in initiatives. However, the currently being developed preventive module in the global medical record for primary care will fill this gap, next to the breast cancer screening initiative.

As described in previous chapters, P4Q can support both medical condition specific as generic quality improvement initiatives. Initiatives like care itineraries, clinical pathways, reference payments for hospitals, breast cancer screening and the preventive module in the global medical record are mostly medical condition specific. Capitation funding revaluation in primary health care centres, providers' accreditation, use of the EPA tool and the global medical record are generic in terms of target population. Finally, the quality and patient safety in hospitals initiative, hospital accreditation, hospital benchmarking, centres of reference, and prescription feedback show medical condition specific and generic components.

Quality targets

One initiative, providers' accreditation, does not define specific quality targets measured. It therefore cannot be used together with P4Q in its current form.

Many initiatives (breast cancer screening, capitation funding revaluation in primary health care centres, reference payments for hospitals, prescription feedback, the global medical record and its future preventive module) only make use of process targets. Two initiatives combine structural and process targets (hospital accreditation, use of the EPA tool).

Although useful for P4Q purposes, the current state of the art includes at least (also) intermediary outcome measures. Care itineraries (process + intermediary outcome), clinical pathways (process + outcome), hospital benchmarking (process + outcome), centres of reference (structure + process + outcome) and the quality and patient safety in hospitals initiative (all target types possible) fit best as P4Q opportunity. As already noted above, the current level of local adaptability should be further considered. When a local organization implements P4Q independently, it is free to design the quality target and measurement system according to internal needs.

However, when P4Q is implemented nationally or regionally, the introduction of a financial incentive should be linked to a set of minimal requirements to ensure that the P4Q system conforms to best practice standards. These standards provide the highest probability of reaching the national/regional priority goals and targets. Too much freedom in target selection induces a risk of free rider behaviour. This means that participants might select targets based on the level of ease to reach them. Thus they focus on receiving financial resources without true quality improvement or maintenance. This would reduce P4Q's health gain potential and be unfair in terms of incentive allocation. The design of standards to reduce this risk can be the responsibility of professional scientific organizations, which represent different medical disciplines.

The number of target indicators available and/or used in the different quality improvement initiatives gives enough input to design a P4Q target set which is not too limited, nor too complex (see for example indicators within the EPA tool and within hospital benchmarking). The exact number should be specified based on all stakeholders' consensus. Breast cancer screening, being a true P4Q initiative, is based on only one target measure, which is quite narrow in approach.

Next to the two efficiency focused initiatives (reference payment for hospitals, capitation revaluation in primary health care centres), and prescription feedback which all address inappropriate care (overuse) exclusively, there are five initiatives with an exclusive appropriate care focus (underuse): care itineraries, breast cancer screening, centres of reference, the global medical record and its planned preventive module in primary care. P4Q requires at least the appropriateness approach, preferably in combination with attention for inappropriate care. This combination can be found in the other six initiatives, although again sometimes dependent on local decisions.

Concerning target selection and definition criteria a SMART configuration is present in most initiatives, except for the reference payment of hospitals, the capitation revaluation, and providers' accreditation.

One of the main weaknesses in most initiatives is the lack of evidence base requirements. Exceptions are the care itineraries, breast cancer screening and the preventive module in the global medical record. These more recent initiatives put the evidence base central to target selection. Most other initiatives (clinical pathways, quality and patient safety in hospitals, EPA tool, hospital benchmarking, etc.) provide an opportunity to participants to make use of EB standards, but participants remain completely free to make use of EB or not. An EB precondition should be reinforced for a P4Q opportunity to succeed.

The same kind of noncommittal approach can be found in terms of room for improvement as a target selection criterion. Many initiatives advise and teach participants to base target selection on existing quality gaps, but in the end the selection is often locally decided upon without any level of supervisory monitoring of compliance with baseline needs.

In addition, initiatives on a national or regional level (care itineraries, preventive module, breast cancer screening) predefine the targets with no attention for local room for improvement.

The level of cost effectiveness as a target selection criterion in existing initiatives lacks evidence and is rarely explicitly focused upon. A P4Q initiative will therefore require additional attention for the degree of health gain per unit of expense as part of the target selection process.

A strength of many initiatives is their dynamical approach. Care itineraries, clinical pathways, the quality and patient safety in hospitals initiative, hospital accreditation, use of the EPA tool, hospital benchmarking, etc. are all meant to evolve and change over time, both regarding targets as regarding the quality improvement process itself.

Quality measurement

Current quality improvement initiatives in Belgian primary and hospital care show that quality is measured based on different methods of data collection. Some initiatives make use of secondary data analysis based on existing administrative data (capitation revaluation, reference payments for hospitals, prescription feedback) or clinical data (breast cancer screening). Others use a sampled approach based on health care records (clinical pathways, EPA tool). Automatic data extraction is planned for more recent initiatives like care itineraries. Finally, a number of initiatives are based on a combined approach of data collection methods (quality and patient safety in hospitals, hospital accreditation, hospital benchmarking). The validity of data is subject to the limitations of all methods used. Probably a combined approach offers most security. A strength and P4Q opportunity is that quality data are being collected broadly throughout initiatives. Data validity deserves specific attention, but both international and national examples show that a valid data foundation is possible when addressed profoundly.

A strength of current initiatives is the attention for case mix adjustment for outcome measures (care itineraries, clinical pathways, hospital benchmarking). This can be based on separate subgroup analyses and comparisons or on a more integrated risk adjustment modelling approach. This experience can be considered a P4Q opportunity.

Currently quality improvement initiatives do not use exception reporting as a way to preserve intended or uncontrollable variability of care. However, the clinical pathway initiative makes use of a similar approach by using analysis of variation. This means that deviations of expected care are recorded continuously as part of daily practice (integrated in the health record system) and are used as a source of information to improve the quality of care systematically. This method does not predefine exception reporting criteria as in the UK QOF framework, but can be used as a starting point to explore exception reporting possibilities.

A weakness of existing quality initiatives in Belgium is the general lack of monitoring for unintended consequences. Differences in equity of care (based on gender, age, ethnicity, socio economical status, presence of co morbidities, etc.) and potential neglecting effects toward other quality priority areas are not exclusively related to the 'pay' component in P4Q. Unintended consequences may be present or arise in quality improvement initiatives in general. Irrespective of P4Q implementation, this issue deserves specific attention in Belgian healthcare.

Figure 11 : P4Q concepts: Quality

Quality	
Strength	*Weakness*
Effectiveness focus, often combined with other domains *Growing attention for integration and coordination*	*No clear focus in all initiatives* *Some initiatives with an exclusive efficiency focus* *Limited attention for Equity*
High level of local adaptation	
Quality targets	
Strength	*Weakness*
Primary care + hospital care initiatives *Preventive + Acute + Chronic care initiatives* *Medical condition specific + generic initiatives* *Availability of different target types (structure, process, intermediate outcome, long term outcome* *High number of target definitions available* *Attention for both appropriate and inappropriate care* *Attention for SMART target definition* *Use of a dynamical approach in most initiatives*	*No measured quality targets in all initiatives* *Lack of evidence base selection requirements* *Lack of room for improvement selection requirements* *Lack of cost effectiveness selection requirements*
High level of local adaptation	
Quality measurement	
Strength	*Weakness*
Different types of data collection available *A combined approach is used in some initiatives* *Attention for case mix adjustment in various initiatives*	*Limited experience with exception reporting* *Lack of monitoring of unintended consequences*
Existing databases can be leveraged for hospital care, and more limited for primary care	

P4Q incentive

A number of quality initiatives are not directly related to financing (clinical pathways, hospital accreditation, EPA tool, centres of reference, hospital benchmarking). Financial resources are not specifically related, except in terms of implicit internal costs (e.g. number of FTE's involved, payment of participation fee) and implicit revenue effects (e.g. patient volume effects due to public recognition or implicit rewarding of a shorter hospital LOS as part of the current financing system). These initiatives are not explicitly incentivized financially. An incentive and its structure would have to be implemented starting from scratch.

Other initiatives are explicitly incentivized, but not directly related to quality performance (care itineraries, quality and patient safety in hospitals). We also include capitation revaluation, providers' accreditation, reference payments of hospitals, prescription feedback, the global medical record and its preventive module in this category. Most of these initiatives are financed based on a fixed fee independent of cost or quality (surpassing efficiency and/or volume performance). Some are however measuring and/or reporting related as a performance criterion (care itineraries, quality and patient safety in hospitals). This can be considered a first stage in a phased approach of incentivizing quality of care. One initiative, breast cancer screening, includes an explicit financial incentive directly related to quality performance.

Incentive structure

The not per se strict P4Q incentives that are used in current initiatives show a number of strengths in terms of incentive structure best practice. The first is the predominantly rewarding nature. The second is the absolute methods use (participants are rewarded independently from each other based on an objective, known and fixed criterion to be pursued and based on each participant's own merits).

Also in terms of stability and simplicity vs. complexity there are promising signs based on existing initiative characteristics. The frequency, which is often yearly, provides sufficient time for practice change and for practical data collection modalities to be feasible. However, in long term a focus on shorter time intervals, with assistance of automatic data collection methods, may prove worthwhile to ensure a sense of urgency and to heighten awareness.

What is often lacking in current quality initiatives is the size of the incentive (well below 10% of total revenues). In addition, the incentivized level is often, as recommended, the individual provider in primary care (e.g. prescription feedback, preventive module). There is no incentive focus on the team level in primary care, except through the complementary goals within the care itineraries initiative, and through the team allocation level within the capitation revaluation initiative.

In hospital care the incentive is in existing initiatives awarded to a hospital as a whole (quality and patient safety in hospitals, reference payments of hospitals). There is a lack of incentive focus towards individual providers and teams of providers improving quality within hospitals. This serves as a further point of attention with regard to P4Q.

Figure 12 : P4Q concepts: Incentives

Incentives	
Strength	*Weakness*
Some pay for reporting initiatives *Incentive structure:* *Predominant use of rewards* *Predominant use of absolute incentive methods* *Attention for incentive stability*	*Lack of explicit financial incentives in some initiatives* *Lack of a relationship with quality performance in some initiatives* *Small incentive size (<10% of revenues)* *In some initiatives focus on individual and/or team level (Strength), but especially in hospital care often focus on organization level*
Focus on simplicity of the incentive structure	
Use of long duration time intervals between incentive allocation (e.g. yearly)	

Implementing and communicating the programme

In general it seems that there is sufficient attention for provider involvement in setting goals in most current initiatives. There is also a high level of communication to raise awareness. These conditions are only lacking in efficiency only focused initiatives. As recommended, most initiatives make use of voluntary participation. In addition, some initiatives are good examples of using a staged approach (care itineraries, clinical pathways, quality and patient safety in hospitals). It is also reassuring that most initiatives make use of an embedded quality supportive approach instead of a stand-alone design. This global experience with implementing and communicating quality improvement initiatives can be leveraged to support P4Q implementation.

Evaluation of the programme

Questions regarding sustainability of change, validation of the programme and reviewing and revising the process are often at present premature to answer, because a number of the studied initiatives haven't reached the evaluation stage yet or evaluation is currently ongoing. However, this doesn't prevent a general assessment of the quality of evaluation.

Some initiatives only make sporadic or exceptional use of evaluation (capitation revaluation, providers' accreditation). Others make regular use of evaluation (clinical pathways, quality and patient safety in hospitals, hospital accreditation, hospital benchmarking), but it is not clear to which extent this evaluation then can lead to possible changes in the programme.

Figure 13: P4Q concepts: Implementation, communication and evaluation of the programme

Implementing and communicating the programme	
Strength	*Weakness*
Attention for involvement of providers in setting goals in most initiatives *High level of communication in most initiatives* *Use of voluntary participation in most initiatives* *Some good examples of using a staged approach* *Use of an embedded quality support approach in most initiatives*	
Evaluation of the programme	
Strength	*Weakness*
	Often premature to assess due to pre evaluation stage or ongoing first time evaluation *Some initiatives make regular use of evaluation (Strength), others sporadic or exceptional*

7.2.3.2 Starting from scratch

The 'starting from scratch' option gives complete freedom to focus on quality dimensions and quality targets which fit best. During a first phase 'effectiveness' can be the main target dimension, with monitoring of other dimensions to prevent unintended consequences (e.g. level of equity, provider experience and cost effectiveness). Later on other quality dimensions can be included (e.g. care continuity and coordination).

P4Q can be of similar value, both for primary care as for hospital care. Based on evidence there is no argument to prefer one of both. P4Q can support quality throughout different settings, with operational modifications wherever necessary.

However, according to the WHO primary care should receive a particular focus since it can have a stronger impact on health if guidelines are more widely spread and applied, and prevention and health promotion are correctly managed.[262]

Rewarding quality goals in general practice can contribute to strengthen primary care, leading to a more balanced health care system.

There is also no reason to focus on preventive, acute or chronic care separately. A mix of targets can be included to emphasize and incentivize the complementary value of an integrated approach.

The quality targets within a Belgian context can consist of a combination of structural, process and intermediate outcome measures. Long term outcome effects are practically assessed by relating them to short term intermediate outcome effects. The different types of indicators are already widely available within Belgian healthcare. Existing knowledge, experience and evidence can therefore guide the target selection and definition process.

The number of targets can be phased in, but to gain sufficient awareness and behavioural effect from the start a too low number (e.g. below five) should be avoided.

One can choose between an immediate appropriateness (tackling current underuse) and inappropriateness (tackling current overuse) approach or also using a phased approach in this respect. In the end both underuse and overuse targets should complement each other in a logical manner. Since Belgium has experience in both, this poses no specific problem. However, it might require that the separate worlds of cost containment and quality improvement come closer together based on supportive evidence.

Within this approach at least the following target selection criteria would be used: the presence of a high level of evidence, SMART configurability, and sufficient room for improvement based on local baseline measurement and indications of target specific cost effectiveness (health gain per unit of expense). Finally, it is important to systemize a continuous dynamical approach of quality improvement. Existing Belgian PDCA based initiatives are a source of input herein.

The biggest operational obstacle of P4Q implementation in Belgium (as elsewhere) is the set of conditions of quality measurement once the targets have been defined. P4Q from scratch can be based on already collected data and/or additionally collected data. Already numerous databases are compiled within Belgian healthcare, some for quality purposes, and others for financial or administrative purposes. Therefore, when possible, available data can be used (combined with a thorough validation process). However, data availability should not replace the target selection criteria as presented above. In short term it may be necessary to collect a limited amount of additional data. The bureaucratic workload may be minimized by the use of a random sampling approach of a provider's health care records. As exemplified abroad, automation of data collection is perfectly feasible when nationally supported (as also in Belgium is already largely the case for financial and administrative data). Primary care should be a particular target, since the current quality of data registration at the practice level is currently very low. Recent national initiatives to standardize and integrate health care record IT support are a first step towards automation.

There is a well founded Belgian tradition of risk adjustment of outcome measures, for example in hospital care based on case mix grouping. A similar approach can be used for primary care, based on basic information about patient and provider characteristics.

Although in Belgian initiatives often exclusion criteria are applied to calculate target performance, exception reporting as defined abroad (e.g. in the UK) is in Belgium not used. One exception is the use of variance tracking and analysis as part of care pathway initiatives. Deviations of what is considered expected care are specifically and systematically reported in some of these projects. The difference with the UK approach is the level of standardization of exception (or deviation) arguments. Whereas in the UK only a few predefined arguments are accepted to apply exception reporting, in the Belgian care pathway approach arguments are not predefined. They are however registered and used as an additional input for quality improvement. The care pathway example illustrates the feasibility of exception reporting in Belgium when a stakeholder consensus on the appropriate methods has been reached.

Belgian quality improvement initiatives in general show a lack of monitoring for unintended consequences with regard to care equity for patients, financial equity for providers and equity in terms of attention to the whole of quality priorities. For the first two types no additional data collection is needed; only a specific comparative attention when analyzing target performance. To safeguard equal treatment of different priorities, data on not incentivized targets should be available. These can often be extracted from already existing databases, e.g. from current benchmarking initiatives. This remains a necessary condition of the starting from scratch option.

The P4Q financial incentive could be combined with existing financial mechanisms and with the even more important non financial incentives which drive provider behaviour. An explicit positive financial reward can be attached to performance, with a minimal size of 10% of provider income. This is gradually staged, based on planned budget increases, to guard national budget equilibrium and predicted growth. Initially cost coverage of quality improvement effort is considered the minimum.

This staging can coincide with a phasing of pay for participation (measuring) towards reporting, and finally, performance. Pay for reporting is in some Belgian quality improvement initiatives already the case.

The P4Q incentive is allocated partly to individual providers and partly to providers as a local team, using for example absolute performance criteria. This means that each person's/team's level of performance is incentivized independently, without competition based on mutual performance comparison. Furthermore, the incentive is partly linked to reaching predefined thresholds of performance for each indicator and partly linked to the size of improvement through time.

Initially, each target has an equal weight in incentive allocation. Afterwards adjustments can be made for target specific workload and expected health gain. It is necessary to gather pilot data to devise this weighting structure. Composite measures and an 'all-or-none' approach aren't used during the initial years of P4Q implementation. These options require a higher level of P4Q experience and additional research.

The incentive is closely linked to feedback of performance and provided two or three times a year. The incentive structure remains stabile. Targets are regularly revised based on local room for improvement.

The above described incentive structure reflects the level of complexity of the incentive programme. This seems transparent enough to fit with provider understanding and awareness. In long term the incentive structure may become more complex, requiring specific attention for front office programme communication.

All stakeholders (state, payers, providers, patients and scientific community) are involved during the P4Q design and implementation process. Together they determine through national or regional representation the balance between programme uniformity (e.g. evidence based requirements) and local adaptability (target selection). Professional groups and medical scientific societies have a key role in setting performance standards. In the end all professional providers can decide democratically on the adoption of the programme or not.

After programme design a direct and intensive communication towards individual providers is set up, as part of a broad information platform. Participation in the programme by providers is voluntary. Modelling and pilot testing precedes a wide dissemination of programme implementation.

The programme is embedded into a wider quality improvement support frame. Tools and education are provided to improve and/or maintain performance. Existing quality initiatives can be used to compile such an integrated support platform.

With regard to programme evaluation, target performance is monitored during a sufficiently long time interval, and regularly resampled afterwards. The same is true for a limited selection of not incentivized key targets. The programme as a whole is evaluated using at least a quasi experimental study design, with concurrent and before after comparison points. The support of the use of (cluster) randomization is determined based on a consensus of involved stakeholders. As exemplified abroad, conducting a P4Q RCT, as the best means to minimize selection bias, is a feasible approach. A two phased trial, in which the control group also receives the P4Q intervention during the second phase, supports recruitment of participants in both groups. Because a team (a GP practice and/or a within hospital team) is part of the P4Q focus, cluster randomization should be used.

The alternative of randomization is the use of an observational design such as a cohort study or an interrupted time series. However, a cohort study does not take selection bias into account with regard to unknown confounder variables. The use of an interrupted time series requires a long term study approach with a sufficient number of measurement points. This design supports rather long term monitoring than the assessment of the effects of implementation before dissemination.

The P4Q programme is optimized based on the evaluation results.

7.3 DISCUSSION AND CONCLUSIONS

This chapter addresses the question of feasibility of P4Q within the context of Belgian healthcare. The components of the conceptual framework were used to study feasibility starting from existing quality improvement initiatives and/or starting from scratch.

The option to start from scratch has the advantage of lacking any historically determined restraints which result from an already ongoing quality initiative. Each existing initiative has its own quality definition, goals, targets and quality measurement system in place. Some combine this already with financial incentives, others don't. Therefore, when adding P4Q to these initiatives this might threaten the original programme purpose when done inadequately. In addition, the existing common practice might in itself become an obstacle to state-of-the-art P4Q implementation. This risk of combining both and coming out with none of both adequately accomplished should be carefully considered. A disruption of an ongoing initiative might undo years of effort and evolution. However, these risks neither imply that an existing quality initiative never might evolve towards P4Q integration. As initiatives are regularly evaluated and modified, each initiative might add P4Q characteristics, based on the internally based consensus of long term involved stakeholders. This process describes a kind of natural evolution of existing initiatives towards P4Q.

As described in the results section, it is feasible to adapt existing quality initiatives to include a P4Q component. Although there are some weaknesses (e.g. lack of monitoring of unintended consequences), there are also many strengths (e.g. high experience in implementing and communicating a programme) to support P4Q implementation. The identified weaknesses and threats can be addressed specifically, based on national and international theory, evidence and stakeholder consensus. The analysis of existing initiatives has shown that there is a substantial body of knowledge and experience that can be leveraged to assist in P4Q implementation. It is not about 'reinventing the wheel'. The option to start from scratch therefore shouldn't result in ignoring all lessons learned, on the contrary. Furthermore, the existing quality initiatives can also be used as a target, because most of them are examples of care management processes. P4Q can be aimed at the use of benchmarking, the use of clinical pathways, the use of safety management tools, etc.

Practical feasibility has been confirmed in this chapter. Many of the so called weaknesses are more related to the will to address them as to the ability to address them. The level of data validity is one example often cited as a potential threat to P4Q implementation, due to other purposes of existing data collection and due to the risk of gaming. This argument serves as a primary example of what P4Q in fact stands for: taking quality of care seriously and acting accordingly, based on verifiable quality demanding standards. Just as financial data are strictly monitored using an elaborate accounting system, quality data deserve a similar ardour. There is no room to apply double standards when addressing quality of care as compared to financing of care. P4Q relates both and reformulates priorities. The idea of 'for financing purposes' becomes equal to 'for quality purposes'. Instead of a permissive approach towards gaming, and using this as an argument against P4Q feasibility, P4Q enforces high standard requirements, and the reallocation of resources and effort to ensure those standards. Currently it is considered common sense in every sector that gaming with financial data equals fraud and that a control system is institutionalized. The same kind of high intensity attention, effort and resources are consumed daily to ensure that one's PC receives regular updates to maintain adequate functioning. This same kind of logic (enforcing minimal standards of how to address quality, monitoring those standards, and updating those standards on a very short time span) is not applied on quality of care, the core value and mission of healthcare. As such these are questions of 'willingness' and not of 'ability'.

Finally, it is clear that society evolves towards a higher demand for accountability. Patients expect to receive quality healthcare. P4Q can be used, both for internal as for external accountability. This P4Q driven accountability, based for example on a one month wage quality bonus, requires a more difficult and fundamental shift in how current 'agents' and 'principals' deal with each other. As exemplified in current quality initiatives, at present quality of care is often regarded as an internal issue, with an internal approach, characterized by a high level of freedom and adaptation of quality goals, targets, the measurement system, etc. When financing is linked to quality an independent body needs to verify both the quality improvement process and its performance. This implies a shift towards an external quality auditing approach. Initiatives such as hospital accreditation and the use of the EPA tool fit best with this independent position.

Because it's more a question of will than of feasibility the biggest threat to P4Q implementation is the perception by stakeholders (providers, payers, policymakers, etc.) of P4Q as a threat. Only with positive engagement and support P4Q implementation can succeed. P4Q, if implemented, should therefore become their shared tool, based on a broad involvement and consensus, to realign healthcare with its central mission.

8 PERCEPTIONS AND OPINIONS OF STAKEHOLDERS IN BELGIUM

In this chapter, opinions, quotes and reflections from 40 high-level stakeholders' interviews were put together in a format that follows the conceptual framework described in chapter 3.

Stakeholders' opinions and reflections presented in this chapter include both the items for which consensus exists as well as the differences as they were reported. Three main parts are presented including: a first part on the present Belgian health care system, a second part referring to items related to quality improvement and a third part including a concrete planning with regard to the implementation of pay for quality in Belgium.

This chapter aims at answering the following research question:

To what extent are the current financing scheme, databases and other tools (guidelines, quality indicators) appropriate to implement P4Q in the Belgian health care setting? What are the most important facilitating and hindering factors?

8.1 PART I: THE BELGIAN HEALTH CARE SYSTEM FACING QUALITY OF CARE.

Despite the stakeholders were selected from the leading persons in the Flemish and Walloon part of the Belgian healthcare system, a first important and general finding is that a substantial number of them are not familiar with the concept of pay for quality. The stakeholders often have a limited view on the concept. Their knowledge and experiences with pay for quality are often related to their own area of expertise, respectively hospitals, primary care, pharmaceutical care or insurance. For this reason, pay for quality has to be considered as a relative new concept to the Belgian healthcare system. All stakeholders were willing to reflect openly upon the strengths, weaknesses, opportunities and threats of a potential pay for quality system in Belgium, which is considered a challenging concept, innovative and promising for some, dangerous, unwanted and of limited use to others.

Most of the stakeholders see quality as an intrinsic element in the present Belgian system. They emphasized that it is a normal requirement for all providers to deliver high quality care, with or without a pay for performance system. They suggested also to primarily invest in a better understanding and overview with regard to the intrinsic quality of the present system, before introducing programmes that additionally pay for quality.

Many stakeholders critically pointed out that at present "the P becomes before the Q" within our Belgian health care system, meaning that discussions on payment have a much higher priority compared to the quality that should be related to it.

Many stakeholders clearly believe that in the future an open reflection on quality will be introduced on the political agenda, if not already the case. As there is a demand for an annual budget growth of 4.5% in the years to come, the quality of care will automatically be questioned as society cannot be asked to pay for bad quality. Budgetary constraints, the ageing of the population, the exponential growth of people with chronic conditions and 'voice of the customer' will further provide an impetus in the search for innovative ways to deal with (poor) quality.

A more positive conviction is also put forward that well conceived and transparent quality initiatives can contribute to the legitimacy and cost-effectiveness of the health care system. Pay for quality might provide a new meaning to "accountability" at both the system and individual level.

Some stakeholders insist that quality improvement initiatives should stay or become part of the basic mandatory health care insurance, to avoid selectivity and inequality. It is considered an unacceptable option if quality would be limited to the additional or private insurance sector.

8.1.1 Health care system characteristics

All stakeholders consider key principles of our Belgian system to be individual freedom for patients and providers and a strong dedication in all actors to deliver high quality of care. It has been pointed out during the interviews that there is probably too much (therapeutic) freedom in our healthcare system and that pay for quality might serve as a solution to better deal with quality in general, and with the negative consequences of (therapeutic) freedom in particular.

Another important characteristic of our health care system is that the payment system which is currently in place, -mainly based on fee for service-, implicitly assumes that quality will be delivered at all time and in all places. There are however very few clear-cut quality criteria available that are related to (different levels of) payment (see the 14 examples of quality related initiatives, in chapter 7).

The Belgian healthcare system is also criticized because of its disease-oriented approach, which favours a specialist and hospital focus, its priority for cure over care and even more over prevention, with little emphasis on primary care. For this reason the relation between primary and secondary care is considered as imbalanced, which could be an important barrier towards a movement of integrated care and the development of a more public health related approach.

Another shortcoming of our system, as perceived by a substantial number of stakeholders, is that there is a lack in the transparency of the decision making process at all levels, with no open communication. As quality gradually becomes a topic that moves and concerns many stakeholders, quality issues should be clearly brought on the public agenda.

Most stakeholders conclude that the current level of care is "suboptimal" leaving substantial room for improvement. But before any system of pay for quality can be installed, some stakeholders consider it important to first define what is meant by quality (within the Belgian context), its dimensions, its critical success factors and how it should be measured and promoted. This could prevent resistance to pay for quality initiatives in the long run.

8.1.1.1 Values of the system

Key values of the Belgian healthcare system, as put forward by a few stakeholders, were "accountability" and "professional autonomy". These elements were seen as two sides of the same coin, in the sense that more accountability would potentially lead to less autonomy, and vice versa. In this context improved "transparency" is seen as a value that could facilitate the acceptance of greater levels of accountability in our healthcare system.

All stakeholders that mentioned the value "accessibility" considered it as a crucial value for our health care system as a whole, and when considering the implementation of pay for quality.

Less agreement was noted between the stakeholders for what concerns the value of "privacy" and "confidentially" of data and medical information. Some stakeholders fear that less privacy would lead to less professional autonomy. Appropriateness of care also seemed to be an important value to some stakeholders. Pay for quality should therefore focus on both appropriate care at the right cost, whilst considering the short-term and long-term outcomes.

8.1.1.2 The Belgian health insurance system

A major obstacle in the achievement of quality of care, mentioned by multiple stakeholders is that the Belgian healthcare system is primarily focused on fee for service, and is not explicitly related to the quality of the services provided. As a consequence we do not have a solid tradition in quality measurement nor do we have a yardstick to define the concept of quality/performance. Most stakeholders agree that changes are necessary in the future. Budgetary context will force us towards an increased level of accountability of all the players in the system and widespread implementation strategies for quality including e.g. benchmarking. Another future driver towards quality is considered the ageing of the population and chronic diseases in particular.

The Belgian system is centrally driven, based on public insurers companies, our so called "mutualities". In this context a big divergence could be noted as to the roles of private and public insurers in the definition and execution of pay for quality programmes. Some stakeholders argue that private insurers shouldn't be associated to pay for quality programmes since there is a danger of 'risk selection' and what is called "managed care". Other stakeholders however want to open the debate on the role of private insurers in our health care system. There is strong agreement that it is of no use to simply copying existing pay for quality initiatives from other countries since all health care systems are different.

8.1.1.3 Type of payment system: FFS, capitation, structural or salary

It is clear for all that, if pay for quality programmes would be installed in the future, an adequate financing system should be put in place, which should be more divers than the present system. Most stakeholders consider the fee-for-service system to be a strong incentive for action and service, as it ensures "availability" but not necessary "compliance to guidelines". Pay for quality is seen as an opportunity to enlarge the diversity of paying systems by combining different systems into one programme. Capitation elements can mainly help primary care practices to act on practice populations rather than on individual complaints of patients. Structural financing for e.g. quality management and data monitoring is considered an important impetus for the implementation of pay for quality.

Stakeholders see it as important that for every payment system a good analysis is required on what incentive types or quantities are desired and what are the potential negative consequences. If all the aforementioned elements become part of a transparent plan, many stakeholders are willing to accept budget shifts within the present budgets, shifting profits and economies in one area to support other areas of the health care system. Most suggest that new financing mechanisms will be needed for new pay for quality initiatives, especially for local and regional practice organisation and capacity building.

8.1.1.4 Competition between different subsystems, different levels of care and different providers.

Important elements with regard to competition between different regions, levels of care and providers were put forward by multiple stakeholders. First, several stakeholders pointed out that the cultural differences in the way health care in general, and more specific quality issues are conceptualized and organised strongly differs between the Flemish and Walloon region of Belgium. From a legal perspective, 'quality' is the responsibility of the regions (Flanders and Wallonia), but since the federal government focuses on performance as well, the line between what is quality and what is performance is sometimes difficult to draw.

Second, the competition between secondary and primary care still exists, and many stakeholders stress that quality should be transmural, with strong emphasis on the integration of both levels when implemented. In this context it is important that the differences between hospital and primary care are recognized. Hospital care is considered to be disease-oriented, whereas primary care targets integration, interpersonal relation and individual information.

Third, as competition is still present between individual caregivers in their practices, and the different hospital services, it is important that pay for quality programmes do not increase the level of competition between individual caregivers.

8.1.1.5 Availability of information systems

Availability of good and performing information systems is considered a critical success factor by several stakeholders for the successful implementation of pay for performance programmes. In Belgium we have a substantial amount of (socio-economic) data. There is however an apparent lack of integration of relevant data as is needed to measure performance, rewarding, and for the detection of duplications and overlap in care delivery. Nowadays, the available data are used for 'ad hoc' budgetary decisions and control, but not for epidemiology utility.

As a consequence, very few quality indicators are derived from Belgian data sources and are used in quality measurement, with the exception of e.g. hospitals that regularly receive feed-back derived from centralised databases. The use of valid indicators is a particular problem in primary care that however has the potential to provide important epidemiological data. Support for computerization and data coding systems within the electronic medical record is felt as urgently needed.

Several stakeholders state that there are at present already enough good and interesting data available for public reporting. They considered public reporting already as a good step towards sensitization on quality and quality related issues. In this context, the notion of "guided transparency" was mentioned by one stakeholder, i.e. the public should be helped to read and understand the meaning of the different data. In conclusion: all stakeholders agree that information systems play an important role to facilitate communication between individual caregivers and different levels of the healthcare system, the organisation of feed-back recall systems.

8.1.2 Payer characteristics

Three essential elements were put forward by the stakeholders that mentioned this point with regard to the vision of the payer. These were a) the potential for bias in the view of the payer, and more in particular b) the discrepancy between payers and providers in their views on pay for quality programmes, and c) the potential impact of pay for quality programmes on the repartition of the budget.

Regarding a), a number of stakeholders have pointed out that insurers (payers) might be biased in their views on pay for quality programmes, since they primarily defend the interests of their members. As a consequence, the cost of pay for quality programmes might be considered the primary focus of the payer.

Regarding c, an important advantage of pay for quality programmes that was noted for the payer's perspective was that these programmes will probably contribute to a better repartition of the budget.

8.1.3 Provider characteristics

8.1.3.1 Awareness, perception, familiarity, agreement, self-efficacy

Stakeholders stress the importance of recognizing existing and future 'attitudes' in providers. To give an example, satisfaction in care providers will probably increase if the latter believe they contribute to the quality of care by adhering to evidence based practices. Pay for quality is in this context considered as a means to add value to the (medical) profession, as it induces a reflective attitude of medical doctors on their own practice, which in itself will induce quality. One stakeholder clearly said that participation at pay for quality programme will automatically lead to changes in attitude.

Particularly stakeholders from unions stress that if pay for quality programmes are perceived as assessment, control and interference in their practice, it will automatically lead to important resistance (to change), and thus resistance to the programmes themselves.

The use and reliability of the proposed quality measurement tools must be clearly understood and positively judged by the providers in order a programme to become successful. Communication, transparency again will be very important.

8.1.3.2 Medical leadership, role of peers, role of industry

A substantial number of stakeholders consider medical leadership as a critical success factor for pay for performance programmes. Medical leadership consists of the definition of good clinical practices and aspects of high quality care and requires the presence of persons that are recognized by their peers as true medical experts. Medical leadership is considered by the stakeholders as a shared responsibility between both medical doctors (specialist and general practitioners), but also universities, hospitals as well as the pharmaceutical industry. The process of peer review and feedback is considered essential to create consensus and trust between the actors involved. A particular role is attributed to GLEMS/LOKS, as well as to local general practitioners groups ('cercles', 'kringen'), as forums where medical leadership is to be developed, and in particular where results of pay for quality programmes should be discussed.

8.1.3.3 Existence / implementation of guidelines, room for improvement

The use of guidelines is subscribed as an important element in the implementation of pay for quality programmes. Some stakeholders, not only from the providers' bench, consider guidelines as an important tool, but point out that guidelines are benchmarks, give clear indication when the provided quality is good, but can not be mandatory.

Critical remarks on clinical guidelines were that they are often unrealistic concerning the targets set and often are poorly adjusted to real life situations in primary care. Primary care not only has to adopt these guidelines, but often has to adapt them. Pay for quality programmes, as they often are transmural, regional and include specialist and general practitioners, can potentially contribute to this mutual adoption process, and diminish the diversity of existing guidelines, leading to better agreement and acceptation.

8.1.3.4 Level of own control on changes

A recurrent remark made by several stakeholders is that pay for quality programmes might be perceived as taking control out of the hands of the individual providers. Moreover, pay for quality might be perceived as a system that induces punishments in different ways. Stakeholders indicate that medical specialists might possibly be more resistant to pay for quality compared to primary care physicians. It therefore is considered of high importance that providers become more effectively stimulated to define their own standards of quality of care and take an increased responsibility to define, choose en disseminate their clinical guidelines.

It is recommended that the existing feedback which until now has been a merely financial feedback changes to what is called "multidimensional self–assessment", starting from clinical data.

Especially the providers' representatives pointed out that providers do not have complete control on the patients' contributions in the achievement of quality. They fear that this could lead to undue penalty. There is no way to oblige the patient to follow a treatment or undertake some technical examinations. It makes the evolution of a disease rather independent from the providers' care.

Again providers' representatives fear that official bodies will use transmitted data to control their practices. They have more confidence in their own scientific institutions.

8.1.4 Patient characteristics

The role of the patient, and more in particular compliance of patients to the medical regimen was considered highly important by most stakeholders when discussing clinical outcomes of care. One stakeholder stated that pay for quality programmes will have a low impact on patients, despite their potential influence on outcomes of care.

8.2 PART II: IMPROVING QUALITY THROUGH P4Q

8.2.1 Quality and its dimensions, as seen by Belgian stakeholders

Overall, stakeholders agree that quality has so far not received sufficient attention within our Belgian healthcare system and by policy makers in particular. Most stakeholders explicitly recognize the importance of quality and have often made reference to the gradual growth in the number of quality initiatives that are launched in both hospitals and primary care. Quality definitely needs to become a priority on the political agenda and this is considered a crucial starting point and even a "*conditio sine qua non"* for pay for quality programmes to reach their full potential.

Many stakeholders consider quality as a multidimensional concept of which the different elements are interconnected. When making reference to quality dimensions, the most common cited dimensions were effectiveness, efficiency and safety of care. There is a consensus amongst the stakeholders that high quality management of health care consequently requires a multidimensional approach.

8.2.1.1 Safety of care

As a crucial aspect of quality of care, patient safety was often referred to in relation to structural and process aspects of care. Pay for quality is considered by the stakeholders to potentially enhance (a culture of) fault reduction amongst providers.

8.2.1.2 Equity and access to care

A wide consensus on the importance of equity as a dimension of quality was expressed. Some stakeholders state that, if quality of care has to be improved within the social context of an increasing gap between rich and poor, it must be for all patients, at an affordable cost, proportional to their income. In this context, pay for quality programmes should pay attention to the underuse of services in particular patient populations for which social criteria need to be applied to ensure access to care.

Yet, although accessibility is a crucial aspect of high quality care, it was only mentioned by a few stakeholders, including insurers, hospital chief executive officers and members of regional/community governments. The latter stakeholders emphasized various dimensions of accessibility, i.e. proximity as well as cultural and financial barriers.

Pay for quality programmes should not forget to focus on particular populations that are at risk of low access to services such as migrants and detainees.

8.2.1.3 Effectiveness

Effectiveness of care is considered a basic property to the functioning of our health care system. Two important dimensions, as defined by Campbell et al. (2000), were addressed. These include effectiveness of clinical care and effectiveness of inter-personal care.

The effectiveness of clinical care in Belgium was questioned in terms of both volume and outcomes (e.g. cancer treatment), and for this reason it was doubted that our system can be top-rated when comparing it to other systems. As quality is scarcely measured in Belgium, the aforementioned finding is partly based on private and international reports.

The second dimension of effectiveness of care, interpersonal care, was most often cited by representatives of patient organizations, and primary and secondary care. In this context, the protection of the privileged patient-doctor relationship in general practice is still considered a core value. Some stakeholders therefore argue that revaluation of the intellectual act is important when speaking about pay for quality, and that it a crucial step towards the development of such programmes. Good claim management in hospitals was also cited as a concern in interpersonal care.

8.2.1.4 Patient centeredness

Patient centeredness was highlighted as a dimension of quality of care by mainly stakeholders representing providers, insurers and patients. Patient centeredness is a concept with a lot of different dimensions and operationalized in different ways by different providers. Some relate it to service to patients, positive communication with patients and taking individual responsibility for the patient. Several aspects of patient centeredness were put forward by the stakeholders including service to patients, open communication with patient, taking individual responsibility for the patient, management of the disease by the patients themselves and information management. For what concerns the latter, some stakeholders argue that patients should be informed on the involvement of providers in quality approaches, (particularly when hospitalized), referral to specialists or when informed consent is needed. The importance of privacy and confidentiality was also cited. Patient-centeredness was not presented by the stakeholders as an approach that starts and finishes with patients' expectations and priorities.

As opposed to this, the Belgian healthcare system is considered by some stakeholders as still having strong paternalistic reflexes where the provider and the system know what is good for the patient without asking him/her.

8.2.1.5 Cost-effectiveness, efficiency of care

Most stakeholders recognized the importance of cost-effectiveness as a dimension of quality of care, but it was especially highlighted by Flemish stakeholders.

8.2.1.6 Continuity of services, coordination of care

The dimensions continuity of services and coordination of care were recognized as important dimensions of quality of care. Especially continuous and integrated care were underlined by some stakeholders as an unattained objective in Belgium. Where the focus on integrated care is somewhat more explicit in primary care compared to other levels of the healthcare system, multidisciplinary coordination needs to be improved at all levels, and especially between hospital and primary care. Our current dominant payment system (fee for service) has led to fragmentation of care to a large extent.

Pay for quality is therefore considered a valuable option to improve continuity of care and coordination of services since it allows for a variation in the sources of income in providers that might lead to a broadening of their look on diseases and related processes.

8.2.2 Patient populations and goals for quality projects

8.2.2.1 Defining goals

A central remark made by several stakeholders is that objectives must be valuable to patients, providers, insurers and the government. Stakeholders highlight the need for a preliminary and global assessment of our current health care system first, rather than setting goals for delimitated problems.

When defining goals for pay for performance programmes, stakeholders made reference to the global definition on health, as defined by the World Health Organization (WHO). The WHO defines health as 'a state of optimal physical, social and mental well-being'. For this reason, both clinical care, health promotion/preventive care are aspects of care for which goals should be defined. As a concrete example, a global approach for cardiovascular diseases should combine both medical risk factor management and a health promotion approach for behavioural changes. In this context, some of the stakeholders suggested that it would be necessary to enhance the practitioners' knowledge and competencies in these domains. Another key feature of pay for quality programmes is that they should target the full care continuum and need to bridge the gap between clinical care and public health.

Global goals of pay for quality programmes that were addressed by the stakeholders were under- and overuse of care services, with an accent on underuse of services. Particular goals that were considered important by the stakeholders included the

reduction of the medical practice variability around guidelines and improvements in appropriateness of care. Stakeholders stress the importance of finding the right balance between the relevance of an indicator and the potential effects in providers since discouragement in providers is to be expected when goals are too difficult to reach and no improvement in attitude is to be expected when indicators are too easy to reach.

For some stakeholders, the objectives of pay for quality programmes should be derived from existing data and problem areas that could be derived from it. This implies the need for a qualitative / quantitative data information system that provides good and recent overviews of hospital or primary care output. Some even think that the mere availability of good data could be sufficient to stimulate local creativity so that initiatives could start from the local needs, based on local epidemiology. For others, goal setting should stay centrally, and should start from defined, mainly problematic quality elements, leading to evidence based targets and related indicators

8.2.2.2 Patient populations

Concerning the patient populations that should be targeted in pay for quality programmes, a large consensus was expressed amongst the stakeholders to include patients with chronic diseases, including e.g. diabetes, cancer, heart failure, cardio-vascular diseases, mental diseases and asthma. It is recognized that these diseases need a more global, continuous and integrated approach of care.

Global criteria that were defined for the selection of target populations were: prevalence, the availability of data, the existence of guidelines and whether or not there is room for improvement in the quality of care. Some stakeholders mentioned the importance to have centres of excellence for particular chronic diseases.

Preventive care was also often considered as a priority domain to include patients from: dental care, immunization, smoking and alcohol cessation as well as nutrition related disorders. Some stakeholders even suggested acute diseases, orphan and rare diseases.

Other classifications that were made by the stakeholders were based on social-economic characteristics of the target population and included disadvantaged persons, elderly and teenagers.

8.2.2.3 Processes

The importance of processes as goals for pay for quality programmes was recognized by several stakeholders. However some stakeholders stated that process improvement is of limited value if it does not lead to outcome improvement. Overall, outcomes seemed to be ranked much higher compared to processes, but some argue that process and outcome indicators should be used in a balanced way. A key process that was mentioned is the integration and coordination of care between the different care levels, and within the primary care level between primary care physicians and nurses. Less cited, but probably of no lesser importance were goals including improvements in adherence to guidelines, knowledge update, rational drug prescription and access to information in patients.

Representatives of patient organizations stress the importance of the Patient's Rights Act that is insufficiently known amongst providers, although quality of care is a patient's right. A widespread campaign on this topic and a better implementation of its principles are amongst the goals patient organisations want to support.

8.2.3 Incentives

The topic of incentives is controversial and stakeholders responded in many different ways on the question whether or not there is a need for incentives. Some stakeholders, especially from patient organizations, argued that providers are already paid to provide the necessary quality. Paying 'bonuses' comes down to rewarding what in fact should be standard of care. Other stakeholders didn't see any problem or even insisted on the need for incentives in response to quality delivered. The crucial question seemed to be 'what type of value for what type of money', and more in particular how can we respond in a cost-efficient way to supplementary gains of quality.

Pay for quality is considered a preferred option over the present situation of direct payment without any definition of the desired quality related to it. Incentives that only lead to better data registration and coding, even a good annual report, are considered insufficient.

When the Belgian accreditation system was evoked, stakeholders were straightforward that this kind of incentive to continuous medical education wasn't able to produce any quality output in its current use.

8.2.3.1 Incentive structure

Financial incentives

Financial incentives are considered by the stakeholders as not only very effective, but probably as the only realistic ones that will lead to quality improvement, at least from the providers' point of view. Different types of financial incentives were reported including direct funding, extra salary, company cars or the direct payment of e.g. clerks, software or equipment.

No single negative consequence has been reported on the use of financial incentives in providers; neither did any stakeholder refer to cost-effectiveness. It is suggested that a sort of macro-economic incentives are needed when launching pay for quality programmes, followed by micro-economic incentives when the system is in place. Some stakeholders (primary care physicians) highlight that incentives sometimes might have negative consequences, and will miss their goal, as it was the case with the reward for hallmarked medical record software. On the other hand if new legal initiatives are launched without the presence of an appropriate budget, this might lead to unintended consequences in the sense it becomes perceived as a penalty rather than an incentive. Some stakeholders in this context referred to what they called the structural under financing of the French-speaking Community, and its consequence, the lack of funding for prevention.

Some stakeholders think that direct financial incentives should never be directed to patients in contrast to indirect advantages that are supported in case the patient participates at the programme.

Quality grants/ Financial awards/ Performance funds / Quality infrastructure grants

A few stakeholders suggested quality grants for specific infrastructure including e.g. a minimal package of equipment at the setup of a physician's practice and the IMPULSEO programme. These examples weren't clearly related to quality goals. Particular attention was given to the idea of paying for the development of quality project rather than paying for the achievement of quality indicators. Creating a positive atmosphere by launching a competition, and publicly rewarding selected projects, like the National Council on Quality Promotion has done, was sometimes mentioned as a good example.

Non financial incentives

What concerns the non financial incentives it is important to mention that some stakeholders considered financial incentives just as only one aspect of a global remuneration package. Elements such as quality of life, satisfaction with work, a good working atmosphere, good coordination of care, safety in the working place were equally considered important incentives. In order to improve the working conditions different proposals were made on non financial incentives. These included increased administrative support, practice assistance, extra training or education, coverage of the costs for continuing education, increased social protection and good medical software. Other incentives mentioned were the provision of information and feedback at both the individual and peer level. Very important was that experiences with quality improvements were considered as an incentive in itself.

Stakeholders that represented hospitals and insurers mentioned a particular incentive that is the ranking of both providers and hospitals, but recognized it would be difficult to implement and control. Such ranking should be based on core elements of care, but wasn't specified what these were.

Some stakeholders think that the publication of a ranking (means and/or outcomes), even limited to the providers, is an incentive by itself. Those who are badly ranked are stimulated to improve their actions.

Other stakeholders were very much opposed to the idea of ranking as they were convinced it would not alter behavioural changes. In addition, ranking is perceived as having a negative effect on providers' perception of quality initiatives, and might possibly discourage providers to adopt an open attitude towards quality assessment. International benchmarking however of hospitals is considered a valuable strategy, allowing high level services to be internationally recognised.

Particular incentives for primary care that were reported included annual activity reports or clinical audit, practice evaluation, the payment of quality managers and promotion of the EPA-tool. Public awards were considered to be able to support this kind of initiative. At last, a link with the current accreditation system was suggested, that would implicate an important reconsideration of its role and way of working.

Penalties

Opinions on penalties were strongly divided amongst the stakeholders. Some stakeholders argue that they consider penalties acceptable under certain conditions, whereas others did not support the idea of penalties at all. There was a global consensus that penalties should not be part of a pay for quality programme since rewarding is considered more effective compared to punishing. In this context, current practices of assessment and/or feedback is experienced as being repressive by some providers. However, stakeholders agreed that in case of important outliers, a system of withholding is supported, but after an acceptable delay to correct for these outliers. Consequently, the non-payment of incentives when objectives are not met is considered an adequate 'punishment'. It was of particular concern to patients' stakeholders to avoid any punishment of the patient in case his/her outcome data would not improve, but this idea was not supported in some stakeholders that considered the patient to play an important role in the achievement of the targets set, especially in chronic conditions.

8.2.3.2 Threshold value and / or improvement

The stakeholders gave more weight to real quality improvement and efforts needed to reach quality compared to quality thresholds as such. The aim is however to reward all those who can improve the quality of care and not only those who have substantial room for improvement.

8.2.3.3 Weight of different quality targets

Very few stakeholders evoked this technical aspect of P4Q which was considered a difficult exercise.

8.2.3.4 Size

With regard to the size of the incentive, most stakeholders suggested it should be a minor part of the providers' income, so they would not become dependent upon it. The amount cited ranged from 20 to 40% of the total income (also for hospitals), which is higher than the amounts cited in the international literature. Especially stakeholders that were general practitioners underlined the need for a substantial incentive as they considered primary care to be underfinanced, and they highlighted that especially in primary care there is a lot of quality on the interpersonal, and population care level that is not properly covered by fee for service payment systems. In this context, the success of the English Quality and Outcomes Framework in primary care was attributed to the major financial incentives that are applied. Stakeholders that acted as insurers stressed their own objectives, i.e. cost control of future pay for quality programmes.

8.2.3.5 Stable and long enough

Most stakeholders were thinking of pay for quality programmes as long-term programmes striving for a longitudinal follow-up.

8.2.3.6 Target unit (individual, group / organisation …)

The majority of stakeholders indicated that it is not the individual level of care that should be targeted in pay for quality programmes. Targeting the individual level would be labour-intensive, and would lead to avoiding strategies by providers and probably endless debates on cause-effect relationships with regard to clinical outcomes. Incentives should be directed at professional groups as a whole or to teams as they will facilitate and improve the debate between peers on how to improve the quality of care. Instruments and strategies that were mentioned by several stakeholders in the context of targeting individuals and groups were feed-back and benchmarking.

No specific numbers were mentioned by the stakeholders with regard to the size of units that should be targeted. Examples of levels that were mentioned by the stakeholders were the loco-regional level and the hospital level.

8.2.4 Target audience for P4Q programmes

As target audience for pay for quality programmes both primary and secondary care were mentioned. In this context a preliminary assessment of the opportunities and threats was considered a critical success factor. For some stakeholders among general practitioners and insurers, the primary care setting is theoretically the first to be considered, as it would fit with the need to globally strengthening primary care. On the contrary, if the aim is to seek for economies, the hospital setting is to be considered as some stakeholders think the available budgets aren't spent in an efficient way.

The coordination between primary and secondary care seems to be an important target for some stakeholders representing hospitals, general practitioners and the NIHDI. The current dissymmetric level of empowerment and available budgets between primary and secondary care makes it however difficult to collaborate.

8.2.5 Quality measurement in P4Q programmes

8.2.5.1 General considerations on quality measurement

Although it is often said that Belgium has the best health care system in the world, stakeholders criticize this statement. The perception that our system delivers high quality of care might result from the existing overuse of services and the absence or limited waiting lists. Shortcomings of our health care system is that we do not have sufficient and reliable data and the fact that indexes of performance in specific care needs are established by private companies abroad (e.g. Health Consumer Powerhouse). Insurers' and trade unions' stakeholders do agree on the opportunity that pay for quality programmes represent in terms of quality improvement in order to ensure legitimacy and cost-effectiveness of the health system. On the other hand many stakeholders assume that pay for quality will be difficult to perform because of e.g. the variety of determining factors of quality, the complexity of clinical care and case management and the delays between actions and outcomes in preventive care. A crossover action of different databases, at the local and regional level, will probably be necessary to avoid under- or overestimation. A remaining problem is the difficulty to delimit the customer or population base of general practitioners, mainly in the French part of Belgium, due to the limited use of the global medical record.

There is a general conviction that good indicators can be built from a current set of measures. The set of indicators used for the KCE study 85B (Comparison of cost and quality of two financing systems in primary health care in Belgium - 2008) represents for some stakeholders the best that is available in Belgium.

For some stakeholders, quality measurement should feed a data collection system, aimed at scientific research and feedback to the providers. In this context, the combination of self-assessment (with easy-to-use tools) and an external accreditation system seems to be effective.

8.2.5.2 *Risk adjustment*

A sufficient case-mix within the targeted populations was important for some stakeholders, mainly for general practitioners, patients, insurers and health care authorities. Thus, the risk-adjustment is considered a preliminary condition for pay for quality programmes. To avoid the exclusion of so called 'unwanted patients', an exception reporting should be foreseen.

8.2.5.3 *Indicators*

General consideration about indicators

Many of the stakeholders knew about existing quality indicators. They think that a global assessment tool is needed for the development of a reliable and evolving data set. One stakeholder suggested the construction of personal indicators in primary care, using the "Lot Quality Assessment Sampling" (LQAS) method, allowing general practitioners to perform a self-assessment on their own databases. The aim would be to support voluntary quality improvement, but also to reduce existing barriers, i.e. the fear of control from the NIHDI.

Structure indicators

Although the quality of structural elements of care is often omitted in current quality programmes, some stakeholders argue that these elements are fundamental to both hospital settings and primary care. For the latter, it is stated that they are the easiest to assess. As primary care has no approach on diseases as such, but in terms of global individual care, the European Practice Assesment tool for primary care focuses includes a lot of mainly structural indicators. Examples that were cited for hospital and primary care were: size and satisfaction of staff, complementary services to patients in hospitals, data structure and coding and level of computerization in primary care, availability of up to date material and systems for updating guidelines.

Process indicators

For some stakeholders (general practitioners, hospitals) process assessment and process indicators would seem to be more acceptable compared to outcome indicators since they are more easily controllable and because of their shorter delay for assessment. Moreover, outcome targets are often appraised as specialists' targets whereas general practitioners think more global, accepting much more diversity, and integrate their services and those from others around a specified patient. Quality assurance in general practice needs a global approach valorising existing routines and patient communication. Examples of structure and process indicators that were given were: medical record management, number of global medical records, antibiotic prescription, guideline adherence, permanence of care, duration of consultations, immunization rate, and access time from admission to diagnosis in hospitals and staff scientific activities.

Outcome indicators

A lot of resistance was expressed on outcome assessment and outcome indicators because of the technical difficulties of this assessment, (especially in primary care), the tendency to come forward with false results (because of the link to financial incentives) and the role of the patient in his/her adherence to the medical treatment regimen. Outcome indicators that were cited were: disease-free survival, quality of life, nosocomial infections rate, accuracy in diagnosis, iatrogenic death and surgical errors. We can refer again to the KCE report 85A on capitation versus fee for service payment, whereby several possible outcomes indicators are listed.

8.2.6 Evaluation of the programme

8.2.6.1 Sustainability of change

One GP stakeholder mentioned that objectives of pay for quality programmes should be realistic, in the sense that the objectives of the programme should attainable by most actors involved. The latter is considered a critical success factor of the programme. Another stakeholder (insurer) stressed the importance of practical learning and sharing of experiences to sustain the programme.

8.2.6.2 Validation of the programme

Some stakeholders stated it is important to define in a precise way the objectives of the programme, its standards and quality dimensions in order to assure a solid evaluation of the programme. Other criteria that were considered important for the evaluation of the programme were the participation rate, satisfaction of the participants as well as the budgetary impact on the target population. Effectiveness and efficiency of the programme could also be measured through reports on morbidity and mortality, and intermediary outputs. Various agencies were put forward to do this evaluation including the KCE, a centralized agency or international experts working for a European agency.

8.2.6.3 Review and revising the process

Stakeholders mentioned the need of an initial assessment of the current situation and continuous monitoring of the programmes to detect positive or unintended effects. This should be coordinated at a central (regional or federal) level and based on good cooperation of the professionals.

8.2.6.4 Financial impact and return on investment

For most stakeholders, considering the huge part of the Gross Domestic Product allocated to health, the cost of P4Q programmes should be included in the current budget of the health system with a re-allocation of the current means e.g. from the accreditation budget. One stakeholder pointed out that a particular part of the budget for primary care that is usually not spent should be used to finance pay for quality programmes in primary care.

Some stakeholders highlighted the need for initial and substantial investments at the start of pay for quality programmes. Return on investment and long-term savings should be considered next. Obviously, a programme should only be undertaken once the expected return on investment and cost-effectiveness has been calculated.

8.2.6.5 Unintended consequences

Patient selection

A major unintended consequence that was highlighted by many stakeholders was the risk for patient selection in providers. A second risk as mentioned by the stakeholders was data gaming. There seems to be an apparent lack of trust in providers as expressed by stakeholders from the government, but the issue of data gaming was however expressed by many stakeholders from all settings. Exception reporting is considered important as one must avoid undue penalties of providers when compliance in patients to the medical regimen is poor or absent (confirmed by poor outcomes in indicators). On the other hand exception reporting seems important to follow-up on 'difficult' patients.

Other potential issues that were raised by a lot of stakeholders were the risk for disinterest for problems which are not linked to incentives, which consequently means that pay for quality programmes should target a wide range of diseases. Second, there is a risk for disintegration of care in the sense that a holistic approach towards the patient is less applied. A strong focus on processes and services and the multidisciplinary aspect of care is considered helpful to overcome this risk.

Widening gaps in performance among providers

Some stakeholders expect two basic attitudes in practitioners: those who want to develop their value and impact by subscribing to pay for quality programmes, and others who don't want to involve in it. Various factors will play a role in this, like e.g. part-time working.

8.2.6.6 Organisational system change and extra cost / time required

Many stakeholders mentioned that the programmes should not be too complicated either at the process level or at the assessment level. Otherwise, additional costs, workload and time required to fulfil the programme would be excessive for both general practitioners and small size hospitals**.** Doubts were raised in GP stakeholders on the ability of the government to keep quality programmes simple. To deal with complexity of programmes several suggestions were made: coordination at the local level within group practice in primary care or coordination of the programmes at the national level to support the providers when performing extra tasks related to the implementation of the programme.

8.2.7 Implementing and communicating the programme

8.2.7.1 Involvement of providers in setting goals

For most stakeholders, the involvement of the professionals in the definition of goals seems to be essential. The rationale behind this idea is that providers directly face problems 'in the field', have access to local data and perfectly know the work that has to be performed. They also represent the most direct link to the patient who is the first target unit. Stakeholders also mentioned that the involvement of providers is in itself a way of promoting quality, motivating providers, and preparing them for self-assessment and external evaluation.

Some stakeholders mentioned the need for an institution to help providers in defining goals and to make the link between the different actors of a programme. Some of the existing institutions have, to their point of view, the competencies and the required organisational framework. Organizations/institutions that were mentioned were NIHDI and the Forum des Associations des Généralistes (FAG).

8.2.7.2 Involvement of patients in setting goals

Some stakeholders think patient's representatives should be part of the board which sets the goals of the P4Q programmes as patients are primarily concerned by the quality of healthcare. Moreover patients' representatives are members of many committees at the Ministry of Health, consumers' associations or the Health Promotion Council, giving them a wide, transversally and realistic overview of health care. It is noted that this opinion wasn't supported by any of the providers' representatives (except one) but only by patients themselves, insurers and regional government.

8.2.7.3 Communication to patients

Stakeholders think that the authorities should clarify the definition, concepts and the aim of quality as a first important step towards the development of pay for quality. Then, a clear communication to the patients is required to maximize their commitment to the programme. Stakeholders agree there is a need for increased public accountability towards the patient, as this indicated in related legislation. Stakeholders however strongly differ in their opinions to what extent information about prices and/or quality should be made available to the public.

Many stakeholders state that at least every citizen should be able to consult up-to-date information whether or not a particular hospital or service applies quality improvement approaches and follows up on its performance. Some stakeholders take a step ahead and suggest that the disclosure of a list of indicators that are related to quality in so called 'reference centres', is a mandatory minimum.

The idea of 'ranking' such as hospitals is not supported by the majority of stakeholders. It is said that there is no evidence that this kind of information would be of additional benefit to the patient. In some cases it can have a pervert effect, giving them dissymmetric and biased information.

Overall it is assumed that quality should be defined in agreement with patient representatives.

8.2.7.4 Mandatory or voluntary participation

Many stakeholders subscribe to the principle of voluntary participation to pay for quality programmes, at least as a first step towards the implementation of such programmes. They evoked various reasons for it including e.g. increased chances for success (strongly related to changes in attitudes), improved awareness on the importance of quality in providers and a greater chance for the acceptance of the programme. In contrast to stakeholders that favour voluntary participation other stakeholders defend the idea of mandatory participation mainly because of methodological requirements. However, the latter stakeholders also admitted it would be very difficult to gain widespread acceptance for mandatory participation. Some softened their opinion in proposing the mandatory participation as a final step of implementation or with a self definition of own goals.

8.2.7.5 Staged approach of implementation

It was commonly declared that a progressive and staged approach should be followed to implement P4Q programmes. The main reason is the time required to change attitudes, develop buy-in and to increase motivation.

Some stakeholders proposed to start on a local basis in those domains were success is not too difficult to achieve which would potentially result into high participation rates in providers and patients. Critical success factors for the implementation of pay for quality programmes are defined as: clear role definition, responsibilities and tasks of the parties involved, the development of a common methodology for the design, evaluation and validation of those programmes. In this context, some stakeholders referred to the methodology of clinical trials and action research. Baseline evaluations and periodic assessments were also mentioned. In this way, Belgian stakeholders comply with the international requirements for a staged implementation of pay for quality programmes.

8.2.7.6 Communication to the providers

Clear communication using professional tools is considered necessary to reach the professionals involved. It is recommended that any confusion with programmes that are badly perceived (e.g. accreditation, feedback) is to be avoided.

Some stakeholders stressed the importance of a clear definition of objectives and standards and of the use of evidence based data.

8.2.7.7 Stand alone or embedded in a broader quality project

It is considered important that pay for quality programmes become part of a global vision and plan on quality improvement of the Belgian health care system. Stakeholders stressed that coordination is required to ensure a common approach of the programmes.

8.3 PART III: CONCRETE PLANNING IN THE BELGIAN SITUATION

8.3.1 The need for an "evidence collecting institute"

All stakeholders agree that pay for quality, if implemented, should be based on sound scientific evidence on standards, indicators, and internationally proven ways to achieve them to provide the best guarantees for success. Stakeholders highlight the need for independent and academic scientific advice to help decision makers on the subject. This could be an extended KCE- or WIV/IP type, or an academic consortium type, acting as Independent Scientific Advisory Institute. Others defend the idea of one single national institute that will finally become responsible for the development and implementation of pay for quality programmes in Belgium. Reference is made by some to existing examples in our neighbouring countries: NIVEL or Dutch Institute for Health and Welfare. Only when concepts and instruments are scientifically validated, decision makers can come to action.

8.3.2 A new role for a revised National Council on Quality Promotion (NCQP)

All stakeholders agree that Pay for Quality should be a governmental-led initiative, responsible for the coordination, implementation and follow up. For some of them, it should be initiated and guided by a multidisciplinary initiative, joining academic and public health scientists, providers' representatives (GPs, specialists), trade unions, scientific societies, the National Council of Hospitals, insurers and patients' representatives. This group should define the goals, the content and the implementation strategy for pay for quality programmes.

There is no unanimity which governmental body should lead it. Some mentioned an extended NIHDI group, bringing in nurses and patient representatives. Many stakeholders refer to the National Council for Quality Promotion, established 8 years ago for this purpose. Many propose that, if that choice is made, adaptations are necessary to make the present council more performant, as there is a under representation of the hospital organisations and its directors, the main focus is currently on medical aspects in primary and ambulatory care. There is also a need for a new management style and new impact possibilities. Some even argue a complete make-over of the present body. Or do we need a new specific agency on Quality Implementation issues?

Patient representation is felt important in the organisation of pay for quality, but it is not clear amongst the stakeholders who should represent them: insurers, government, mutualities or specific patient organisations. There is debate on the role of the mutualities in this: some see this as a new function for their future, some state that a conflict of interest will arise.

8.3.3 Think global, act local: the crucial need for organizing at the local level, and paying for local support initiatives

Centrally-led pay for quality initiatives are felt needed, but not to be sufficient by many stakeholders. The idea of decentralisation in pay for quality programmes was supported by multiple stakeholders. Ownership is an important aspect whilst therapeutic freedom is a sensitive issue in Belgian healthcare. Initiatives that are locally developed in individual practices, hospitals or regional networks are considered to provide the best guarantees for success. Best practice examples have to relate to the local situation, opportunities and context. A decentralised pool of interest is necessary in the real local environment. Support, knowledge and qualification should be available at the local level.

Quality managers in hospital as well as in primary care are judged by many stakeholders to play a crucial role in the management of the intermediate level. If this level is important, there should also be foreseen an important budget that includes training, implementation, data-collection, feedback and reflection.

Some stakeholders suggest a sort of intermediate level, between centrally-led and local bodies, to be responsible for the design and implementation of pay for quality programmes, and suggest on giving this task to their scientific societies. In Belgium, although a legal framework exists for a national quality coordinator this is not yet made operational.

8.3.4 How and where to start in Belgium

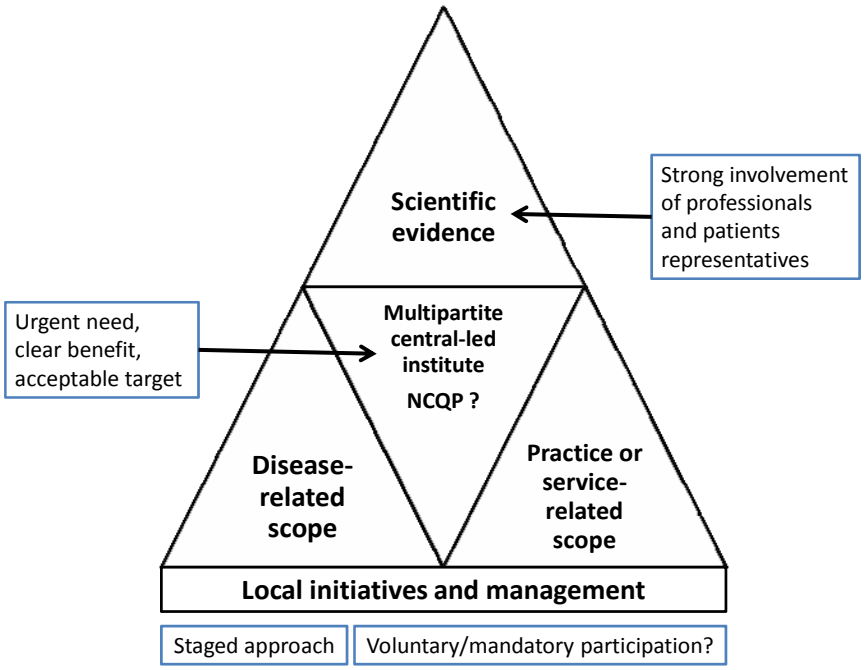
A stepwise approach for the development of a global pay for quality programme is suggested by most stakeholders. As a first step, a broad range of potential quality improving initiatives and projects (that have proven their success in the past) should be listed by the institute that coordinates pay for quality in Belgium. It is suggested that domains for which the needs are urgently displayed, the benefits clearly documented and the targets widely accepted should be considered first. These are often domains where international experience and results is available too. Hospital care and primary care, and in this a target population of medical doctors is considered a priority choice.

There is however no consensus on the type of initiatives that should be launched first. Many propose to start with initiatives that have a traditional disease-related scope, but some advocate a practice or service-related scope. Chronic diseases seem to be the most obvious choice, as the domain of prevention is expected to present important measurement difficulties. Some stakeholders clearly state that mental health is not yet ready for pay for quality systems since the measurement of outcomes is hardly developed.

8.3.5 Schematic view

The figure below visualizes the point of views of stakeholders with regard to the organization of pay for quality in Belgium.

Figure 14: Structure for implementation of P4Q in Belgium



9 DISCUSSION

The aim of this study was to comprehend the large body of evidence related to pay for quality (P4Q), in order to assess feasibility and make suggestions for its implementation in Belgium. We defined P4Q as the use of explicit and direct financial incentives to improve quality of health care.

The concept fits in an evolution from "pay for doing things" (fee for service) towards "pay for doing things right" and "pay for doing the right things" (pay for structure, process and outcome related aspects) and even to "pay to do right" (in which an important equity element is present).

In our country so far, the only pilot based application of this direct financial incentive is related to the improvement of breast cancer screening participation in one Flemish sub region. Other initiatives, however with a less direct link between the financial incentive and improved quality of care exist as well, as described in Chapter 7.

The international literature is overwhelmed by the Anglo-Saxon world and the body of evidence is largely increasing. In addition to that, interviews with international experts and with local Belgian stakeholders (chapter 8) have finally led to some observations, as described in the current chapter.

In general terms, we can conclude that the available evidence from the literature shows that P4Q is no magic bullet. It has the potential to work, with effects sizes mostly varying between 0 and rather positive, and with a very limited number of negative results. The methodological robustness of the evidence shows to be variable, ranging from very weak evidence to strong evidence.

In contrast to evidence on improving average quality, there is lack of evidence on the reduction of *variability* in care, yet this is an objective that should be an integral part of the definition of quality, as also stated by several Belgian stakeholders.

In general, not so many authors have already demonstrated whether P4Q shows undesired effects.

The impact on equity in health care has only been investigated in the Quality Outcomes Framework in the UK. The extent to which different patient groups benefit from P4Q tends to vary and to be highly dependent on the type and complexity of the indicator(s) under study, the observed patient groups (age groups, males versus females, socioeconomic groups or ethnic groups), the characteristics of the study (design, level of analysis, covariates, …) and the level of detail of the studied indicators. Hence, it is difficult to draw firm conclusions regarding equity. In general, all citizens benefit from the improvements in quality of care and the extent to which they benefit determines whether the existing health gap narrows (when the least off have a larger growth than the best off) or increases (if the least off have a smaller growth than the best off). More studies suggest a narrowing of the gap, however, for some indicators new gaps arise. For example a significant difference between the most and least deprived patients emerged after implementing P4Q for the recording of blood pressure, the recording of smoking status and giving smoking advice. Also "pre-P4Q" diabetic women were as likely as men to have their HbA1c, blood pressure, serum creatinine and cholesterol recorded where post-P4Q inequities in these indicators appeared. Further research is needed to understand the mechanisms behind these observations.

Cost-effectiveness of P4Q has not been studied widely either, although from a payer's perspective it is crucial to know whether the money spent in P4Q was well spent money.

To answer the main research question of this project, we can state that implementing P4Q is possible taking into account the considerations made in this report. P4Q programmes can have value if "organised and implemented in a correct way", and as an add-on to other payment mechanisms.

More specifically, our conclusions can be framed according to the conceptual model that was presented in Chapter 3. This model was developed to represent all relevant aspects of Pay for Quality and its application in practice. Central in the model is the relationship between the desired **quality** increase (or maintenance if already a high level was achieved) and the **incentive** that is to be paid. The nature of the incentive and the way that quality is defined and measured can be quite different between programmes. As important is the relationship between the **payer** (who will execute the incentive) and the health care **provider** (who is the target for the incentive and who is supposed to increase or maintain quality of care). Here again, the characteristics of these stakeholders and their relationship will affect the success of a P4Q programme. The model moreover emphasized the need to account for characteristics of the patients, as well as the overall **health system** (i.e. social security or NHS, type of prevailing physician payment system, etc…).

The **implementation** of the programme must follow a "Plan Do Check Act" (PDCA) logic, in which room and efforts are foreseen to regularly seek input from all stakeholders involved and undertake a continuous evaluation of the programme's effects.

The reader should note that conclusions based on evidence from P4Q studies should be regarded with caution because the effect of a P4Q programme can depend from so many factors (incentive size, choice of indicators, involvement of practitioners …) that it is difficult to assign success or failure to one specific aspect of a programme. Modern healthcare organisation should be considered a complex network. Actual complexity research shows that single focused interventions never show simple linear effects.

9.1 QUALITY

9.1.1 Defining quality

In this study, we have defined quality in all its aspects: patient safety, clinical effectiveness, patient centeredness, timeliness, equity and access of care, efficiency and finally continuity and integration. However, only two quality domains are mainly focused upon in P4Q evaluation studies: clinical effectiveness and – to a lesser extent - equity of care. The latter is moreover poorly defined in most studies.

Only a few studies focussed on continuity and integration, with positive effects. Also Belgian stakeholders stated that quality is a transmural concept with strong emphasis on integration.

A disease related focus differs from a more generic focus. Most reported initiatives focus on specific diseases or defined medical problem areas, like diabetes or breast cancer screening. Some are more global and support a more generic involvement in quality issues, like the EPA-initiative in primary care and the Quality and Safety framework for hospitals.

Indicators for quality can relate to structure (e.g. availability of sufficient staff), process (e.g. timely measurement of blood parameters) and outcome (the actual patient's health results).

Current P4Q studies make mostly use of process and intermediate outcome indicators. Structural indicators are used to a lesser extent. Long term outcome indicators are used very rarely. Most evidence comes from primary care (with recently an increase in hospital based studies) and most studies are observational in nature (with a limited number of comparative interventional studies).

The choice of the quality target has a substantial influence on the effect of P4Q. Whereas structural and process targets show in general a more positive effect of P4Q, this is more difficult to reach for intermediate outcome targets (such as HbA1c < 7.4% in diabetes patients). Yet, on the various types of intermediate outcomes often also positive effects were found. This contrasts with long term outcome targets for which no significant effects were detected, which can possibly be explained by the fact that these studies lacked sufficient power, because of the almost zero prevalence of long term complications combined with the time gap in effect. Long term outcomes require long term large scale studies or case control studies.

Almost no study indicated that a previously detected quality problem (high variability or low performance) on a specific target was the reason to include it in a P4Q programme. Most studies implicitly referred to general lack of quality without assessing this in a local context as a first step. A number of studies show an already high performance at baseline (e.g. 80 to 90% achievement on certain included measures), which compromises the effects of the programme, and should be avoided in a possible Belgian implementation. It is generally accepted by experts and stakeholders that quality indicators should be evidence based. But an important remark is made by several stakeholders in that evidence is sometimes too theoretical, and that guidelines are developed by key opinion leaders. In all day medical practice the theoretical objectives are not always easy to achieve, and the full meaning of evidence based (not only scientific evidence, but also context and preferences) is to be taken into account.

Targets should be selected taking into account health care system characteristics and values. For instance, regarding payments systems, in a fee for service system, with inherent risk for overuse, incentives could be related to tackling such overuse. Another example, regarding heath care objectives/values, in a context of prevention (which is a regional responsibility in Belgium), the focus should be on tackling clinical inertia (which is defined as a lack of treatment initiation or intensification in a patient that is not achieving evidence-based goals of care).

It is striking that most studies are focused on the correction of underuse of appropriate care, with varying P4Q results, whereas only two studies focus on the overuse of inappropriate care (lab testing prescription, medical imaging prescription, drug prescription, etc.). One may state that, if both goals could be better balanced this could improve the cost-effectiveness of a possible P4Q programme.

There is an evolution in the number of targets and indicators which are included in P4Q programmes. Programs during the nineties included often only one or a few targets. Later, this number expanded gradually with the initiation of new programmes (cfr. The Quality Outcomes Framework (QOF) in the UK with almost 150 indicators). The effect of simplicity vs. complexity of the P4Q programme is difficult to assess, based on the included studies, and an optimal number of indicators is not really described.

Different weights can be assigned to different indicators according to the workload related with achieving the target, the potential health gain or cost effectiveness. In the UK, thus far, only the workload has been taken into account.

According to the international experts, frequent revisions of the indicator set are necessary, hence targets that are reached can be adjusted and other priority targets can be included to redirect quality improvement resources.

9.1.2 Measuring quality

The way that quality is measured in a P4Q programme is crucial since it will determine the incentive consequences. In most studies, data validity and acceptance were reported as sufficient, perhaps because the decision to initiate a P4Q programme was only taken in settings with presumed or verified (QOF) data validity and acceptance.

An important concern from the provider's perspective is the risk that the data are not sufficiently adjusted for practice/hospital and patient characteristics. As a solution for this, in the UK the concept of exception reporting was introduced. Yet, there are some concerns that exception reporting may be "overused" in order to polish the results somewhat, and that this goes at the cost of equity.

A related concern from the payer's perspective is the risk of "gaming", i.e. wrong or biased reporting. The data collection should therefore be organized as such that gaming becomes almost impossible, for instance by extracting data automatically out of the electronic health record and by setting up an audit system. Because of the workload associated with data collection, it is preferred to make use of existing data as much as possible.

Quality measurement also involves attention for other unintended consequences such as patient selection or shifting attention away from other, unincentivized, healthcare quality priorities. Monitoring of unintended consequences remains therefore important in any new initiative.

It should be clear that careful selection of indicators, together with structural investments in data monitoring and quality management are necessary conditions for P4Q to become successful, as was also stated by the Belgian stakeholders. With this regard, the current lack of integration of existing data, the difficulties to use these for epidemiological purposes and the cumbersome process of obtaining data for research are a key problem, as stated by several stakeholders.

9.2 INCENTIVE

The cornerstone of any P4Q project is the incentive itself. This incentive can be characterized by different aspects such as the size, the nature (bonus or penalty), the frequency, etc…

Incentives of a purely positive nature (rewards) seem to have generated more positive effects than incentive schemes using a competitive approach (in which there are winners and losers). Both international experts as most Belgian stakeholders support rewards rather than penalties. This seems quite logic, but it should be realised that applying rewards entails automatically higher investments, and it should be assessed whether these investments are cost-effective.

Another issue is whether to reward best improvers (those who make the best progress versus the baseline) or best performers (those who achieve a fixed threshold, e.g. >80% of patients with HbA1c below a given level). It is felt by both international experts and Belgian stakeholders that both should be rewarded. The evidence from literature is mixed in this regard.

At present the included studies do not enable to make a further distinction on the effects of different incentive structures (bonus, fee schedule, withhold, regular payment increases, and quality grants), nor is there a clear relationship between incentive size and reported P4Q results. Workable incentive sizes seem to add 5 to 10% to the current income level of the health care providers. Not surprisingly, stakeholders representing provider organisations argued for substantive rewards, whereas stakeholders representing payers stressed more the need for cost control (hence small incentives).

There is an absence of evidence with regard to the choice between direct income stimuli and quality improvement investment stimuli, due to a lack of programmes and studies including the second option. The Quality Outcomes Framework in the UK, which led to mainly positive effects, is based on a combination of both. Practices receive a bonus as part of their operational revenues and can use it to reinforce the practice resources, tools and infrastructure and/or to allocate additional income to individual physicians. This has led to major investments by practices in staffing.

A key question is how frequent incentives should be given and for how long. There is not much evidence with that regard. Based on expert and stakeholder findings, it is advised to opt for a programme in which incentives are given on predefined time points with a sufficient duration over time. In this light it will be interesting to follow up the long term effects of the current Flemish initiative related to breast cancer screening.

9.3 THE OVERALL HEALTH CARE SYSTEM

The overall health care organisation (NHS, managed competition, social security, etc…) may influence P4Q success. National P4Q decision making leads to more uniform P4Q results, as illustrated by the UK example, in contrast with the USA, where different P4Q schemes are not aligned, making it difficult for providers to distinguish the different indicators and involved targets. Belgian stakeholders, focusing on the Belgian healthcare system, where there is (too) much freedom for providers, state that P4Q may serve as a solution to deal with the negative consequences of this freedom. In a system where there is currently too much focus on cure rather than prevention, P4Q may help to improve prevention programmes. With low coordination on chronic conditions, P4Q can offer better task definition and supportive actions.

Our conceptual framework suggested that the general system would be of influence on P4Q results. The limited literature with this regard shows mixed evidence. Many other variables act as concurrent mediators, which might cloud theoretically expected relationships.

One study reports that higher competition between providers for attracting patients is positively related to both incentives to increase desired services as to incentives to decrease undesired services.

In Belgium, it will be important to take into account the existing competition and imbalance between primary and secondary care. P4Q programmes should aim at integrating 1) both levels of care, and 2) cure and prevention. . Chronic diseases can be a good target area to achieve this.

9.4 PAYER CHARACTERISTICS

Within a given health care system, payer characteristics can encompass a wide variety of aspects.

A pre-existing focus of payers on quality before undertaking a P4Q programme is considered important. It is encouraging to observe that in Belgium many different quality initiatives are currently ongoing, as described in Chapter 7. Adding a payment strategy to most of these programmes that show the intrinsic capability for a true pay for quality programme, can bring us a step forward. For instance, the existing EPA-tool in primary care could be used as a basis for additional payment to GPs when they comply with certain international agreed standards in their structure and organisation of practices. Already existing funds could be redirected in this respect, avoiding new investments. Physicians participating in care itineraries ("zorgtrajecten/trajets de soins") could receive a bonus if targets for process indicators or intermediate outcome parameters are achieved.

The typology of the payer (e.g. private/public) to P4Q results is not reported to be significant in the current P4Q evaluation studies, but admittedly this has received little attention.

It could be expected that "for-profit" payers in a competitive environment would be focussed more on cost savings in the short term, and for instance less on equity, while NHS type payers would be focussed more on health targets and equity.

9.5 PROVIDER CHARACTERISTICS

There has been a lack of attention for the effects of (dis)congruence of P4Q with professional culture and with physicians' internal motivation. Through high involvement and democratic decision making when implementing of P4Q it seems that these issues can be addressed, as the UK example shows. But its impact in terms of P4Q results remains unclear, as compared to programmes in other countries where P4Q sometimes was imposed on care providers.

Belgium has a tradition of weak confidence between payers and providers. In the opinion of physicians, the design of and the communication about P4Q programmes should avoid a perception of control, interference in practice and punishment.

According to several stakeholders P4Q may improve a "reflective" attitude among providers, but one should pay attention not to negatively affect the intrinsic motivation of providers to deliver good quality.

Whatever the level of involvement, it is clear that a history of engagement by providers – as a group – with quality improvement activities positively influences P4Q results.

Programs aimed at the individual provider (or small provider group practices) level, rather than at a hospital level or at a regional or national association, report in general positive results. Incentives that are given on a too high level could create a moral hazard problem. This is in contrast to the opinion of most stakeholders who elicit a preference for group based incentives.

Within the small target units, group practices perform better on P4Q than single handed practices according to some studies. However, smaller practice size is also related to other factors such as having patients with poorer health, being located in a deprived area, having more patients from minority ethnic groups, etc. These interrelationships have to be taken into account when assessing the practice size characteristic and its P4Q effects. In the UK there is some evidence that the performance gap between large versus small practices which existed before QOF implementation has disappeared afterwards.

There is mixed evidence on the role of the specialty of the provider and little is known about the hospital sector in general. Only a teaching status of a hospital is positively related to P4Q performance, according to one study.

9.6 PATIENT CHARACTERISTICS

Both experts and Belgian stakeholders emphasize the importance of taking into account patient characteristics when implementing and assessing P4Q programmes, although this is largely under investigated. For instance, there is a lack of research and evidence on the effects of patient educational status and insurance status. There is also a lack of evidence that patient behaviour in terms of lifestyle, cooperation and therapeutic compliance affects P4Q results or is affected by P4Q programmes. Especially providers point to the fact that they have no complete control on the patient's role in achieving quality and they express concerns that this may affect their reward.

Several stakeholders consider the protection of the patient-doctor relationship in general practice as a core value, and some argue that a revalorisation of intellectual tasks (as opposed to technical tasks) may be achieved partially through P4Q.

9.7 IMPLEMENTATION

A necessary condition for implementing P4Q is having funds available, at least by the start of implementation. As the preliminary cost effectiveness results indicate, continuously adding additional funding is no option in the long term.

According to the experts it can be useful to model the costs related to a P4Q programme in advance so that cost-effectiveness can be estimated and payers are not confronted with unpleasant surprises afterwards.

Another requirement is the stepwise introduction of P4Q. In the UK, this was not the case, which has led to the need to make a number of corrections afterwards on a national scale. In the other countries demonstration projects have been used (or are ongoing) before considering national implementation. Some programmes made use of pay for participation and pay for reporting as a first step. It is however at present too early to tell whether the lessons learned in such a phased approach leads to a higher positive impact of P4Q as a result.

A number of studies relate absence of P4Q effects to an absent or insufficient awareness and poor communication. Studies that used more extensive and direct transparent communication about P4Q to the involved providers found more positive effects. Involvement of all stakeholders, in the first place the providers themselves, when developing the P4Q programme proved important, but findings remain mixed (sometimes positive, sometimes no effect). Experts and Belgian stakeholders clearly confirm this need for communication, creating awareness and obtaining involvement.

For successful implementation, stakeholders insist that centrally conceived programmes should be supported by decentralised availability of knowledge, qualified support and local organised input like quality management skills in hospital and primary care, and performing IT systems at all levels.

The role of medical leadership in supporting the P4Q programme is potentially influencing motivation and therefore effectiveness of a P4Q programme, but is rarely reported upon in the studies. Belgian stakeholders considered the role of leadership crucial.

9.8 TOWARDS P4Q IN BELGIUM?

Although many P4Q studies show design problems, leading to mixed evidence on several aspects, elements that ideally should be taken into account when undertaking P4Q initiatives in Belgium can be listed.

P4Q should not be started simply as a nice new idea. It should be made clear why to start, what the current quality issue is, and how it could be addressed with a P4Q programme. Already in this process, all stakeholders should be involved and consulted.

A first key aspect is the definition of quality. The following aspects appear to be of importance, either based on theoretical grounds, or supported by evidence, experts or stakeholders:

- Quality is more than just clinical effectiveness; the different dimensions of quality should be kept in mind. The quality definition and conceptualisation should be in line with both the health system and the provider values.
- When translating quality into indicators, structure, process, and intermediate outcome indicators should all be considered, since they all have their own value (e.g. IT adoption enhancement as a structural goal). But they should all be supported by evidence, and by evidence on room for improvement.
- Consider both increasing appropriate care (reducing clinical inertia) and reducing inappropriate care. In some circumstances, maintaining quality can also be an option.
- Plan already next targets when current targets have been largely achieved (i.e. following a plan do check act approach).

The way that quality achievement will be measured, must also be planned. The following could be taken into account:

- Make use of accurate and validated data, for instance by investing in IT development, and making data collection automatic
- Make use of data already available as much as possible
- Monitor potential unintended consequences (especially in care equity, patient experience and provider experience)
- Apply case mix adjustment on intermediate outcome measures.
- Apply exception reporting to guard individualized care.
- Provide an audit system to prevent and detect gaming. This, together with most of the above elements, requires a well established health information system.
- Include both baseline and comparison group measurements (in the initial phase)

- Take into account the experience of the patient as part of the targets, and during programme development, implementation and evaluation.

Several reflections about the type of incentive can be made:

- Make use of rewards. Of course this also means no payment in case of any performance. The size of this reward could be up to 10% of total payment.
- Find an acceptable balance between rewarding high achievement and rewarding improvement → reward both best performers and best improvers.
- Incentives should be weighted in accordance with the supplementary effort needed to achieve the supplementary quality but also with societal importance (health impact/cost-effectiveness).
- Target incentives at least at the individual provider level, but combine individual incentives with team based incentives when appropriate (to stimulate inter provider collaboration).
- Keep the incentive scheme as simple as possible and easy to communicate
- Make use of a non competitive approach. Budget control can be guarded by applying a corrective factor on all P4Q incentive payments, equal in size for all participants.

The next challenge is to implement gradually the programme, taking into account health care system, payer, provider and patient characteristics. The following could be recommended with this regard:

- Avoid a "one shot", but make use of a cyclical, dynamical quality improvement approach.
- Use a phased approach, i.e. start with a pilot programme, of which lessons can be drawn in preparation of the full programme. "Pilot" can mean a limited region, or it can also mean starting with a limited number of indicators.
- Develop P4Q schemes starting from or together with other quality improvement initiatives.
- Provide a uniform P4Q system (in which local target priorities may vary) from all payers to all participating providers to support transparency, awareness and a sufficient incentive size.
- Make support (for instance IT support), knowledge and qualification available at the local level.
- When implementing P4Q, the medical profession needs to be involved from the start.
- Take into account the level of congruence with professional culture, but realize that P4Q may also support a cultural shift.
- Communicate and create awareness around the planned programme.

Finally, no programme should be started without a guarantee for a correct assessment of its overall impact. It is no use to invest in a P4Q programme when the invested money does not proportionally lead to benefits (either savings or improved health). Therefore:

- Build in a mechanism to avoid exceeding the budget. For instance, a fixed budget to spend in the form of rewards, whereby the reward per provider is smaller when more providers meet the objectives.
- Estimate the cost-effectiveness of the programme already before starting
- Build in a post hoc evaluation of effectiveness and cost-effectiveness of the programme using scientifically valid methods.

In conclusion, we could state, along with several Belgian stakeholders, that well conceived and transparent quality initiatives can contribute to the legitimacy and cost-effectiveness of the health care system. Pay for quality might therefore provide a new meaning to accountability at both the system and individual level.

Both experts and Belgian stakeholders believe that it is important that government and clinical leadership recognize that quality is variable and improvement is important. For a P4Q programme to be successful a cultural shift towards this recognition is needed. P4Q programmes should thereby be seen as part of a range of quality improvement activities and not as stand-alone initiatives. Finally it must be stressed out that for a P4Q programme to work, it is necessary to take into account the lessons learned from past P4Q programmes.

Future research on Pay for Quality should focus on the effects of P4Q on currently less informed domains such as continuity of care, patient and provider experience, and cost effectiveness of program use, as compared to the use of existing implicit financial incentives. To date potential issues in regard to intrinsic motivation effects have been largely neglected in current healthcare research and need specific attention in future studies, making use of survey and qualitative methods. Knowledge from economics, psychology and social sciences can be leveraged further to refine P4Q design recommendations. The international comparison of P4Q within different health systems should be broadened to include a wider implicit incentive scope and to provide explicit attention to the use of incentives in developing countries. Within the context of the Belgian healthcare system, if policy makers decide to refocus incentives toward quality of care, research should shift from an exploratory to a pilot testing phase, making use of demonstration projects.

10 REFERENCES

1. Baker GR, Norton PG, Flintoft V, Blais R, Brown A, Cox J, et al. The Canadian Adverse Events Study: the incidence of adverse events among hospital patients in Canada. Canadian Medical Association Journal. 2004;170(11):1678-86.
2. Institute of Medicine. Rewarding provider performance. Aligning incentives in Medicare. Washington, D.C.: the National Academies Press; 2007.
3. Kohn LT, Corrigan JM, Donaldson MS. To err is human: building a safer health system. Washington DC National Academic Press; 1999.
4. Landon BE, Normand SLT, Lessler A, O'Malley AJ, Schmaltz S, Loeb JM, et al. Quality of care for the treatment of acute medical conditions in US hospitals. Archives of Internal Medicine. 2006;166(22):2511-7.
5. McGlynn EA, Asch SM, Adams J, Keesey J, Hicks J, DeCristofaro A, et al. The quality of health care delivered to adults in the United States. New England Journal of Medicine. 2003;348(26):2635-45.
6. Vincent C, Neale G, Woloshynowych M. Adverse events in British hospitals: preliminary retrospective record review. British Medical Journal. 2001;322(7285):517-9.
7. Zhan C, Miller MR. Excess length of stay, charges, and mortality attributable to medical injuries during hospitalization. JAMA. 2003;290(14):1868-74.
8. Jacques J, Gillain D, Fecher F, Van De Sande S, Vrijens F, Ramaekers D. Studie naar praktijkverschillen bij electieve chirurgische ingrepen in Belgi‰. Brussel: Federaal Kenniscentrum voor de gezondheidszorg (KCE); 2006.
9. Van den Heede K, Sermeus W, Diya L, Lesaffre E, Vleugels A. Adverse outcomes in Belgian acute hospitals: retrospective analysis of the national hospital discharge dataset. International Journal for Quality in Health Care. 2006;18(3):211-9.
10. Altman DE, Clancy C, Blendon RJ. Five years after the IOM report. New England Journal of Medicine. 2004;351(20):2041-3.
11. Porter ME, Teisberg EO. How physicians can change the future of health care. Jama-Journal of the American Medical Association. 2007;297(10):1103-11.
12. Gillen R, Tennen H, McKee T. The impact of the inpatient rehabilitation facility prospective payment system on stroke program outcomes. American Journal of Physical Medicine & Rehabilitation. 2007;86(5):356-63.
13. Gosden T, Forland F, Kristiansen IS, Sutton M, Leese B, Giuffrida A, et al. Capitation, salary, fee-for-service and mixed systems of payment: effects on the behaviour of primary care physicians. Cochrane.Database.Syst.Rev. 2000(3):CD002215.
14. Gray B. Do Medicaid physician fees for prenatal services affect birth outcomes? Journal of Health Economics. 2001;20(4):571-90.
15. Kane RL, Johnson PE, Town RJ, Butler M. Economic incentives for preventive care. Evid.Rep.Technol.Assess.(Summ.). 2004(101):1-7.
16. Stone EG, Morton SC, Hulscher ME, Maglione MA, Roth EA, Grimshaw JM, et al. Interventions that increase use of adult immunization and cancer screening services: a meta-analysis. Ann.Intern.Med. 2002;136(9):641-51.
17. Lewandowski S, O'Connor PJ, Solberg LI, Lais T, Hroscikoski M, Sperl-Hillen JM. Increasing primary care physician productivity: A case study. American Journal of Managed Care. 2006;12(10):573-6.
18. Formulier Va voor het beoordelen van een systematische review van randomised controlled trials. Dutch Cochrane Collaboration Centre 2008.
19. EPOC Methods Paper. Including interrupted time series (ITS) designs in a EPOC review. Cochrane Effective Practice and Organisation of Care Group (EPOC); 1998.
20. Undertaking systematic reviews of research on effectiveness: CRD's guidance for those carrying out or commissioning reviews. York: University of York; 2001.
21. EPOC specific resources for review authors. Cochrane Effective Practice and Organisation of Care Group (EPOC); 2008.

22. Formulier II voor het beoordelen van een randomised controlled trial (RCT). Dutch Conchrane Collaboration Centre; 2008.
23. Formulier III voor het beoordelen van een cohortonderzoek. Dutch Cochrane Collaboration Centre 2008.
24. Boutron I, Moher D, Altman DG, Schulz KF, Ravaud P. Extending the CONSORT statement to randomized trials of nonpharmacologic treatment: Explanation and elaboration. Annals of Internal Medicine. 2008;148(4):295-309.
25. Campbell MK, Elbourne DR, Altman DG. CONSORT statement: extension to cluster randomised trials. British Medical Journal. 2004;328(7441):702-8.
26. Deeks JJ, Dinnes J, D'Amico R, Sowden AJ, Sakarovitch C, Song F, et al. Evaluating non-randomised intervention studies. Health Technol.Assess. 2003;7(27):iii-173.
27. Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. Journal of Epidemiology and Community Health. 1998;52(6):377-84.
28. Higgins JPT, Green S. Cochrane Handbook for Systematic Reviews of Interventions 5.0.0. updated February 2008. The Cochrane Collaboration; 2008.
29. Sanderson S, Tatt LD, Higgins JPT. Tools for assessing quality and susceptibility to bias in observational studies in epidemiology: a systematic review and annotated bibliography. International Journal of Epidemiology. 2007;36(3):666-76.
30. Van de Voorde C, L,onard C. Search for evidence and critical appraisal: health services research (HSR). Brussels Belgian Health Care Knowledge Centre (KCE); 2007.
31. Cleemput I, Van Wilder P, Vrijens F, Huybrechts M, Ramaekers D. Guidelines for Pharmacoeconomic Evaluations in Belgium. Health Technology assessment (HTA). Bruxelles: Belgian Health Care Knowledge Centre (KCE); 2008. KCE Reports 78C (D/2008/10.273/27)
32. Weinstein MC, O'Brien B, Hornberger J, Jackson J, Johannesson M, McCabe C, et al. Principles of good practice for decision analytic modeling in health-care evaluation: report of the ISPOR Task Force on Good Research Practices--Modeling Studies. Value Health. 2003;6(1):9-17.
33. van der Knaap L, Leeuw F, Bogaerts S, Nijssen L. Combining Campbell Standards and the Realist Evaluation Approach. American Journal of Evaluation. 2008;29(1):48-57.
34. Chien AT, Conti RM, Pollack HA. A pediatric-focused review of the performance incentive literature. Curr.Opin.Pediatr. 2007;19(6):719-25.
35. Conrad DA, Christianson JB. Penetrating the "black box": financial incentives for enhancing the quality of physician services. Med.Care Res.Rev. 2004;61(3 Suppl):37S-68S.
36. Dudley RA, Miller RH, Korenbrot TY, Luft HS. The impact of financial incentives on quality of health care. Milbank Quarterly. 1998;76(4):649-+.
37. Dudley RA, Frolich A, Robinowitz DL, Talavera JA, Broadhead P, Luft HS. Strategies to support quality-based purchasing: A review of the evidence. Rockville, MD: Prepared by the Stanford-University of California San Francisco Evidence-based Practice Center; 2004.
38. Frolich A, Talavera JA, Broadhead P, Dudley RA. A behavioral model of clinician responses to incentives to improve quality. Health Policy. 2007;80(1):179-93.
39. Magnus SA. Physicians' financial incentives in five dimensions: A conceptual framework for HMO managers. Health Care Management Review. 1999;24(1):57-72.
40. Town R, Kane R, Johnson P, Butler M. Economic incentives and physicians' delivery of preventive care: a systematic review. Am.J Prev.Med. 2005;28(2):234-40.
41. Town R, Wholey DR, Kralewski J, Dowd B. Assessing the influence of incentives on physicians and medical groups. Med.Care Res.Rev. 2004;61(3 Suppl):80S-118S.
42. Williams TR. Practical design and implementation considerations in pay-for-performance programs. American Journal of Managed Care. 2006;12(2):77-80.
43. Young GJ, Conrad DA. Practical issues in the design and implementation of pay-for-quality programs. Journal of Healthcare Management. 2007;52(1):10-8.
44. Lindenauer PK, Remus D, Roman S, Rothberg MB, Benjamin EM, Ma A, et al. Public reporting and pay for performance in hospital quality improvement. New England Journal of Medicine. 2007;356(5):486-96.
45. Petersen LA, Woodard LD, Urech T, Daw C, Sookanan S. Does pay-for-performance improve the quality of health care? Ann.Intern.Med. 2006;145(4):265-72.

46. Rosenthal MB, Landon BE, Normand ST, Frank RG, Epstein AM. Pay for performance in commercial HMOs. New England Journal of Medicine. 2006;355(18):1895-902.
47. Bonner SE, Sprinkle GB. The effects of monetary incentives on effort and task performance: theories, evidence, and a framework for research. Accounting Organizations and Society. 2002;27(4-5):303-45.
48. Deci EL, Koestner R, Ryan RM. A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. Psychological Bulletin. 1999;125(6):627-68.
49. Gagne M, Deci EL. Self-determination theory and work motivation. Journal of Organizational Behavior. 2005;26(4):331-62.
50. Prendergast C. The provision of incentives in firms. Journal of Economic Literature. 1999; 37 7-63.
51. Rynes SL, Gerhart B, Parks L. Personnel psychology: performance evaluation and pay for performance. Annu.Rev.Psychol. 2005;56:571-600.
52. Lavy V. Using performance-based pay to improve the quality of teachers. Future.Child. 2007;17(1):87-109.
53. Salber PR, Bradley BE. Adding quality to the health care purchasing equation. Health Aff.(Millwood.). 2001;Suppl Web Exclusives:W93-W5.
54. Young GJ, White B, Burgess JF, Berlowitz D, Meterko M, Guldin MR, et al. Conceptual issues in the design and implementation of pay-for-quality programs. American Journal of Medical Quality. 2005;20(3):144-50.
55. Roland M. Linking physicians' pay to the quality of care--a major experiment in the United kingdom. N.Engl.J.Med. 2004;351(14):1448-54.
56. Vlayen J, Van De Water G, Camberlin C, Paulus D, Leys M, Ramaekers D. Klinische kwaliteitsindicatoren. Brussel: Federaal Kenniscentrum voor de Gezondheidszorg (KCE); 2006.
57. Institute of Medicine. Crossing the Quality Chasm: a new health system for the 21st century. 2001.
58. Whitehead M, Dahlgren G. Concepts and principles for tackling inequalities in health: Levelling up Part 1. W.R.O.F. Europe. Copenhagen, World Health Organisation. 2006.
59. Aday LA, Begley CE, Lairson DR, Slater CH, Richard AJ, Montoya ID. A framework for assessing the effectiveness, efficiency, and equity of behavioral healthcare. Am.J.Manag.Care. 1999;5 Spec No:SP25-SP44.
60. Braveman P. Health disparities and health equity: concepts and measurement. Annu.Rev.Public Health. 2006;27:167-94.
61. Chang WC. The meaning and goals of equity in health. J.Epidemiol.Community Health. 2002;56(7):488-91.
62. Whitehead M, Dahlgren G, Evans T. Equity and health sector reforms: can low-income countries escape the medical poverty trap? Lancet. 2001;358(9284):833-6.
63. Goddard M, Smith P. Equity of access to health care services: theory and evidence from the UK. Soc.Sci.Med. 2001;53(9):1149-62.
64. Ward PR. The relevance of equity in health care for primary care: creating and sustaining a 'fair go, for a fair innings'. Qual.Prim.Care. 2009;17(1):49-54.
65. Willems S. The socio-economic gradient in health: a never-ending story? A descriptive and explorative study in Belgium: Department of General Practice and Primary Health Care. Ghent University; 2003.
66. Bradshaw J. The conceptualisation and measurement of need. In: PopayJ, G WI, editors. Researching the people's Health. London: Routledge; 1994. p. 45-57.
67. Duckett SJ. Design of price incentives for adjunct policy goals in formula funding for hospitals and health services. BMC Health Serv.Res. 2008;8:72.
68. Hackbarth G, Milgate K. Using quality incentives to drive physician adoption of health information technology. Health Affairs. 2005;24(5):1147-9.
69. Rosenthal MB, de Brantes FS, Sinaiko AD, Frankel M, Robbins RD, Young S. Recognizing High-Quality Care: Analysis of Physician Quality and Resource Use. American Journal of Managed Care. 2008;14(10):670-7.

70. Bradley EH, Herrin J, Elbel B, McNamara RL, Magid DJ, Nallamothu BK, et al. Hospital quality for acute myocardial infarction: correlation among process measures and relationship with short-term mortality. JAMA. 2006;296(1):72-8.

71. Griffith JR, Knutzen SR, Alexander JA. Structural versus outcomes measures in hospitals: a comparison of Joint Commission and Medicare outcomes scores in hospitals. Qual.Manag.Health Care. 2002;10(2):29-38.

72. Mant J. Process versus outcome indicators in the assessment of quality of health care. International Journal for Quality in Health Care. 2001;13(6):475-80.

73. Werner RM, Bradlow ET. Relationship between Medicare's hospital compare performance measures and mortality rates. JAMA. 2006;296(22):2694-702.

74. Rosenthal MB, Frank RG, Li ZH, Epstein AM. From concept to practice. Jama-Journal of the American Medical Association. 2005;294(14):1788-93.

75. Custers T, Hurley J, Klazinga NS, Brown AD. Selecting effective incentive structures in health care: A decision framework to support health care purchasers in finding the right incentives to drive performance. BMC.Health Serv.Res. 2008;8:66.

76. Eggleston K. Multitasking and mixed systems for provider payment. Journal of Health Economics. 2005;24(1):211-23.

77. Shahian DM, Edwards FH, Ferraris VA, Haan CK, Rich JB, Normand SLT, et al. Conceptual framework and measure selection. Annals of Thoracic Surgery. 2007;83(4):S3-S12.

78. Dudley R.A, Rosenthal M.B. Pay for Performance: A Decision Guide for Purchasers. 2006.

79. Sachdeva RC. Electronic healthcare data collection and pay-for-performance: translating theory into practice. Ann.Health Law. 2007;16(2):291-311, table.

80. Kuttner R. The risk-adjustment debate. New England Journal of Medicine. 1998;339(26):1952-6.

81. Meddings JA, McMahon LF. From 'one-size-fits-all' measures to individual patient risk-reduction scores. Disease Management & Health Outcomes. 2008;16(4):205-16.

82. Mehta RH, Liang L, Karve AM, Hernandez AF, Rumsfeld JS, Fonarow GC, et al. Association of patient case-mix adjustment, hospital process performance rankings, and eligibility for financial incentives. Jama-Journal of the American Medical Association. 2008;300(16):1897-903.

83. Tabak YP, Johannes RS, Silber JH. Development and validation of six disease-specific mortality predictive models for pay-for-performance. Medical Care. 2007;45(8):789-805.

84. Doran T, Fullwood C, Reeves D, Gravelle H, Roland M. Exclusion of patients from pay-for-performance targets by english physicians. New England Journal of Medicine. 2008;359(3):274-84.

85. Larriviere DG, Bernat JL. Invited article: Threats to physician autonomy in a performance-based reimbursement system. Neurology. 2008;70(24):2338-42.

86. Lewis JM, Marjoribanks T. The impact of financial constraints and incentives on professional autonomy. International Journal of Health Planning and Management. 2003;18(1):49-61.

87. Mannion R, Davies H. Payment for performance in health care. British Medical Journal. 2008;336(7639):306-8.

88. Werner RM, Asch DA. Clinical concerns about clinical performance measurement. Annals of Family Medicine. 2007;5(2):159-63.

89. Forrest CB, Villagra VG, Pope JE. Managing the metric vs managing the patient: The physician's view of pay for performance. American Journal of Managed Care. 2006;12(2):83-5.

90. Ganz DA, Wenger NS, Roth CP, Kamberg CJ, Chang JT, MacLean CH, et al. The law of unintended consequences? Medical Care. 2007;45(1):8-18.

91. Global Health Workforce Alliance. Guidelines: Incentives for Health Professionals. Available from: http://www.who.int/workforcealliance/documents/Incentives_Guidelines%20EN.pdf

92. Chaix-Couturier C, Durand-Zaleski I, Jolly D, Durieux P. Effects of financial incentives on medical practice: results from a systematic review of the literature and methodological issues. Int J Qual.Health Care. 2000;12(2):133-42.

93. Rodwin MA. Have their place but need to be evaluated and used to promote appropriate goals. British Medical Journal. 2004;328(7452):1328-9.

94. Robinson JC. Theory and practice in the design of physician payment incentives. Milbank Quarterly. 2001;79(2):149-+.
95. Averill RF, Vertrees JC, McCullough EC, Hughes JS, Goldfield NI. Redesigning medicare inpatient PPS to adjust payment for post-admission complications. Health Care Financing Review. 2006;27(3):83-93.
96. Pronovost PJ, Goeschel CA, Wachter RM. The wisdom and justice of not paying for "preventable complications". Jama-Journal of the American Medical Association. 2008;299(18):2197-9.
97. Rosenthal MB. Nonpayment for performance? Medicare's new reimbursement rule. New England Journal of Medicine. 2007;357(16):1573-5.
98. Cromwell J, Drozd EM, Smith K, Trisolini M. Financial gains and risks in pay-for-performance bonus algorithms. Health Care Financing Review. 2007;29(1):5-14.
99. Nolan T, Berwick DM. All-or-none measurement raises the bar on performance. Jama-Journal of the American Medical Association. 2006;295(10):1168-70.
100. Reeves D, Campbell SM, Adams J, Shekelle PG, Kontopantelis E, Roland MO. An evaluation of different analytic approaches. Medical Care. 2007;45(6):489-96.
101. Caldis T. Composite health plan quality scales. Health Care Financing Review. 2007;28(3):95-107.
102. Ferreira-Gonzalez I, Busse JW, Heels-Ansdell D, Montori VM, Akl EA, Bryant DM, et al. Problems with use of composite end points in cardiovascular trials: systematic review of randomised controlled trials. British Medical Journal. 2007;334(7597):786-8A.
103. Freemantle N, Calvert M, Wood J, Eastaugh J, Griffin C. Greater precision but with greater uncertainty? Jama-Journal of the American Medical Association. 2003;289(19):2554-9.
104. Freemantle N, Calvert M. Composite end points may mislead - and regulators allow it to happen. British Medical Journal. 2007;334(7597):756-7.
105. Montori VM, Permanyer-Miralda G, Ferreira-Gonzalez I, Busse JW, Pacheco-Huergo V, Bryant D, et al. Validity of composite end points in clinical trials. British Medical Journal. 2005;330(7491):594-6.
106. O'Brien SM, Delong ER, Dokholyan RS, Edwards FH, Peterson ED. An example from coronary artery bypass surgery. Circulation. 2007;116(25):2969-75.
107. Nahra TA, Reiter KL, Hirth RA, Shermer JE, Wheeler JRC. Cost-effectiveness of hospital pay-for-performance incentives. Medical Care Research and Review. 2006;63(1):49S-72S.
108. Eisenhardt KM. Agency theory:Anassessment and review. The Academy ofManagementReview. 1989;14(1):57-74.
109. Figueras J, Robinson R, Jakubowski E. Purchasing to improve health systems performance. Open University Press; 2005.
110. Buetow S. What motivates health professionals? Opportunities to gain greater insight from theory. J Health Serv.Res.Policy. 2007;12(3):183-5.
111. Marshall M, Harrison S. It's about more than money: financial incentives and internal motivation. Qual.Saf Health Care. 2005;14(1):4-5.
112. Adams O, Hicks V. Pay and Non-pay Incentives, Performance and Motivation Prepared for WHO's December 2000 Global Health Workforce Strategy Group, Geneva. Human Resources for Health Development Journal (HRDJ). 2000;4(3):126-45.
113. Wazana A. Physicians and the pharmaceutical industry: is a gift. Journal of the American Medical Association. 2000;283:373-80.
114. Berenson RA, Ginsburg PB, May JH. Hospital-physicians relations: cooperation, competition, or separation? Health Aff.(Millwood.). 2007;26(1):w31-w43.
115. Berwick DM, DeParle NA, Eddy DM, Ellwood PM, Enthoven AC, Halvorson GC, et al. Paying for performance: Medicare should lead. Health Aff.(Millwood.). 2003;22(6):8-10.
116. Safavi K. Aligning financial incentives. J.Healthc.Manag. 2006;51(3):146-51.
117. Gallagher TH, Levinson W. A prescription for protecting the doctor-patient relationship. American Journal of Managed Care. 2004;10(2):61-8.
118. Boyd CM, Darer J, Boult C, Fried LP, Boult L, Wu AW. Implications for pay for performance. Jama-Journal of the American Medical Association. 2005;294(6):716-24.

119. Tinetti ME, Bogardus ST, Agostini JV. Potential pitfalls of disease-specific guidelines for patients with multiple conditions. New England Journal of Medicine. 2004;351(27):2870-4.

120. Walter LC, Davidowitz NP, Heineken PA, Covinsky KE. Lessons learned from a VA performance measure. Jama-Journal of the American Medical Association. 2004;291(20):2466-70.

121. Grol RP, Bosch MC, Hulscher ME, Eccles MP, Wensing M. Planning and studying improvement in patient care: the use of theoretical perspectives. Milbank Q. 2007;85(1):93-138.

122. McGuire W. Attitudes and Attitude Change. In: The Handbook of Social Psychology 1985. p. 233-46.

123. Pine M, Jordan HS, Elixhauser A, Fry DE, Hoaglin DC, Jones B, et al. Enhancement of claims data to improve risk adjustment of hospital mortality. Jama-Journal of the American Medical Association. 2007;297(1):71-6.

124. Hippisley-Cox J, O'Hanlon S, Coupland C. Association of deprivation, ethnicity, and sex with quality indicators for diabetes: population based survey of 53 000 patients in primary care. British Medical Journal. 2004;329(7477):1267-9.

125. Lynch M. Effect of Practice and Patient Population Characteristics on the Uptake of Childhood Immunizations. British Journal of General Practice. 1995;45(393):205-8.

126. MacBride-Stewart SP, Elton R, Walley T. Do quality incentives change prescribing patterns in primary care? An observational study in Scotland. Family Practice. 2008;25(1):27-32.

127. McGovern MP, Boroujerdi MA, Taylor MW, Williams DJ, Hannaford PC, Lefevre KE, 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;25(1):33-9.

128. McGovern MP, Williams DJ, Hannaford PC, Taylor MW, Lefevre KE, Boroujerdi MA, 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. 2008;25(9):1083-9.

129. McLean G, Guthrie B, Sutton M. Differences in the quality of primary medical care services by remoteness from urban settlements. Quality & Safety in Health Care. 2007;16(6):446-9.

130. McLean G, Guthrie B, Sutton M. Differences in the quality of primary medical care for CVD and diabetes across the NHS: evidence from the quality and outcomes framework. Bmc Health Services Research. 2007;7.

131. Millett C, Car J, Eldred D, Khunti K, Mainous AG, Majeed A. Diabetes prevalence, process of care and outcomes in relation to practice size, caseload and deprivation: national cross-sectional study in primary care. Journal of the Royal Society of Medicine. 2007;100(6):275-83.

132. Ritchie LD, Bisset AF, Russell D, Leslie V, Thomson I. Progress and the 1990 Contract. British Medical Journal. 1992;304(6830):816-9.

133. Saxena S, Car J, Eldred D, Soljak M, Majeed A. Practice size, caseload, deprivation and quality of care of patients with coronary heart disease, hypertension and stroke in primary care: national cross-sectional study. Bmc Health Services Research. 2007;7.

134. 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(9):2354-60.

135. Simpson CR, Hannaford PC, McGovern M, Taylor MW, Green PN, Lefevre K, et al. Are different groups of patients with stroke more likely to be excluded from the new UK general medical services contract? A cross-sectional retrospective analysis of a large primary care population. Bmc Family Practice. 2007;8.

136. Sutton M, McLean G. Determinants of primary medical care quality measured under the new UK contract: cross sectional study. British Medical Journal. 2006;332(7538):389-90.

137. Wang YY, O'Donnell CA, Mackay DF, Watt GCM. Practice size and quality attainment under the new GMS contract: a cross-sectional analysis. British Journal of General Practice. 2006;56(532):830-5.

138. Williams PH, De Lusignan S. Does a higher 'quality points' score mean better care in stroke? An audit of general practice medical records. Inform.Prim.Care. 2006;14(1):29-40.

139. Cupples ME, Smith SM, Murphy AW. How effective is prevention in coronary heart disease? Heart. 2008;94(11):1370-1.

140. Ashworth M, Armstrong D, de Freitas J, Boullier G, Garforth J, Virji A. The relationship between income and performance indicators in general practice: a cross-sectional study. Health Serv.Manage.Res. 2005;18(4):258-64.

141. Ashworth M, Lloyd D, Smith RS, Wagner A, Rowlands G. Social deprivation and statin prescribing: a cross-sectional analysis using data from the new UK general practitioner 'Quality and Outcomes Framework'. Journal of Public Health. 2007;29(1):40-7.

142. Ashworth M, Seed P, Armstrong D, Durbaba S, Jones R. The relationship between social deprivation and the quality of primary care: a national survey using indicators from the UK Quality and Outcomes Framework. British Journal of General Practice. 2007;57(539):441-8.

143. Campbell S, Reeves D, Kontopantelis E, Middleton E, Sibbald B, Roland M. Quality of primary care in England with the introduction of pay for performance. New England Journal of Medicine. 2007;357(2):181-90.

144. Coleman K, Reiter KL, Fulwiler D. The impact of pay-for-performance on diabetes care in a large network of community health centers. Journal of Health Care for the Poor and Underserved. 2007;18(4):966-83.

145. Coleman T, Lewis S, Hubbard R, Smith C. Impact of contractual financial incentives on the ascertainment and management of smoking in primary care. Addiction. 2007;102(5):803-8.

146. Doran T. Lessons from early experience with pay for performance. Disease Management & Health Outcomes. 2008;16(2):69-77.

147. Doran T, Fullwood C, Gravelle H, Reeves D, Kontopantelis E, Hiroeh U, et al. Pay-for-performance programs in family practices in the United Kingdom. New England Journal of Medicine. 2006;355(4):375-84.

148. 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;372(9640):728-36.

149. Downing A, Rudge G, Cheng Y, Tu YK, Keen J, Gilthorpe MS. Do the UK government's new Quality and Outcomes Framework (QOF) scores adequately measure primary care performance? A cross-sectional survey of routine healthcare data. Bmc Health Services Research. 2007;7.

150. Gray J, Millett C, Saxena S, Netuveli G, Khunti K, Majeed A. Ethnicity and quality of diabetes care in a health system with universal coverage: Population-based cross-sectional survey in primary care. Journal of General Internal Medicine. 2007;22(9):1317-20.

151. Gulliford MC, 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;24(5):505-11.

152. Jaiveer PK, Jaiveer S, Jujjavarapu SB, Morrissey J, White J, Gadsby R, et al. Improvements in clinical diabetes care in the first year of the new General Medical Services contract in the UK. The British Journal of Diabetes and Vascular Disease. 2006;6 45-8.

153. Langham S, Gillam S, Thorogood M. How Have Changes in Financial Incentives Affected Health Promotion Activity in General-Practice. British Journal of General Practice. 1995;45(401):665-8.

154. Millett C, Gray J, Bottle A, Majeed A. Ethnic disparities in blood pressure management in patients with hypertension after the introduction of pay for performance. Ann.Fam.Med. 2008;6(6):490-6.

155. 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. 2007;4(6):1087-93.

156. 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. Canadian Medical Association Journal. 2007;176(12):1705-10.

157. Millett C, Gray J, Wall M, Majeed A. Ethnic Disparities in Coronary Heart Disease Management and Pay for Performance in the UK. J Gen.Intern.Med. 2008.

158. Millett C, Netuveli G, Saxena S, Majeed A. Impact of pay for performance on ethnic disparities in intermediate outcomes for diabetes: longitudinal study. Diabetes Care. 2008.

159. Shohet C, Yelloly J, Bingham P, Lyratzopoulos G. The association between the quality of epilepsy management in primary care, general practice population deprivation status and

epilepsy-related emergency hospitalisations. Seizure-European Journal of Epilepsy. 2007;16(4):351-5.

160. Sigfrid LA, Turner C, Crook D, Ray S. Using the UK primary care Quality and Outcomes Framework to audit health care equity: preliminary data on diabetes management. Journal of Public Health. 2006;28(3):221-5.
161. Srirangalingam U, Sahathevan SK, Lasker SS, Chowdhury TA. Changing pattern of referral to a diabetes clinic following implementation of the new UK GP contract. British Journal of General Practice. 2006;56(529):624-6.
162. Steel N, Bachmann M, Maisey S, Shekelle P, Breeze E, Marmot M, et al. Self reported receipt of care consistent with 32 quality indicators: national population survey of adults aged 50 or more in England. British Medical Journal. 2008;337(7667).
163. Steel N, Maisey S, Clark A, Fleetcroft R, Howe A. Quality of clinical primary care and targeted incentive payments: an observational study. Br.J Gen.Pract. 2007;57(539):449-54.
164. Strong M, Maheswaran R, Radford J. Socioeconomic deprivation, coronary heart disease prevalence and quality of care: a practice-level analysis in Rotherham using data from the new UK general practitioner Quality and Outcomes Framework. Journal of Public Health. 2006;28(1):39-42.
165. Tahrani AA, McCarthy M, Godson J, Taylor S, Slater H, Capps N, et al. Diabetes care and the new GMS contract: the evidence for a whole county. British Journal of General Practice. 2007;57(539):483-5.
166. Tahrani AA, McCarthy M, Godson J, Taylor S, Slater H, Capps N, et al. Impact of practice size on delivery of diabetes care before and after the Quality and Outcomes Framework implementation. British Journal of General Practice. 2008;58(553):576-9.
167. Vaghela P, Ashworth M, Schofield P, Gulliford MC. Population intermediate outcomes of diabetes under pay for performance incentives in England from 2004 to 2008. Diabetes Care. 2008.
168. Wright J, Martin D, Cockings S, Polack C. Overall Quality of Outcomes Framework scores lower in practices in deprived areas. British Journal of General Practice. 2006;56(525):277-9.
169. Ashworth M, Medina J, Morgan M. Effect of social deprivation on blood pressure monitoring and control in England: a survey of data from the quality and outcomes framework. BMJ. 2008;337:a2030.
170. Armour BS, Friedman C, Pitts MM, Wike J, Alley L, Etchason J. The influence of year-end bonuses on colorectal cancer screening. American Journal of Managed Care. 2004;10(9):617-24.
171. Greene RA, Beckman H, Chamberlain J, Partridge G, Miller M, Burden D, et al. Increasing adherence to a community-based guideline for acute sinusitis through education, physician profiling, and financial incentives. American Journal of Managed Care. 2004;10(10):670-8.
172. Pearson SD, Schneider EC, Kleinman KP, Coltin KL, Singer JA. The impact of pay-for-performance on health care quality in Massachusetts, 2001-2003. Health Affairs. 2008;27(4):1167-76.
173. Shenkman E, Tian LL, Nackashi J, Schatz D. Managed care organization characteristics and outpatient specialty care use among children with chronic illness. Pediatrics. 2005;115(6):1547-54.
174. Glickman SW, Ou FS, Delong ER, Roe MT, Lytle BL, Mulgund J, et al. Pay for performance, quality of care, and outcomes in acute myocardial infarction. Jama-Journal of the American Medical Association. 2007;297(21):2373-80.
175. Grossbart SR. What's the return? Assessing the effect of "pay-for-performance" initiatives on the quality of care delivery. Medical Care Research and Review. 2006;63(1):29S-48S.
176. Karve AM, Ou FS, Lytle BL, Peterson ED. Potential unintended financial consequences of pay-for-performance on the quality of care for minority patients. American Heart Journal. 2008;155(3):571-6.
177. Kouides RW, Lewis B, Bennett NM, Bell KM, Barker WH, Black ER, et al. A Performance-Based Incentive Program for Influenza Immunization in the Elderly. American Journal of Preventive Medicine. 1993;9(4):250-5.

178. Kouides RW, Bennett NM, Lewis B, Cappuccio JD, Barker WH, LaForce FM. Performance-based physician reimbursement and influenza immunization rates in the elderly. American Journal of Preventive Medicine. 1998;14(2):89-95.

179. Mentari EK, Deoreo PB, O'Connor AS, Love TE, Ricanati ES, Sehgal AR. Changes in Medicare reimbursement and patient-nephrologist visits, quality of care, and health-related quality of life. American Journal of Kidney Diseases. 2005;46(4):621-7.

180. Fairbrother G, Friedman S, Hanson KL, Butts GC. Effect of the vaccines for children program on inner-city neighborhood physicians. Archives of Pediatrics & Adolescent Medicine. 1997;151(12):1229-35.

181. Fairbrother G, Hanson KL, Friedman S, Butts GC. The impact of physician bonuses, enhanced fees, and feedback on childhood immunization coverage rates. American Journal of Public Health. 1999;89(2):171-5.

182. Fairbrother G, Hanson KL, Butts GC, Friedman S. Comparison of preventive care in Medicaid managed care and medicaid fee for service in institutions and private practices. Ambulatory Pediatrics. 2001;1(6):294-301.

183. Hillman AL, Ripley K, Goldfarb N, Nuamah I, Weiner J, Lusk E. Physician financial incentives and feedback: Failure to increase cancer screening in Medicaid managed care. American Journal of Public Health. 1998;88(11):1699-701.

184. Hillman AL, Ripley K, Goldfarb N, Weiner J, Nuamah I, Lusk E. The use of physician financial incentives and feedback to improve pediatric preventive care in Medicaid managed care. Pediatrics. 1999;104(4):931-5.

185. Pourat N, Rice T, Tai-Seale M, Bolan G, Nihalani J. Association between physician compensation methods and delivery of guideline-concordant STD care: Is there a link? American Journal of Managed Care. 2005;11(7):426-32.

186. Rittenhouse DR, Robinson JC. The use of care management processes for chronic illness and preventive care. Medical Care. 2006;44(1):47-54.

187. Young GJ, Meterko M, Beckman H, Baker E, White B, Sautter KM, et al. Effects of paying physicians based on their relative performance for quality. Journal of General Internal Medicine. 2007;22(6):872-6.

188. Morrow RW, Gooding AD, Clark C. Improving physicians' preventive health care behavior through peer review and financial incentives. Arch.Fam.Med. 1995;4(2):165-9.

189. Casale AS, Paulus RA, Selna MJ, Doll MC, Bothe AE, McKinley KE, et al. "ProvenCare(SM)" a provider-driven pay-for-performance program for acute episodic cardiac surgical core. Annals of Surgery. 2007;246(4):613-23.

190. Larsen DL, Cannon W, Towner S. Longitudinal assessment of a diabetes care management system in an integrated health network. J Manag.Care Pharm. 2003;9(6):552-8.

191. Levin-Scherz J, DeVita N, Timbie J. Impact of pay-for-performance contracts and network registry on diabetes and asthma HEDIS (R) measures in an integrated delivery network. Medical Care Research and Review. 2006;63(1):14S-28S.

192. Shortell SM, Zazzali JL, Burns LR, Alexander JA, Gillies RR, Budetti PP, et al. The role of market pressures, compensation incentives, and culture in physician organizations. Medical Care. 2001;39(7):I62-I78.

193. Weber V, Bloom F, Pierdon S, Wood C. Employing the electronic health record to improve diabetes care: A multifaceted intervention in an integrated delivery system. Journal of General Internal Medicine. 2008;23(4):379-82.

194. O'Malley AS, Pham HH, Reschovsky JD. Predictors of the growing influence of clinical practice guidelines. Journal of General Internal Medicine. 2007;22(6):742-8.

195. Grady KE, Lemkau JP, Lee NR, Caddell C. Enhancing mammography referral in primary care. Preventive Medicine. 1997;26(6):791-800.

196. An LC, Bluhm JH, Foldes SS, Alesci NL, Klatt CM, Center BA, et al. A randomized trial of a pay-for-performance program targeting clinician referral to a state tobacco quitline. Archives of Internal Medicine. 2008;168(18):1993-9.

197. Amundson G, Solberg LI, Reed M, Martini EM, Carlson R. Paying for quality improvement: compliance with tobacco cessation guidelines. Jt.Comm J Qual.Saf. 2003;29(2):59-65.

198. Li R, Simon J, Bodenheimer T, Gillies RR, Casalino L, Schmittdiel J, et al. Organizational factors affecting the adoption of diabetes care management process in physician organizations. Diabetes Care. 2004;27(10):2312-6.

199. McMenamin SB, Schmittdiel J, Halpin HA, Gillies R, Rundall TG, Shortell SA. Results from a national study. American Journal of Preventive Medicine. 2004;26(4):259-64.

200. Mehrotra A, Pearson SD, Coltin KL, Kleinman KP, Singer JA, Rabson B, et al. The response of physician groups to P4P incentives. Am.J Manag.Care. 2007;13(5):249-55.

201. Roski J, Jeddeloh R, An L, Lando H, Hannan P, Hall C, et al. The impact of financial incentives and a patient registry on preventive care quality: increasing provider adherence to evidence-based smoking cessation practice guidelines. Preventive Medicine. 2003;36(3):291-9.

202. Schmittdiel J, McMenamin SB, Halpin HA, Gillies RR, Bodenheimer T, Shortell SM, et al. The use of patient and physician reminders for preventive services: results from a National Study of Physician Organizations. Preventive Medicine. 2004;39(5):1000-6.

203. Williams TR, Raube K, Damberg CL, Mardon RE. Pay for performance: its influence on the use of IT in physician organizations. J Med.Pract.Manage. 2006;21(5):301-6.

204. Safran DG, Rogers WH, Tarlov AR, Inui T, Taira DA, Montgomery JE, et al. Are they related to primary care performance? Archives of Internal Medicine. 2000;160(1):69-76.

205. Chung RS, Chernicoff HO, Nakao KA, Nickel RC, Legorreta AP. A quality-driven physician compensation model: four-year follow-up study. J Healthc.Qual. 2003;25(6):31-7.

206. Gilmore AS, Zhao YX, Kang N, Ryskina KL, Legorreta AP, Taira DA, et al. Patient outcomes and evidence-based medicine in a preferred provider organization setting: A six-year evaluation of a physician pay-for-performance program. Health Services Research. 2007;42(6):2140-59.

207. Casalino L, Gillies RR, Shortell SM, Schmittdiel JA, Bodenheimer T, Robinson JC, et al. External incentives, information technology, and organized processes to improve health care quality for patients with chronic diseases. Jama-Journal of the American Medical Association. 2003;289(4):434-41.

208. Simon JS, Rundall TG, Shortell SM. Adoption of order entry with decision support for chronic care by physician organizations. Journal of the American Medical Informatics Association. 2007;14(4):432-9.

209. McMenamin SB, Schauffler HH, Shortell SM, Rundall TG, Gillies RR. Results from a National Study. Medical Care. 2003;41(12):1396-406.

210. LeBaron CW, Mercer JT, Massoudi MS, Dini E, Stevenson J, Fischer WM, et al. Changes in clinic vaccination coverage after institution of measurement and feedback in 4 states and 2 cities. Archives of Pediatrics & Adolescent Medicine. 1999;153(8):879-86.

211. Bhattacharyya T, Mehta P, Freiberg AA. Hospital characteristics associated with success in a pay-for-performance program in orthopaedic surgery. Journal of Bone and Joint Surgery-American Volume. 2008;90A(6):1240-3.

212. Reschovsky JD, Hadley J, Landon BE. Effects of compensation methods and physician group structure on physicians' perceived incentives to alter services to patients. Health Services Research. 2006;41(4):1200-20.

213. Cameron PA, Kennedy MP, McNeil JJ. The effects of bonus payments on emergency service performance in Victoria. Medical Journal of Australia. 1999;171(5):243-6.

214. Bond L, Davie G, Carlin JB, Lester R, Nolan T. Increases in vaccination coverage for children in child care, 1997 to 2000: an evaluation of the impact of government incentives and initiatives. Australian and New Zealand Journal of Public Health. 2002;26(1):58-64.

215. Martens JD, Werkhoven MJ, Severens JL, Winkens RAG. Effects of a behaviour independent financial incentive on prescribing behaviour of general practitioners. Journal of Evaluation in Clinical Practice. 2007;13(3):369-73.

216. Twardella D, Brenner H. Effects of practitioner education, practitioner payment and reimbursement of patients' drug costs on smoking cessation in primary care: a cluster randomised trial. Tobacco Control. 2007;16(1):15-21.

217. Cattaneo A, Borgnolo G, Simon G. Breastfeeding by objectives. European Journal of Public Health. 2001;11(4):397-401.

218. Gene-Badia J, Escaramis-Babiano G, Sans-Corrales M, Sampietro-Colom L, Aguado-Menguy F, Cabezas-Pena C, et al. Impact of economic incentives on quality of professional life and on end-user satisfaction in primary care. Health Policy. 2007;80(1):2-10.

219. Ettner SL, Thompson TJ, Stevens MR, Mangione CM, Kim C, Steers WN, et al. Are physician reimbursement strategies associated with processes of care and patient satisfaction for patients with diabetes in managed care? Health Services Research. 2006;41(4):1221-41.

220. Greenberg MR, Weinstock M, Fenimore DG, Sierzega GM. Emergency department tobacco cessation program: staff participation and intervention success among patients. J Am.Osteopath.Assoc. 2008;108(8):391-6.

221. Reiter KL, Nahra TA, Alexander JA, Wheeler JR. Hospital responses to pay-for-performance incentives. Health Serv.Manage.Res. 2006;19(2):123-34.

222. Beaulieu ND, Horrigan DR. Putting smart money to work for quality improvement. Health Services Research. 2005;40(5):1318-34.

223. Herrin J, Nicewander D, Ballard DJ. The effect of health care system administrator pay-for-performance on quality of care. Jt.Comm J Qual.Patient Saf. 2008;34(11):646-54.

224. Mason A, Walker S, Claxton K, Cookson R, Fenwick E, Sculpher M. The GMS Quality and Outcomes Framework: Are the Quality and Outcomes Framework (QOF) indicators a cost-effective use of NHS resources. 2008.

225. Curtin K, Beckman H, Pankow G, Milillo Y, Greene RA. Return on investment in pay for performance: A diabetes case study. Journal of Healthcare Management. 2006;51(6):365-74.

226. Fleetcroft R, Parekh S, Steel N, Swift L, Cookson R, Howe A. Potential population health gain of the Quality and Outcomes Framework. Report to the department of health 2008. 2008.

227. Fleetcroft R, Cookson R. Do the incentive payments in the new NHS contract for primary care reflect likely population health gains? J Health Serv.Res.Policy. 2006;11(1):27-31.

228. McElduff P, Lyratzopoulos G, Edwards R, Heller RF, Shekelle P, Roland M. Will changes in primary care improve health outcomes? Modelling the impact of financial incentives introduced to improve quality of care in the UK. Quality & Safety in Health Care. 2004;13(3):191-7.

229. Kahn CN, Ault T, Isenstein H, Potetz L, Van Gelder S. Snapshot of hospital quality reporting and pay-for-performance under Medicare. Health Affairs. 2006;25(1):148-62.

230. McLean G, Sutton M, Guthrie B. Deprivation and quality of primary care services: evidence for persistence of the inverse care law from the UK Quality and Outcomes Framework. Journal of Epidemiology and Community Health. 2006;60(11):917-22.

231. Armour BS, Pitts MM, Maclean R, Cangialose C, Kishel M, Imai H, et al. The effect of explicit financial incentives on physician behavior. Arch.Intern.Med. 2001;161(10):1261-6.

232. Giuffrida A, Gosden T, Forland F, Kristiansen IS, Sergison M, Leese B, et al. Target payments in primary care: effects on professional practice and health care outcomes. Cochrane.Database.Syst.Rev. 2000(3):CD000531.

233. Sabatino SA, Habarta N, Baron RC, Coates RJ, Rimer BK, Kerner J, et al. Systematic reviews of provider assessment and feedback and provider incentives. American Journal of Preventive Medicine. 2008;35(1):S67-S74.

234. Schatz M. Does pay-for-performance influence the quality of care? Curr.Opin.Allergy Clin.Immunol. 2008;8(3):213-21.

235. Victora CG, Vaughan JP, Barros FC, Silva AC, Tomasi E. Explaining trends in inequities: evidence from Brazilian child health studies. Lancet. 2000;356(9235):1093-8.

236. Department of Health. Quality and Outcomes Framework. Available from: http://www.dh.gov.uk/en/Healthcare/Primarycare/Primarycarecontracting/QOF/index.htm

237. Integrated Healthcare Association (IHA). P4P 2009 Measurement Year / 2010 Reporting and Payment Year... Available from: http://www.iha.org/p4py7.htm

238. Medicare Australia. Practice Incentives Program (PIP). Available from: http://www.medicareaustralia.gov.au/provider/incentives/pip/index.jsp

239. Kaiser Permanente. Get to know Kaiser Permanente. Available from: https://members.kaiserpermanente.org/kpweb/aboutus.do

240. Corens D. Health system review: Belgium. Health systems in transition. 2007.

241. Borgermans L, Goderis G, Van Den Broeke C, Buelens K, Reynders C, Heyrman J, et al. Beleidsconclusies vanuit het dubbelproject Leuven Aalst. Rijksinstituut voor Ziekte- en Invaliditeitsverzekering (RIZIV); 2007.

242. Goderis G, Borgermans L, Heyrman J, Broeke CV, Grol R, Boland B, et al. Type 2 diabetes in primary care in belgium: need for structured shared care. Exp Clin Endocrinol Diabetes. 2009;117(8):367-72.

243. Devriese S, Lambert M-L, Eyssen M, Van De Sande S, Poelmans J, Van Brabandt H, et al. Prospectief bepalen van de honoraria van ziekenhuisartsen op basis van klinische paden en guidelines: makkelijker gezegd dan gedaan. Report. Brussel: Federaal Kenniscentrum voor de Gezondheidszorg (KCE); 2005. KCE reports 18A (D/2005/10.273/19) Available from: http://kce.fgov.be/Download.aspx?ID=271

244. Panella M, Brambilla R, Marchisio S, Di Stanislao F. Reducing stroke in-hospital mortality: organized care is a complex intervention. Stroke. 2008;39(11):e186.

245. Vanhaecht K, De Witte K, Bellemans J, Diya L, Lesaffre E, Sermeus W. Does the organization of care processes affect outcomes in patients undergoing total joint replacement. J Eval Clin Pract. 2008;In Print.

246. Heyrman J, Lemye R, Moens M, Van den Oever R. Verslag van de resultaten van de schriftelijke enquête onder de LOK's ter evaluatie van de accreditering als systeem. Rijksinstituut voor Ziekte- en Invaliditeitsverzekering (RIZIV). Available from: http://www.riziv.fgov.be/care/nl/doctors/accreditation/news/pdf/repport.pdf

247. Heyrman J, Lemye R, Moens M, Van den Oever R. Rapport des résultats de l'enquête écrite menée auprès des GLEM et relative a l'évaluation de l'accréditation en tant que système Institut national d'assurance maladie-invalidité (INAMI). Available from: http://inami.fgov.be/care/fr/doctors/accreditation/news/pdf/rapport.pdf

248. Annemans L, Closon J-P, Closon M-C, Heymans I, Lagasse R, Mendes da Costa E, et al. Comparaison du coût et de la qualité de deux systèmes de financement des soins de première ligne en Belgique. Health Services Research (HSR). Bruxelles: Centre fédéral d'expertise des soins de santé (KCE); 2008. KCE reports 85B (D/2008/10.273/50)

249. Remmen R, Seuntjens L, Pestiaux D, Leysen P, Knops K, Lafontaine J-B, et al. Quality development in general practice in Belgium: status quo or quo vadis ? Good Clinical Practice (GCP). Bruxelles: Belgian Health Care Knowledge Centre (KCE); 2008. KCE Reports 76C (D/2008/10.273/49)

250. Engels Y, Campbell S, Dautzenberg M, van den Hombergh P, Brinkmann H, Szecsenyi J, et al. Developing a framework of, and quality indicators for, general practice management in Europe. Fam Pract. 2005;22(2):215-22.

251. TOPAS-Europe. European Practice Assessment - EPA: Easy to use and scientifically developed quality management for general practice. 2008. Available from: http://www.topaseurope.eu/files/EPA-Information-Paper-English-vs11_0.pdf

252. Remmen R, Seuntjens L, Paulus D, Pestiaux D. Kwaliteitszorg in de huisartsgeneeskunde. Inleiding en een snapshot van basisbegrippen. Huisarts Nu. 2008;37:547-53.

253. Remmen R, Seuntjens L, Knops K, Leysen P, Paulus D, Van den Bruel A, et al. Kwaliteitsverbetering in Belgische huisartspraktijken; kan een Europees instrument de missing link opvullen? Huisarts Nu. 2008;37:555-9.

254. Remmen R, Seuntjens L, Knops K, Leysen P, Pestiaux D, Philips H, et al. Mesurer la qualité de l'organisation des pratiques: une utopie en Belgique? La Revue de la Médecine Générale 2009;260:70-3.

255. Remmen R, Seuntjens L, Paulus RA, Pestiaux D, Knops K, Van den Bruel A. Can we import quality tools? A feasibility study of European practice assessment in a country with less organised general practice. BMC Health Services Research. 2009;9(1):183.

256. Van Linden A, Heymans I, Mambourg F, Leys M, De Prins L, Dieleman P, et al. Feedback: onderzoek naar de impact en barrières bij implementatie – Onderzoeksrapport: deel 1. Report. Brussel: Federaal Kenniscentrum voor de Gezondheidszorg (KCE); 2005 Avril. KCE reports 9A (D/2005/10.273/01) Available from: http://kce.fgov.be/Download.aspx?ID=259

257. Heymans I, Van Linden A, Mambourg F, Leys M. Feedback : onderzoek naar de impact en barrières bij implementatie – Onderzoeksrapport : deel II. Brussel: Federaal Kenniscentrum voor de Gezondheidszorg (KCE); 2006. KCE reports 32A (D2006/10.273/19)

258. de Walcque C, Seuntjens B, Vermeyen K, Peeters G, Vinck I. Comparative study of hospital accreditation programs in Europe. Health Services Research (HSR). Brussels: Belgian Health Care Knowledge Centre (KCE); 2008. KCE Reports 70C (D/2008/10.273/03)

259. Vlayen J, Van De Water G, Camberlin C, Paulus D, Leys M, Ramaekers D, et al. Klinische kwaliteitsindicatoren. Objective Elements - Communication (OEC). Brussel: Federaal Kenniscentrum voor de gezondheidszorg (KCE); 2006. KCE reports 41A (D/2006/10.273/43)

260. Agency for Healthcare Research and Quality (AHRQ). Hospital Survey on Patient Safety Culture. Available from: http://www.ahrq.gov/qual/patientsafetyculture/hospsurvindex.htm

261. Van De Sande S, KCE, IMA, UCL, UZ-GENT, TCT RESEARCH TEAM. Evaluatie van de referentiebedragen. Report. Brussel: Federaal Kenniscentrum voor de Gezondheidszorg (KCE); 2005. KCE reports 17A (D/2005/10.273/17) Available from: http://kce.fgov.be/Download.aspx?ID=270

262. World Health Organization. The World Health Report 2008 - primary Health Care (Now More Than Ever). Available from: http://www.who.int/whr/2008/en/index.html

This page is left intentionally blank.

Legal depot : D/2009/10.273/52

KCE reports

33 Effects and costs of pneumococcal conjugate vaccination of Belgian children. D/2006/10.273/54.
34 Trastuzumab in Early Stage Breast Cancer. D/2006/10.273/25.
36 Pharmacological and surgical treatment of obesity. Residential care for severely obese children in Belgium. D/2006/10.273/30.
37 Magnetic Resonance Imaging. D/2006/10.273/34.
38 Cervical Cancer Screening and Human Papillomavirus (HPV) Testing D/2006/10.273/37.
40 Functional status of the patient: a potential tool for the reimbursement of physiotherapy in Belgium? D/2006/10.273/53.
47 Medication use in rest and nursing homes in Belgium. D/2006/10.273/70.
48 Chronic low back pain. D/2006/10.273.71.
49 Antiviral agents in seasonal and pandemic influenza. Literature study and development of practice guidelines. D/2006/10.273/67.
54 Cost-effectiveness analysis of rotavirus vaccination of Belgian infants D/2007/10.273/11.
59 Laboratory tests in general practice D/2007/10.273/26.
60 Pulmonary Function Tests in Adults D/2007/10.273/29.
64 HPV Vaccination for the Prevention of Cervical Cancer in Belgium: Health Technology Assessment. D/2007/10.273/43.
65 Organisation and financing of genetic testing in Belgium. D/2007/10.273/46.
66. Health Technology Assessment: Drug-Eluting Stents in Belgium. D/2007/10.273/49.
70. Comparative study of hospital accreditation programs in Europe. D/2008/10.273/03
71. Guidance for the use of ophthalmic tests in clinical practice. D/200810.273/06.
72. Physician workforce supply in Belgium. Current situation and challenges. D/2008/10.273/09.
74 Hyperbaric Oxygen Therapy: a Rapid Assessment. D/2008/10.273/15.
76. Quality improvement in general practice in Belgium: status quo or quo vadis? D/2008/10.273/20
82. 64-Slice computed tomography imaging of coronary arteries in patients suspected for coronary artery disease. D/2008/10.273/42
83. International comparison of reimbursement principles and legal aspects of plastic surgery. D/200810.273/45
87. Consumption of physiotherapy and physical and rehabilitation medicine in Belgium. D/2008/10.273/56
90. Making general practice attractive: encouraging GP attraction and retention D/2008/10.273/66.
91 Hearing aids in Belgium: health technology assessment. D/2008/10.273/69.
92. Nosocomial Infections in Belgium, part I: national prevalence study. D/2008/10.273/72.
93. Detection of adverse events in administrative databases. D/2008/10.273/75.
95. Percutaneous heart valve implantation in congenital and degenerative valve disease. A rapid Health Technology Assessment. D/2008/10.273/81
100. Threshold values for cost-effectiveness in health care. D/2008/10.273/96
102. Nosocomial Infections in Belgium: Part II, Impact on Mortality and Costs. D/2009/10.273/03
103 Mental health care reforms: evaluation research of 'therapeutic projects' - first intermediate report. D/2009/10.273/06.
104. Robot-assisted surgery: health technology assessment. D/2009/10.273/09
108. Tiotropium in the Treatment of Chronic Obstructive Pulmonary Disease: Health Technology Assessment. D/2009/10.273/20
109. The value of EEG and evoked potentials in clinical practice. D/2009/10.273/23
111. Pharmaceutical and non-pharmaceutical interventions for Alzheimer's Disease, a rapid assessment. D/2009/10.273/29
112. Policies for Orphan Diseases and Orphan Drugs. D/2009/10.273/32.
113. The volume of surgical interventions and its impact on the outcome: feasibility study based on Belgian data
114. Endobronchial valves in the treatment of severe pulmonary emphysema. A rapid Health Technology Assessment. D/2009/10.273/39
115. Organisation of palliative care in Belgium. D/2009/10.273/42
116. Interspinous implants and pedicle screws for dynamic stabilization of lumbar spine: Rapid assessment. D/2009/10.273/46
117. Use of point-of care devices in patients with oral anticoagulation: a Health Technology Assessment. D/2009/10.273/49.

118. Advantages, disadvantages and feasibility of the introduction of 'Pay for Quality' programmes in Belgium. D/2009/10.273/52.

This list only includes those KCE reports for which a full English version is available. However, all KCE reports are available with a French or Dutch executive summary and often contain a scientific summary in English.