# Getting the incentives right

## The design of value-based consumer and provider payments in health care

Daniëlle Cattel

# Getting the incentives right

## The design of value-based consumer and provider payments in health care

**Daniëlle Cattel**

ISBN 9789490420642
Lay-out and printing by Optima Grafische Communicatie (www.ogc.nl)

**Getting the incentives right**

The design of value-based consumer and provider payments in health care

**Op weg naar betere prikkels**

De vormgeving van waardegedreven betalingen door zorggebruikers
en beloning van zorgaanbieders

Thesis

to obtain the degree of Doctor from the
Erasmus University Rotterdam
by command of the
rector magnificus

Prof. dr. A.L. Bredenoord

and in accordance with the decision of the Doctorate Board.

The public defence shall be held on

December 17, 2021 at 13.00 hrs

by

**Daniëlle Cattel**

born in Gouda, the Netherlands

**Erasmus University Rotterdam**

**PROMOTIECOMMISSIE**

**Promotor**
Prof. dr. F.T. Schut

**Overige leden**
Prof. dr. ir. C.T.B. Ahaus
Prof. dr. C.A. Baan
Prof. dr. M.P.M.H. Rutten-van Mölken

**Copromotor**
Dr. F. Eijkenaar

*Ik heb het nog nooit gedaan, dus ik denk dat ik het wel kan*
– Pippi Langkous –

**CONTENTS**

## LIST OF PUBLICATIONS AND SUBMISSIONS

Chapters 2, 3, 4, 5, and 6 are based upon the following articles.

**Chapter 2**
Cattel, D., R.C. van Kleef & R.C.J.A. van Vliet. 2016. 'A method to simulate incentives for cost containment under various cost sharing designs: An application to a first-euro deductible and a doughnut hole.' *European Journal of Health Economics* 18: 987–1000.

**Chapter 3**
Cattel, D., F. Eijkenaar & F.T. Schut. 2020. 'Value-based provider payment: Towards a theoretically preferred design.' *Health Economics, Policy and Law* 15(1): 94-112.

**Chapter 4**
Cattel, D. & F. Eijkenaar. 2020. 'Value-based provider payment initiatives combining global payments with explicit quality incentives: A systematic review.' *Medical Care Research and Review* 77(6): 511-537.

**Chapter 5**
Cattel, D. & F. Eijkenaar. 2020. 'How to manage financial risk for capitated primary care providers? The impact of care package, risk adjustment, risk sharing, and patient panel size.' *Preparing for submission.*

**Chapter 6**
Cattel, D., F. Eijkenaar & R.C. van Kleef. 2021. 'Getting the incentives right: Simulating the effects of residual-based risk-sharing for primary care providers under global payment.' *Preparing for submission.*

# Chapter 1

## General introduction

## 1. BACKGROUND

Despite substantial contributions of healthcare systems to life expectancy and quality of life, in many countries it is widely recognized that there remains considerable room for improvements in the quality and efficiency of health care. In particular, across OECD countries, a significant share of spending on healthcare is wasteful (OECD 2017). In addition, adherence to clinical guidelines is often low, current healthcare systems are ill-equipped to provide well-coordinated, integrated care, and focus is on treating health problems instead of preventing them (Pronovost 2013; Tsiachristas 2015). Therefore, realizing more 'value' in health care has increasingly become a focal point in health policy in the last decade, with value being defined in many ways. In the past, value has been narrowly defined (see, for example, Porter 2010), while more recently, good attempts have been made to provide more comprehensive descriptions of value (see, for example, European Commission 2019). In this dissertation, value is considered a multifaceted concept, comprising not only quality of care at the lowest possible costs, but also efficient coordination of care, cost-effective innovation, and prevention (IOM 2001; Berwick et al. 2008; Porter 2009; Porter 2010; Conrad 2015; Eijkenaar & Schut 2015; European Commission 2019).

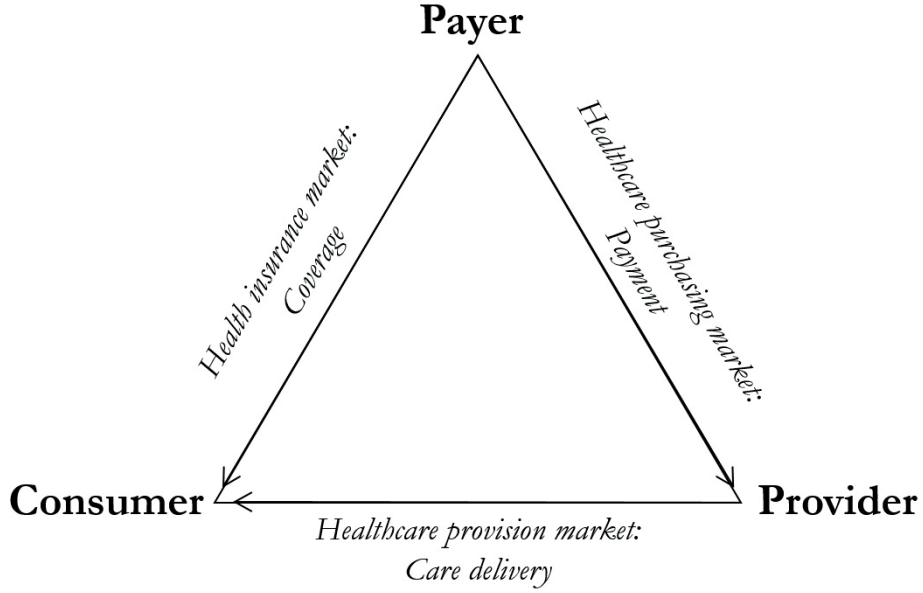
Improving value requires a thorough understanding of the main drivers of suboptimal value. A wide range of evidence suggests that suboptimal value is (at least partly) caused by perverse financial incentives in healthcare influencing behavior (McGuire 2000; McGuire 2011; Evans 1974; Newhouse 1993; Pauly 1968; Gaynor et al. 2004). Perverse incentives exist in all three healthcare markets: (1) the healthcare provision market where the consumer (here: the patient) interacts with the provider and care is delivered, (2) the health insurance market where the consumer (here: the insured) purchases health insurance from the payer (i.e., government or health insurer) in exchange for coverage, and (3) the healthcare purchasing market where the provider is contracted and paid for care delivery by the payer.[1] Figure 1.1 graphically displays these interactions between the consumer, provider, and payer in the three markets in health care. [2] In this dissertation the focus is on *incentives for consumers* who buy insurance coverage on the health insurance market and on *incentives for providers* who are reimbursed by the payer on the healthcare purchasing market for delivering care services on the healthcare provision market.

---

1 By provider we mean individual health practitioners as well as organizations, including hospitals, post-acute care providers, physicians, and other practitioners. By health professional we specifically mean individual health practitioners (typically physicians).

2 Note that the three markets are not present in each healthcare system. The healthcare purchasing market is only relevant for countries with social health insurance (i.e., a 'Bismarck system') and/or private health insurance (e.g., the United States of America; US), and does not exist in countries with a classical National Health Service (e.g., the United Kingdom; UK). In this dissertation, however, the focus is on a contract model in which providers and/or payers may or may not compete on price and in which the three markets can be distinguished (Van de Ven et al. 1994) although findings are also relevant for non-competitive healthcare systems.

**Figure 1.1.** Interactions between the consumer, provider, and payer in the three markets in health care




Health insurance is a popular instrument to establish universal access to otherwise unaffordable care and to realize welfare gains in risk-averse societies (Pauly 1968; Nyman 1999; Rosenthal 2004). An important drawback of health insurance, however, is that it may result in moral hazard, which means that individuals use more or more expensive medical services than they would without insurance because they do not bear the complete marginal costs of care (Arrow 1963; Pauly 1968; Zweifel & Manning 2000). This is a problem because of scarce resources and because it may impose a welfare loss on society in the case of excess utilization of health care (Pauly 1968; Feldstein 1973). Empirical research has shown that moral hazard is not merely a theoretical concern (see, for example, Newhouse 1993; Van Vliet 2004; Bakker 1997; Baicker et al. 2013).

In many countries' health insurance markets, out-of-pocket payments by consumers (often referred to as cost sharing) have been implemented to increase consumers' perceived price of health care and thereby reduce moral hazard (Qingyue et al. 2011). These direct payments by consumers ideally reduce inefficient use of health care without putting consumers at too much financial risk. Three types of payments where consumers pay part of the bill are co-insurance, co-payments, and deductibles. With co-insurance, consumers pay a percentage of total healthcare spending out-of-pocket, for example 10% of a €150 bill. With co-payments consumers are required to pay a certain amount per service out-of-pocket, for example €5 per specific prescription of medication. A deductible implies that, up to a certain amount, consumers pay 100% of their healthcare spending out-of-pocket after which the insurer covers additional expenses.

Each conventional cost-sharing method has its own disadvantages. With co-insurance and co-payments, for example, out-of-pocket spending for individuals with a relatively poor health

and high utilization of healthcare services accumulate and total out-of-pocket spending can be substantial. An important drawback of co-payments is that consumers are not sensitive for the price differences between providers and in case of a deductible, that consumers have no continuous incentive to act cost-consciously and reduce inefficient use of health care. Up to the deductible amount, consumers pay their healthcare spending out-of-pocket, stimulating consumers to behave as completely uninsured. However, after reaching the deductible amount, the insurer takes over and fully reimburses excess healthcare spending which stimulates consumers to behave as being completely insured. In sum, incentives for cost-conscious behavior for consumers emanating from the conventional cost-sharing methods are suboptimal.

Financial incentives for providers are stemming from the specific payment models through which they are reimbursed. Based on an extensive review of the literature, McGuire (2000; 2011) concludes that providers can influence quantity in practice and sometimes do so in their own interest. An important concern with predominant provider payment models is that these models are ill-aligned with value. In practice, payment is often positively related to the number of care services (known as fee-for-service (FFS)). FFS stimulates providers to increase the quantity of services because this results in higher income. Providing additional services can be in the best interest of the patient, for example when a physician persuades a non-compliant patient with diabetes to use specific medication for controlling the disease. By improving medication adherence, the provider reduces the health risk for the patient by shifting the demand curve outwards. In this case, the demand curve of a fully informed patient equals the demand curve induced by the provider. However, providing additional services might also harm patients, for example when a physician persuades the patient to demand extra (not cost-effective or unnecessary) tests. In this case, the provider's income rises, without any benefit for the patient and possibly even clinical risks for the patient. Here, the demand curve of the fully informed patient deviates from the demand curve constructed by the provider.

Two other provider payment models that are often used in practice are capitation (in which providers receive a periodical fixed amount per patient) and salary (in which providers receive a fixed amount irrespective of the number of patients). In contrast to FFS, these models have no link with the volume of care at all. Because there is no direct relation between effort and payment, capitation and salary might stimulate underprovision of care. In addition, providers might be inclined to select favorable (i.e., low-cost and 'easy' or low-effort) patients and might skimp on quality. An additional, important concern of predominant provider payment models in general is that none of the models reward well-coordinated, high-quality care. Although pursu-

ing integrated, high-quality care is in the best interest of patients, providers are financially not encouraged to 'walk the extra mile' required to significantly improve health care.[3]

## 2. CENTRAL AIM AND RESEARCH QUESTION

Financial incentives have been convincingly shown to influence consumer and provider behavior and predominant payment systems are ill-aligned with value (section 1). Therefore, stakeholders have been exploring alternative payment models for consumers and providers containing financial incentives facilitating value. These innovative payment models pursuing value will henceforth be referred to as value-based payment (VBP) incentives (Bazemore et al. 2018; Struijs et al. 2019; APMF FPT Work Group 2016; Chernew et al. 2020). To date, however, little is known about what VBP incentives for consumers and providers should look like and what this would entail in practice. The theoretical basis of VBP incentive design is fragmented and the relationship between what a healthcare system ideally pursues in terms of value and what is required in terms of VBP incentive design to achieve this remains poorly understood. In addition, little is known about the effectiveness of alternative payment incentives in improving value. Against this background, the central research question of this dissertation is:

**How can financial incentives in consumer and provider payment be designed to facilitate value in health care?**

This dissertation aims to provide insights into key issues in the design of VBP incentives for consumers and providers, and in associated tradeoffs and incentive effects. In doing so, we contribute to the body of knowledge concerning smarter choices in payment system design. Insights may help stakeholders with (re)designing existing and future consumer and provider payment models. Results are relevant for all countries seeking to increase value in health care by reforming financial incentives in consumer and provider payment systems.

In the next section of this general introduction, the problem of undesired consumer and provider behavior will be positioned within the theory of agency. In section 4 and 5 of this chapter a theoretical framework on financial incentives for consumers and providers, respectively, will

3 Note that the extent to which providers show behavior that does not stimulate value, depends on the utility function of a provider. If providers pursue an optimal combination of net income and patient health, the utility derived from a higher income is has to be balanced against the disutility of demand inducement and other perverse provider behavior that is not in the best interest of the patient (Evans 1974; McGuire & Pauly 1991). Social norms, medical ethics, altruistic preferences, and leisure time are examples of factors that are relevant for the doctor's utility function too (Robinson 2001a).

be provided and the specific research questions in this dissertation will be specified. In the last section, the structure of this dissertation will be presented.

## 3. AGENCY PROBLEMS IN HEALTH CARE

In health care, information is not equally spread among the consumer, payer and provider. In the relationship between the consumer and payer, the former generally is the relatively well-informed party and the latter the relatively ill-informed party, when it comes to the consumer's health status, risk, and behavior (Harris & Raviv 1978; Arrow 1986). In the relationship between the provider and the payer and the provider and the patient, the provider as the clinical professional is the relatively well-informed party (Vermaas 2006). Agency theory, as part of contract theory, helps to understand how information asymmetry may result in agency problems and provides tools for dealing with those problems.

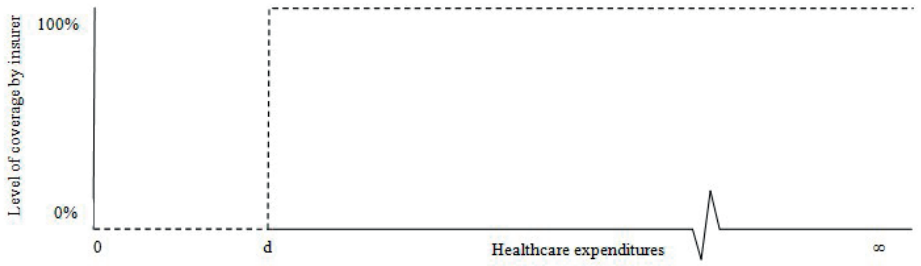
Agency theory studies the contractual relationship between two parties: the agent and the principal. The agent is the relatively well-informed party and the principal the relatively ill-informed party. According to this theory, information asymmetry between agents and principals is not a problem if interests are aligned (Laffont & Martimort 2002). In case of conflicting interests, however, agents might exploit their information surplus for their own (financial) benefit (Jensen & Meckling 1976; Richardson 1981). Two main types of problems may evolve (Arrow 1986). The first problem is called the adverse selection problem and may occur before the contract is concluded (Eisenhardt 1989). Here, the actions of agents are observable, but the information used by agents to take these actions is not. As a result, principals do not know whether the agents' actions suit the principals' interests best. Adverse selection is beyond the scope of this dissertation. The second problem is called the moral hazard problem and may appear after the contract is concluded. In this case, principals cannot monitor the agents' actions and only have some information about the outcomes of the actions. As a result, the principals do not know whether these outcomes are optimal given the agents' knowledge. Supplier-induced moral hazard is an example of the moral hazard problem and implies that a provider induces demand in the knowledge that the patient has insurance and resulting costs are covered by the insurance policy (Vermaas 2006). An example of supplier-induced moral hazard is that of a physician prescribing expensive drugs covered by insurance instead of inexpensive drugs not covered by insurance. To address moral hazard, an important strategy offered by agency theory entails 'controlling' agents by means of contracts – including financial incentives – to align agents' interests with those of principals' (Vermaas 2006). In this dissertation we focus on a specific type of the contractual agreement, namely the design of VBP incentives for consumers and providers.

## 4. VALUE-BASED PAYMENT INCENTIVES FOR CONSUMERS

This dissertation focuses on a specific type of consumer cost sharing: the mandatory deductible. The reason is that mandatory deductibles are widespread, administrative costs are relatively low, and that different choices in the deductible design have different consequences for consumers. In this dissertation the impact on incentives for cost-conscious behavior of three specific designs for the mandatory deductible is studied: (1) a first-euro deductible, (2) a shifted deductible (also known as a 'doughnut hole' deductible) with a uniform starting point, and (3) a shifted deductible with a risk-adjusted starting point.

A *first-euro deductible* (Figure 1.2) is the most commonly applied deductible and means that consumers pay the first €d out-of-pocket before the payer takes over and reimburses all excess medical spending covered by the benefit package. In Figure 1.2 spending in the interval [0, d] is the responsibility of the consumer while spending in interval [d, ∞] is the responsibility of the payer. An example of a first-euro deductible can be found in the Dutch basic health insurance system where individuals have a mandatory deductible of €385 per person in 2021.

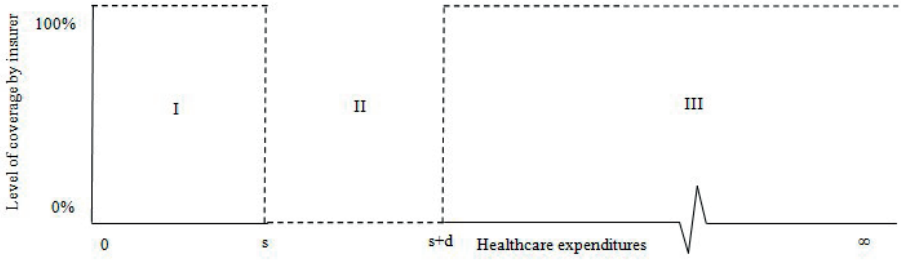
**Figure 1.2.** Insurance under a first-euro deductible with range [0, d]




A first-euro deductible has several drawbacks in terms of efficiency and equity (Van Kleef et al. 2009; Van Kleef et al. 2010; Van Kleef et al. 2011). First, for the high-risk individuals, for instance the chronically ill or elderly, the first-euro deductible is not effective in reducing moral hazard. These individuals know ex ante that their annual healthcare spending will exceed the deductible range. As a result, they are not price sensitive and lack any incentive to contain costs because cost-conscious behavior will not prevent them from having to pay the maximum out-ot-pocket payment at the end of the contract period. In addition, high-risk individuals do not have any incentive to opt for a voluntary deductible on top of the mandatory deductible since they are not likely to benefit financially from this (Van de Ven & Schut 2010). Finally, under a first-euro deductible, high-risk individuals on average pay more out-of-pocket than low-risk individuals. If high-risk individuals do not receive sufficient compensation, this can be seen as a decrease in risk solidarity compared to a situation without consumer cost sharing.

A possible remedy to reduce the drawbacks of a first-euro deductible in terms of efficiency and equity is a shifted deductible (Van Kleef et al. 2009). A shifted deductible is a deductible that starts at a higher level of healthcare spending than €0. Thus, the consumer experiences a coverage gap that begins at a predefined level of healthcare spending. Figure 1.3 shows that full coverage is provided for spending ranging from €0 to €s (interval I [0, s]). Then, consumers experience a gap in coverage – also labeled as the doughnut hole – as healthcare spending from €s until €s+d must be paid out-of-pocket (interval II [s, s+d]). Full coverage is again provided by the payer if healthcare spending exceeds point s+d (interval III [s+d, ∞]). Under a shifted deductible, the probability of having maximum out-of-pocket spending may reduce, which in turn may lead to increased price sensitivity (Van Kleef et al. 2009).

**Figure 1.3.** Insurance under a shifted deductible with range [s, s+d]




An interesting question is where to locate starting point *s*. One possibility is to use a uniform shifted starting point (i.e., *a shifted deductible with a uniform starting point*). The doughnut hole is fixed for all individuals and set, for example, at the mean actual spending in the population in the previous year. An example of such a deductible design can be found in the Medicare Part D coverage system that was implemented in 2006 in the US. In theory, however, a shifted deductible with a uniform starting point does not provide optimal incentives either, as incentives for cost-conscious behavior are weak for low-risk individuals with low expected spending. The reason is that for these individuals, the probability of reaching the starting point of the doughnut hole and paying the full deductible amount concentrates near 0. Because there is little uncertainty about their out-of-pocket spending, incentives for cost-conscious behavior are weak (Van Kleef et al. 2009).

To overcome the problems of both a first-euro deductible and a shifted deductible with a uniform starting point, *a shifted deductible with a risk-adjusted starting point* has been proposed (Van Kleef et al. 2009). Under this specific design, the starting point is dependent on individuals' risk of spending, providing everyone with appropriate financial incentives by correcting for heterogeneity in terms of individuals' ex-ante health status (Zhang et al. 2009; Roblin & Maciejewski 2011). Specific individual-level risk characteristics such as demographics, chronic conditions, and prior healthcare utilization can be used to predict healthcare spending and determining the

starting point (Van Kleef et al. 2009). Under a shifted deductible with a risk-adjusted starting point, high-risk and low-risk individuals have, on average, the same out-of-pocket spending. A shifted deductible with a risk-adjusted starting point has not (yet) been implemented in practice.

In theory, a shifted deductible with a risk-adjusted starting point results in more value than a first-euro deductible and a shifted deductible with a uniform starting point because this design of the deductible provides (1) stronger incentives for cost-conscious behavior and thus less risk of moral hazard since the perceived price of health care in the population is higher, (2) a higher probability that high-risk insured opt for a voluntary deductible and (3) more risk solidarity between high-risk and low-risk individuals. On the other hand, transaction costs may be high due to, for example, higher information costs or administration costs and transparency may decrease as everyone has an individualized starting point according to their risk-characteristics, making it more difficult for consumers to compare insurance policies.

The theoretical basis of the first-euro deductible and shifted deductible in general is quite well established in the literature. However, the relative effects of a first-euro deductible, a shifted deductible with a uniform starting point, and a shifted deductible with a risk-adjusted starting point on incentives for cost-conscious behavior will mostly be absent. This dissertation aims to reduce this knowledge gap. Against this background the first research question of this dissertation is:

**Q1: How can incentives for cost-conscious behavior under various deductible designs be compared?**

A simulation model is developed to approximate the relative effects of different deductible designs on consumers' incentives for cost-consciousness and compare these incentives under various deductible designs. In addition, we empirically illustrate this simulation model for a first-euro deductible and a shifted deductible with various starting points. Results are presented for the total population and separately for low-risk and high-risk individuals and can be used by stakeholders to underpin decisions on the design of effective consumer cost sharing in health insurance.

## 5. VALUE-BASED PAYMENT INCENTIVES FOR PROVIDERS

As briefly discussed in section 1, predominant provider payment models discourage the provision of efficient and well-coordinated care that is of high quality. Below, the incentives for providers generated by five payment models that are commonly used in practice are discussed in more depth (Miller 2009).

### 5.1 Payment per item-of-service (FFS)

A widespread provider payment model in healthcare systems worldwide is payment per-item-of-service. Under this model, which is frequently referred to as fee-for-service (FFS), individual providers are paid a predetermined amount for each discrete service, like an office visit or diagnostic test. The most important drawback of this payment model is that volume is rewarded, which provides incentives for overprovision. In addition, FFS may trigger providers to incorrectly classify patients in treatment categories with higher fees (i.e., upcoding). Furthermore, because the focus is on remunerating individual providers delivering single care activities, no incentives for coordination and collaboration among providers exist, resulting in fragmented care. This is particularly problematic for the increasing number of patients with multiple health problems, who would especially benefit from an integrated care approach. Finally, as preventing health problems will lead to less demand and a decrease in provider's income, incentives for health promotion and prevention are weak. On the other hand, under the assumption that marginal fees exceed marginal costs, providers have no incentives to withhold patients from necessary and good quality care and high productivity is rewarded. In addition, there is an incentive to pursue high levels of patient satisfaction, because satisfied patients are more likely to return.

### 5.2 Payment per case (case rate)

Under this payment model providers receive a single payment for all the services needed by a patient during an episode of care, such as pregnancy and delivery or a heart attack. No matter if the patient has a hospital stay of one or ten days or has five of fifty tests, one set price is paid. This single payment is commonly referred to as a case rate. A payment per case is broader than a payment per item-of-service. In case of a payment per case, providers are financially accountable for the difference between the payment and actual spending during the episode of care. As a result, the provision of unnecessary (expensive) care services per episode of care is discouraged. To contain costs, providers might, however, also be inclined to behave strategically and select financially attractive patients, shift costs to other providers, or skimp on quality. If the services of multiple providers are covered by payment per case, coordination of multiple providers is encouraged. There is, however, no incentive for providers to orchestrate the whole care process because multiple payments for episodes of care and conditions might apply. Another disadvantage of a payment per case is that it still is a volume-based payment model that might stimulate a 'more-is-better culture' and discourages primary prevention (i.e., the preventing episodes of care or conditions from occurring).

### 5.3 Payment per condition (DRG)

Under this payment model providers receive a single payment for a coherent set of care activities (usually hospital services) associated with a specific condition. A payment per condition is broader than a payment per case. A well-known example of a payment per condition is the diagnosis-related group (DRG) payment system in which payments to the hospital (excluding physician-fees) are bundled. The Dutch equivalent of the DRG system is the diagnosis-treatment combination (in Dutch abbreviated as DBC). Like under a payment per case, providers are financially accountable for the difference between the payment and actual spending. On the one hand, this discourages the provision of unnecessary (expensive) care services per condition and provides incentives to control the number of unnecessary episodes of care per condition. On the other hand, incentives for perverse provider behavior such as risk selection and quality skimping might evolve. Incentives for well-coordinated care and cooperation between providers are strong for those services covered by the payment. However, care still is fragmented for patients with multimorbidity since multiple payments for conditions might apply. Finally, there is no incentive to prevent conditions from occurring.

### 5.4 Payment per person (global payment, capitation, or population-based payment)

Another widespread payment model, especially in primary care sectors in many European countries, is a periodic payment per person. Under this payment model, a provider receives a prospectively determined, fixed amount for the provision of a specified care package for each person enrolled with the healthcare provider during the relevant period. A payment per person is broader than a payment per condition. Under a payment per person, the provider receives the payment, irrespective of whether the individual uses healthcare services. Again, the provider is financially responsible for the difference between the payment and actual spending, providing incentives for cost control but also for strategic provider behavior. Unlike a payment per item-of-service, case, or condition, a payment per person stimulates primary prevention and health maintenance because a healthier population is financially rewarding. Assuming the payment per person applies to one provider type (as generally is the case in practice), this payment does not stimulate well-coordinated care across the continuum of care.

### 5.5 Payment per period (salary or budget)

Under a payment per period, providers receive a fixed, periodical lump sum (salary or budget) for providing a set of predefined care services. In contrast with a payment per person, the provider is not accountable for a specific population. In general, this payment model discourages high productivity and may result in waiting lists. In addition, efforts to increase quality of care or boost innovation are not rewarded. An advantage is that administrative costs can be relatively low.

An important conclusion of the above analysis is that despite several advantages, all common payment models have significant drawbacks. None of the payment models are optimally aligned

with value. Therefore, worldwide, stakeholders are exploring alternative payment strategies to help steering healthcare systems towards value. Over the past decade, there has been much experimentation with various types of VBP, especially in the US. A prominent example of an alternative payment model is pay-for-performance (P4P). Under this model, providers receive explicit financial incentives for performing well on specific, measurable aspects of value, often related to quality. Examples of P4P-initiatives from practice are the Hospital Value-Based Purchasing Program in the US and the Quality and Outcomes Framework in the UK. Another type of VBP is bundled payment. Bundled payments are a predetermined reimbursement for services related to a condition or procedure over a defined period (CMS 2020). Payments per condition (e.g., DRGs) are essentially bundled payments for hospital services categorized by diagnosis, but under recent bundled payments a more comprehensive care package is covered by the payment (IBM 2017). In case of a patient suffering from severe arthritis requiring a hip replacement, for example, all charges associated with an inpatient stay related to the hip replacement from the time of admission to discharge are covered under a payment per condition, whereas under a bundled payment, also physicians fees, the costs of rehabilitation care, and of treatment of possible complications would be included in the bundle and covered by the payment. Bundled payment rewards multidisciplinary cooperation among multidisciplinary providers, sometimes even from different organizations and settings. Examples from practice are the Bundled Payment for Care Improvement Initiative and the Acute Care Episode Demonstration, both implemented in US Medicare. The various payment options in the public and private Accountable Care Organizations (ACOs) in the US are a final type of an alternative VBP model. ACOs are multidisciplinary groups or networks of providers that have voluntarily agreed to be held accountable for the cost and quality of care for a patient population assigned to them. Multiple payment options exist, but a frequently used model is a global payment with risk sharing. Under this model, ACOs share in realized savings (and potentially losses too) with the payer, conditional on reaching certain quality targets. In contrast to traditional capitation, under this payment model high-quality care is rewarded and risk-mitigating measures such as reinsurance provisions are included.

In the Netherlands, VBP reform is also high on the (political) agenda. In 2018 the Dutch Ministry of Health, Welfare, and Sport introduced a program to stimulate value in health care by – amongst other things – investing in alternative payment models (in Dutch: 'Programma uitkomstgerichte zorg 2018-2022'). In addition, in their advice on the future of provider payments in secondary care, the Dutch Healthcare Authority recommends stakeholders to invest in alternative payment contracts and reward high-value care (in Dutch: 'Advies bekostiging medisch-specialistische zorg: Belonen van zorg die waarde toevoegt'). Furthermore, in 2018 a working group composed of stakeholders in the field of provider payment reform (i.e., providers, patients, insurers, regulators, and scientists) was formed in the context of the national Linnean Initiative, with the goal of accelerating the uptake of VBP in the Netherlands (in Dutch: 'Werkgroep bekostiging). Another example is the 2015 reform of the payment system for primary care, which since then includes the option to negotiate with insurers a bundled payment for diabetes,

chronic obstructive pulmonary disease, vascular risk management, and asthma (in Dutch: 'keten-DBC's'), and the option to explicitly reward innovation and improving outcomes (i.e., P4P).[4] Also in the hospital sector, several bundled payment initiatives were recently started (Cattel et al. 2021). Finally, in 2013 several experiments with regional population health management were initiated (in Dutch: 'regionale proeftuinen'). Although the ambition of many of these initiatives was to introduce alternative (population-based) payment models, an evaluation of nine different initiatives showed that this ambition has not been realized yet in practice (Drewes et al. 2018).

In sum, VBP for providers is 'hot and happening'. However, despite substantial literature on the theory and implementation of provider payment incentives (e.g., McGuire 2000; McGuire 2011; Conrad et al. 2014; Conrad 2015; Conrad et al. 2016), little is known about what VBP models for providers should look like. Specifically, the relationship between what a healthcare system ideally pursues in terms of value and what is required in terms of VBP design to stimulate the desired provider behavior, has not been explicated. Therefore, the second research question of this dissertation is:

**Q2: What are the key design elements of a theoretically preferred value-based payment model?**

Based on key theoretical and empirical studies on provider behavior and payment incentives, we describe how an 'optimal' provider payment system in theory looks like given our five-dimensional definition of value in health care (section 1). The insights from this paper are of practical relevance for stakeholders who are responsible for (re)designing existing and future VBP initiatives.

After constructing a conceptual framework of a theoretically preferred VBP, an interesting question is whether initiatives that come close to this theoretically 'optimal' design have been implemented in practice and if so, how the payment models are designed in these initiatives and to which extent they are effective in improving value. Therefore, the third research question of this dissertation is:

**Q3: Which initiatives exist in practice that come close to a theoretically 'optimal' VBP model, how are they designed, and what is their impact on value?**

To provide an answer to this question, a systematic review of the literature is conducted. By systematically identifying and describing initiatives from practice that match the definition of a theoretically 'optimal' VBP model, a comprehensive overview of the design and effects of these initiatives is provided. In doing so, we aim to provide stakeholders with insight in promising and

4 Bundled payments for these conditions were already broadly introduced in 2010 on an experimental basis (De Bakker et al. 2012).

practically feasible modalities of VBP reform. This could support innovation and facilitate future provider payment model comparison.

A growing number of provider payment reform initiatives rely on global payments applied in primary care settings. Under these reforms, primary care providers (PCPs) receive a prospectively determined fixed amount for each registered or assigned individual in their patient panel, covering a specified care package for a defined period. In contrast to a traditional payment per person to PCPs, the payment does not only pertain to primary care services but also to other types of care, such as prescription medication and medical specialist care. A key characteristic of these type of payments is that – because of their prospective nature and the care package stretching beyond single services, diseases or treatments in primary care – PCPs are exposed to greater amounts of financial risk for medical spending than under conventional payment models in primary care. As providers become to some extent accountable for discrepancies between spending and payments, incentives for cost control increase. A potential disadvantage, however, is that without ancillary measures providers may be exposed to excessive financial risk, which might result in low rates of provider participation in the payment program, unwanted bankruptcies, and/or strategic provider behavior such as risk selection. Thus, an important question is how financial risk can be kept manageable for PCPs under global payments, while maintaining incentives for cost control. Answering this question requires insight in the key determinants of financial risk and the interplay between these determinants. Therefore, the fourth research question of this dissertation is:

**Q4: Which determinants of financial risk related to global payment design can be distinguished and what is their relative impact on the financial risk of primary care providers subjected to global payments?**

To answer research question 4, we empirically simulate prospective global payments for PCPs using rich administrative data on medical spending and risk characteristics of over 4.2 million individuals enrolled with a large Dutch health insurer. We examine the relative impact on PCPs' financial risk of key determinants of that risk related to the design of the global base payment. This research contributes to the body of knowledge concerning smarter choices in provider payment design and could help those involved in primary care payment reform in making better-informed decisions regarding payment design and appropriate levels of financial risk for providers.

Risk adjustment and risk sharing are important measures to reap the benefits of global provider payments while mitigating adverse effects related to risk selection and excessive financial risk. With risk adjustment, provider payments are based on predicted spending of a population given a predefined set of population characteristics (such as age, gender and morbidity). With risk sharing, provider payments are (partly) based on observed spending. Unfortunately, these two measures are not without drawbacks. Risk adjustment based on prior utilization and diagnoses might confront providers with incentives for upcoding (i.e., incentives to overstate measured pa-

tient risk in order to increase payments), while risk sharing creates a direct link between payments and spending and thereby reduces incentives for cost control. Designing risk adjustment and risk sharing for (global) provider payment thus involves a tradeoff between incentives for cost control, incentives for risk selection, incentives for upcoding, and excessive financial risk. In the light of this tradeoff, the developing field of provider payment might benefit from insights from the field of health plan payment. Specifically, an innovative form of risk sharing that was recently proposed in that field – risk sharing based on residual spending after risk adjustment – may well be an interesting option in the context of provider payment. Under this approach, providers receive extra payments for those individuals most heavily underpaid by the risk-adjustment model and must make repayments for heavily overpaid individuals. At least in theory, residual-based risk sharing substantially reduces incentives for risk selection, incentives for upcoding, and excessive losses/profits for providers, while the reduction in incentives for cost control is limited. Despite its potential, this form of risk sharing has not been studied in the context of provider payment and insight into the incentive effects and tradeoffs associated with the design of residual-based risk sharing is lacking. Therefore, the last research question of this dissertation is:

**Q5: What is the effect of residual-based risk sharing for providers on (1) incentives for cost control, (2) incentives for risk selection, (3) incentives for upcoding, and (4) excessive losses/profits for providers.**

Using rich administrative data on medical spending and risk characteristics of over 4.4 million individuals enrolled with a large Dutch health insurer, we simulate risk-adjusted global payments for primary care providers (PCPs) for a comprehensive care package, and apply various residual-based risk-sharing modalities that differ in the funds devoted to risk sharing and in whether only residual-based payments or both payments and repayments are used. We simulate the effects on cost-control incentives, risk selection incentives, upcoding incentives, and excessive provider-level losses/profits and provide an answer to research question 5. The resulting insights in incentive effects and associated tradeoffs are expected to be of substantial value for providers, purchasers, and policymakers in designing better provider payment models.

## 6. STRUCTURE OF THIS DISSERTATION

This dissertation is structured as follows. In part I (VBP incentives for consumers), chapter 2 compares cost-containment incentives for consumers under three different deductible designs (Q1). Part II focuses on VBP incentives for providers. In chapter 3 a conceptual framework of a theoretically 'optimal' VBP design is presented (Q2). Chapter 4 summarizes the results of an extensive systematic review of the literature on 'optimal VBP' in practice (Q3). Chapters 5 and 6 contain the results of two empirical simulation studies on key determinants of financial risk for

PCPs subjected to global payments (Q4 and Q5). Finally, in chapter 7 the main findings of the preceding chapters are summarized and discussed. In addition, the implications for policy and practice of the findings are discussed as well as important topics for further research.

# Part I

## Value-based payment incentives for consumers

# Chapter 2

## A method to simulate incentives for cost containment under various cost sharing designs: An application to a first-euro deductible and a doughnut hole

With Richard van Kleef and René van Vliet

European Journal of Health Economics, 2016, 18: 987-1000

## ABSTRACT

Many health insurance schemes include deductibles to provide consumers with cost containment incentives (CCI) and to counteract moral hazard. Policymakers are faced with choices on the implementation of a specific cost sharing design. One of the guiding principles in this decision process could be which design leads to the strongest CCI. Despite the vast amount of literature on the effects of cost sharing, the relative effects of specific cost sharing designs—e.g., a traditional deductible versus a doughnut hole—will mostly be absent for a certain context. This paper aims at developing a simulation model to approximate the relative effects of different deductible modalities on the CCI. We argue that the CCI depends on the probability that healthcare expenses end up in the deductible range and the expected healthcare expenses given that they end up in the deductible range. Our empirical application shows that different deductible modalities result in different CCI and that the CCI under a certain modality differs across risk-groups.

## 1. INTRODUCTION

There is a vast amount of literature on the effects of consumer cost sharing on moral hazard (Arrow 1963; Pauly 1968; Zweifel & Manning 2000). The RAND experiment, for example, has shown that a higher level of cost sharing generally results in less moral hazard (Newhouse 1993). It is therefore not surprising that most health insurance schemes include cost sharing arrangements to provide consumers with incentives for cost containment and counteract moral hazard (Baicker & Goldman 2011; Hartman et al. 2015; Qingyue et al. 2011; Stabile et al. 2013; Zare & Anderson 2013). Policymakers are faced with choices on the implementation of a specific cost sharing design. Should, for example, a first-euro deductible[5] (i.e., up to the deductible amount, insured are obliged to pay 100% of their healthcare expenses out-of-pocket in the contract period, generally a calendar year) be favored rather than a 'doughnut hole' (i.e., insured experience a gap in coverage starting after they have incurred a fixed amount of healthcare expenses)? In this case, policymakers decide on the timing of onset of a deductible during the contract period. Under a first-euro deductible, the timing is initial, while under a 'doughnut hole' the timing of onset is delayed, since individual healthcare expenses are required before this modality comes into effect. One of the guiding principles in this decision process on the cost sharing design could be which specific cost sharing design is expected to lead to the strongest incentives for cost containment.

Despite the vast amount of literature on the effects of cost sharing, the relative effects of specific cost sharing designs will mostly be absent. In these situations, methods to simulate incentives for cost containment under various cost sharing designs may be helpful for policymakers to underpin decisions on the design of effective consumer cost sharing in health insurance. To the best of our knowledge, such a method is not yet described in the literature. This paper focuses on the deductible as a cost sharing mechanism and aims at developing a simulation model to approximate the relative effects of different deductible modalities on incentives for cost containment. We simulate the individual's cost containment incentives (henceforth referred to as the CCI) as expected at the start of the contract period, given the individual's expected healthcare expenses. We focus solely on the CCI at the start of the insurance contract—rather than on the evolution of the CCI during the contract period—since benefit design decisions are usually made prior to the start of the insurance contract. In addition to developing a simulation method, we empirically illustrate this method for a first-euro deductible and a doughnut hole.[6] In this illustration we will simulate average CCIs for the total population and, separately, CCIs for groups of low-risk individuals and high-risk individuals.

5 Or a first-dollar deductible.

6 In this paper we do not pursue optimization of the deductible design. We use designs from practice to illustrate the methodology to simulate the CCI. Nevertheless, the framework can be used as a tool to gain insight in the properties of other deductible modalities and compare deductible designs in terms of the CCI.

Our method is based on the classical economic theory that consumers act like a homo economicus and possess traits such as perfect self-interest, rationality, and information. For the homo economicus the CCI is affected by the marginal out-of-pocket expenses given the individual's expected spending in the contract period. We will argue that these marginal out-of-pocket expenses depend on two parameters. The first parameter is the probability that individual healthcare expenses end up in the deductible range. Ceteris paribus, the CCI is expected to decrease with this probability. The explanation is that individuals will hardly experience any incentives for cost-conscious behavior when they expect their expenses to (far) exceed the deductible range; any savings will reduce the insurance claim, but not their out-of-pocket expenses (Keeler et al. 1977; Newhouse 1993). Given that expenses of an individual end up in the deductible range (hypothetically speaking), there is a second parameter of concern: the total expected expenses in the deductible range.[7] The higher the total expected expenses—given that they end up in the deductible range—the higher the savings potential is, and the stronger the CCI will be.

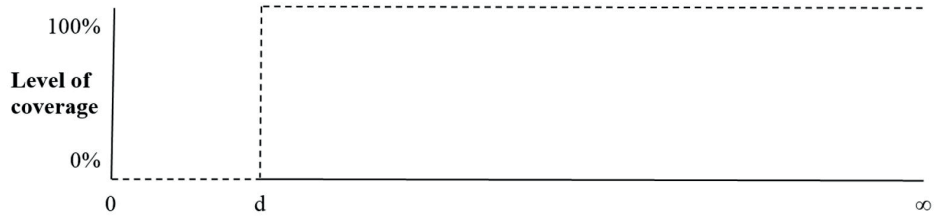
The structure of this paper is as follows. In the next section, the two deductible modalities under study are introduced followed by a section in which the relevant parameters for approximating the CCI are specified. Section 4 briefs about the conceptual framework to simulate the CCI. Data and methods are described in sections 5 and 6. Results are presented before the concluding section. Finally, in section 8 conclusion and discussion are summarized.

## 2. DEDUCTIBLE MODALITIES

In our conceptual model and empirical illustration, we study two deductible modalities applied in practice: (1) a first-euro deductible and (2) a doughnut hole. A first-euro deductible is the most commonly applied deductible modality and implies that patients pay the first €d of healthcare expenses out of their own pocket, before the insurer takes over and reimburses all excess healthcare expenses covered by the benefit package. The timing of onset of this deductible is initial. In Figure 2.1 expenses in the interval [0, d] are borne by the insured, while expenses in the interval [d, ∞] are borne by the insurer. First-euro deductibles can be, for example, found in the US, the Netherlands and Switzerland.

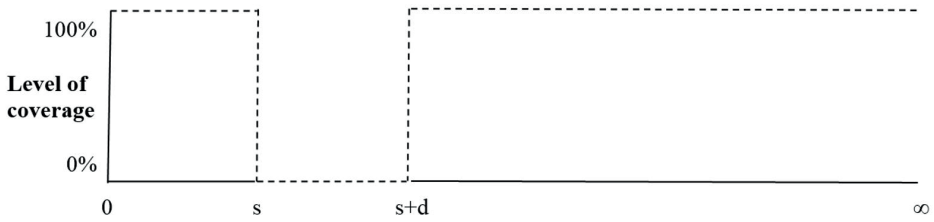
7 Expected expenses are considered to be the total expected healthcare expenses that fall under the basic benefit package.

**Figure 2.1.** Insurance under a first-euro deductible with range [0, d]




A doughnut hole is a deductible that starts at a higher level of healthcare expenses than €0. In contrast to a first-euro deductible, the timing of onset of this deductible modality is delayed, since individual healthcare expenses are required before this modality comes into effect. A 'doughnut hole' can be seen as a 'shifted' deductible with a uniform starting point. The latter means that the starting point of the doughnut hole is fixed for all individuals and set, for example, at the mean of actual healthcare expenses in the population in the previous year. Figure 2.2 shows that full coverage is provided for those expenses ranging from 0 to the starting point of the doughnut hole (interval [0, s]). Then, insured enter the doughnut hole and experience a gap in coverage. Healthcare expenses from the starting point of the deductible s until the endpoint s + d must be paid out-of- pocket (interval [s, s+d]). Full coverage is again provided by the insurer if healthcare expenses exceed the doughnut hole (interval [s+d, ∞]). An example of this modality can be found in the Medicare drug coverage system that was implemented in 2006 in the US (part D).

**Figure 2.2.** Insurance under a doughnut hole with range [s, s+d]




## 3. INCENTIVES FOR COST CONTAINMENT: WHAT ARE THE RELEVANT PARAMETERS?

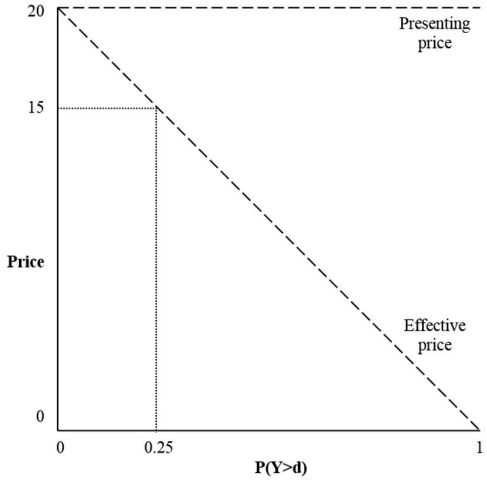
Our framework starts from the idea that consumers behave rationally. Though this assumption is probably unrealistic and oversimplistic, it provides a theoretical starting point for the development of our framework. As we will discuss in the end of this paper, we believe it is possible to extend the framework with other assumptions on consumer behavior that may follow from (future) empirical studies. The central point of our framework is that for a perfectly rational consumer

the CCI in a deductible plan depends on the marginal out-of-pocket expenses given the expected spending in the contract period. More specifically, we will argue that the CCI depends on (1) the probability that individual healthcare expenses end up in the relevant deductible range and (2) the total expected expenses given that they end up in the relevant deductible range. The relevant deductible range represents the interval where the individual, instead of the insurer, bears the costs. In sections 3.1 and 3.2 we discuss these two parameters in more depth.

### 3.1 The probability that individual healthcare expenses end up in the relevant deductible range

Theory predicts that, in case of a first-euro deductible, the price sensitivity of an individual is negatively correlated with the probability that healthcare expenses exceed the deductible amount, ceteris paribus (Keeler et al. 1977; Newhouse 1993). For a doughnut hole, the price sensitivity of an individual is expected to be negatively correlated with the probability that healthcare expenses do not fall in the deductible range, keeping other things equal. This principle can be illustrated by the following anecdotal example from Newhouse [1993:81]: "Consider a consumer on the Experiment plan with a 50% coinsurance plan and a $1000 maximum dollar expenditure (MDE). In any year, this person will have free care after spending $2000 on healthcare services. Suppose the person knows in advance that she will spend at least $2000; then any additional care she decides to purchase today is, in effect, free. Alternatively, suppose the person knows that she will not spend as much as $2000; then any additional care she decides to purchase today will cost 50 cents on the dollar because she will not anticipate free care later in the year." This example implies that a utility-maximizing consumer uses the presenting price of a visit (i.e., the real price) minus the product of the probability to exceed the MDE and the presenting price to determine whether a visit is worth its costs. This can be defined as the effective price (Newhouse 1993). For example, if the probability of exceeding the deductible amount is 0.25, the effective price for healthcare to the insured of a €20 visit is €15 (€20 minus the product of 0.25 and €20). The principle of varying effective prices with the probability of having 'free' healthcare is shown in Figure 2.3.

**Figure 2.3.** Presenting price versus effective price under a deductible




Note. P = probability; Y = healthcare expenses; d = deductible amount.

The theory of effective prices suggests that, in some cases, an individual perceives himself as completely insured or completely uninsured and thus experiences a weak or strong CCI. For example, if for a first-euro deductible the probability that healthcare expenses exceed the deductible amount approximates 0, the individual perceives himself as completely uninsured and the effective price equals the presenting price, which suggests a relatively strong CCI. In contrast, if for a first-euro deductible the probability that healthcare expenses exceed deductible amount is close to 1, the individual perceives himself as completely insured and the effective price is €0 which implies a relatively weak CCI. In the latter case, cost-conscious behavior will not prevent the individual from reaching the maximum on out-of-pocket expenses (Newhouse 1993; Van Kleef et al. 2009; Van Kleef et al. 2011). Under a first-euro deductible, an individual thus perceives himself as completely uninsured if he knows for sure—hypothetically speaking—that total healthcare expenses end up in the interval [0, d]. Under a doughnut hole, this is the case if an individual knows for sure that total healthcare expenses end up in the doughnut hole (interval [s, s+d]). In contrast, an individual perceives himself as completely insured under a first-euro deductible, if he knows for sure that total healthcare expenses will end up in the interval [d, ∞]. Under a doughnut hole, this is the case if the individual knows for sure that total expenses end up in the intervals [0, s] or [s+d, ∞]. Though it is unrealistic to assume that individuals know for sure whether or not healthcare expenses end up in a specific deductible interval, the aforementioned examples illustrate how the CCI depends on the probability to end up in the deductible range.

Theoretically, the probability that an individual's healthcare expenses end up in the deductible range depends on three parameters: (1) the amount of healthcare that is already used in the contract period, (2) the number of days remaining in the contract period, and (3) the expected healthcare expenses for the remainder of the contract period (Keeler et al. 1977). Since we focus

on the CCI at the start of the contract period (and not on how the CCI evolves through the contract) the first two parameters are not relevant here.[8] This implies we will solely focus on the link between expected spending and the CCI. In general, higher expected spending at the start of the contract period implies a higher probability to exceed the deductible.

### 3.2 The total expected expenses given that they end up in the relevant deductible range

As discussed in the previous section, the probability that healthcare expenses end up in the deductible range is an important determinant in approximating the CCI. Nevertheless, we argue it is not the only relevant parameter. Consider the following hypothetical situation where two individuals are subject to a first-euro deductible of €500. Both individuals know with certainty that healthcare expenses remain below this deductible amount.[9] Assume that person A has expected expenses in the deductible range of €100 and person B has expected expenses in the deductible range of €400. In this case, it would be inaccurate to conclude that the CCIs for these individuals are equal. In this specific case, B has a stronger CCI than A, since the expected expenses for which the individual is price sensitive due to the probability of not exceeding the deductible are higher for B than for A. In other words, B has a higher savings potential than A. Building on this example, we state that the expected healthcare expenses given that they end up in the deductible range is a relevant parameter for the CCI too.

## 4. A METHOD TO SIMULATE INCENTIVES FOR COST CONTAINMENT

In this section we build a conceptual framework to simulate the CCI under different deductible modalities at the start of the contract period. We describe our method for a first-euro deductible and a doughnut hole.

### 4.1 First-euro deductible

Under a first-euro deductible, the deductible range where the individual bears the costs equals [0, d]. Accordingly, the CCI under a first-euro deductible can be simulated by combining the probability P that individual healthcare expenses Y remain below the deductible amount d and the expected expenses E(Y) given that expenses Y remain below the deductible amount d:

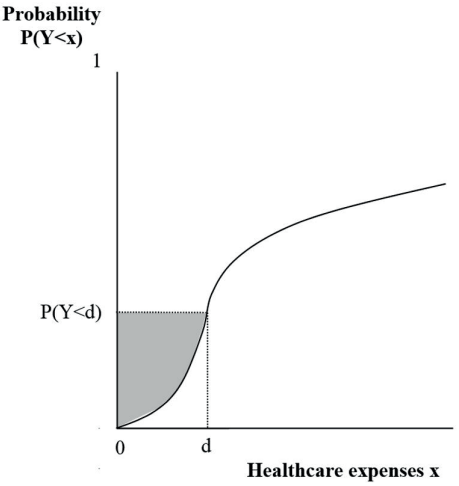
$$CCI_{\text{first-euro deductible}} = P(Y < d) * E(Y|Y < d) \quad (1)$$

8 Nevertheless, the conceptual framework can be refined to facilitate simulation of the CCI during the contract period. By determining the CCI on multiple moments (i.e., by repeating the procedure that is described in this paper), the other two parameters can be taken into account.

9 Or: both individuals have an equal probability that healthcare expenses exceed the deductible amount [i.e., P(Y<d)<1].

The essence of the CCI can be graphically illustrated with Figure 2.4. Consider the curve in Figure 2.4 to represent the probability of an individual's healthcare expenses to remain below amount x. For an infinite value of x, this probability equals 1, which means that all expenses are in the interval [0, x]. In this extreme case E(Y|Y<x) equals E(Y) and the outcome of equation (1) exactly represents the total area above the curve. This is no longer true, however, when P(Y<x) is smaller than 1, which is the case for x = d. Since P(Y<d) is smaller than 1 and E(Y|Y<d) is smaller than E(Y), the outcome of equation (1) no longer represents the total area above the curve, but shrinks to the shaded area. Here we come to the essence of our method: when the shaded area of deductible modality A is larger than that of deductible modality B, the CCI is expected to be stronger under modality A than under modality B.

**Figure 2.4.** CCI under a first-euro deductible




## 4.2 Doughnut hole

Under a doughnut hole, the endpoint of the deductible range is marked by s+d. P(Y<s+d) and E(Y|Y<s+d) are higher compared to P(Y<d) and E(Y|Y<d) under a first-euro deductible with deductible amount d. Consequently, the CCI for the interval [0, s+d] will be stronger than the CCI for the interval [0, d]. It is incorrect, however, to assume that the CCI under a doughnut hole equals the CCI for the complete interval [0, s+d]. This can be illustrated with an infinite value for s: here both P(Y<s+d) and P(Y<s) equal 1. In this case it would be inaccurate to conclude that the CCI equals P(Y<s+d) * E(Y|Y<s+d), since all expenses are in the interval [0, s] and are fully reimbursed by the insurer. In other words, no expenses appear in the interval [s, s+d] where the individual bears the costs. So, we argue that, in this specific example, the CCI should equal 0 and, in general, the negative effect of interval [0, s] on the CCI should be incorporated in the calculation of the CCI. The latter implies that when determining the CCI under a doughnut hole, the

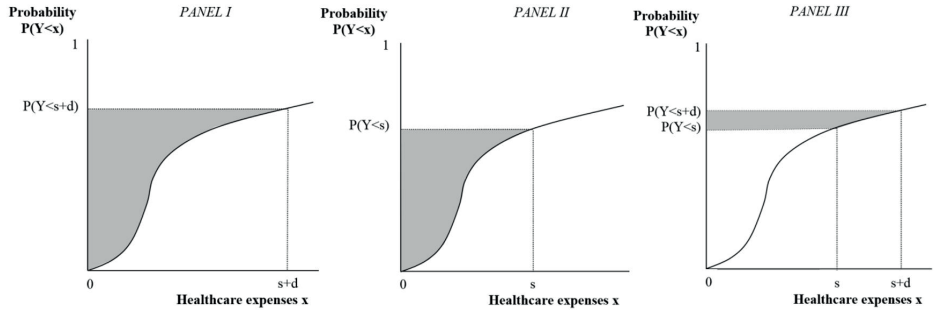
focus should be on the expenses where the insured are price sensitive due to the probability of entering the doughnut hole but not reaching the endpoint of the doughnut hole.

This reasoning implies that the CCI under a doughnut hole can be approximated by the product of P(Y<s+d) and E(Y|Y<s<d) minus the product of P(Y<s) and E(Y|Y<s). Accordingly, the CCI under a doughnut hole can be calculated by equation (2).

$$CCI_{\text{doughnut hole}} = [P(Y < s+ d) * E(Y|Y < s+ d)] - [P(Y < s) * E(Y|Y < s)] \quad (2)$$

This procedure is graphically illustrated in Figure 2.5 where the shaded area in panel I represents P(Y<s+d) * E(Y|Y<s+d), the shaded area in panel II represents P(Y<s) * E(Y|Y<s), and the shaded area in panel III represents the outcome of equation (2).

**Figure 2.5.** The CCI under a doughnut hole




## 5. DATA

For the empirical application of our method, we used administrative data from Dutch insurers operating under the Health Insurance Act. We used a sample of 500,000 individuals who were randomly selected from the total Dutch population of 18 years and older and enrolled in the basic health insurance for a complete calendar year (2011). The sample is similar to the total Dutch population regarding mean, standard deviation, minimum, and maximum.

The dataset includes individual-level risk-information on healthcare expenses and risk-characteristics. The risk characteristics are age-gender classes, diagnoses cost groups (DCGs), pharmacy-based cost groups (PCGs), high-cost groups (HCGs) and multiple prior years high costs (MHCs). In the Netherlands this information is used in the Dutch risk-equalization system. Further information on these risk characteristics can be found in previous work (see, for example, Van Veen et al. 2015a). In addition to information on risk characteristics, the dataset includes information on total healthcare expenses in 2011 that are covered by the Dutch basic health insurance (e.g., costs for general practitioner care, hospital care, pharmaceutical care and mental care). Based on visual inspection, we excluded 10 insured with extremely high healthcare

expenses ranging from €223,184 till €467,722 from the full sample of 500,000 insured, because they appeared to negatively affect our expenditure model. On average in the selected sample of 499,990 individuals, the actual healthcare expenses were €2257 with a standard deviation of €6124, a minimum of €1, a median of €593 and a maximum of €217,566.

## 6. METHODS

To empirically illustrate our method for simulating the CCI under different deductible modalities we follow a four-step procedure:

1. Estimate an expenditure model;
2. Approximate the probability that healthcare expenses end up in the deductible range;
3. Approximate the expected expenses given that they end up in the deductible range;
4. Simulate the CCI.

In this paper we are interested in the CCI under a specific deductible modality relative to others; absolute figures of the CCI are of little significance. Empirical results are intended as an illustration of the method developed. First, we derive the CCI under a first-euro deductible of €500, €1000, €2000, €3000, €4000, €5000 and €10,000 in order to examine the effects of the deductible amount. After that, we examine the CCI under a doughnut hole of €1000 with a uniform starting point at €500, €1000, €2000, €2257 (i.e., the mean of actual healthcare expenses in the selected sample of 499,990 individuals), €3000, €4000 and €5000 in order to compare the CCI between a first-euro deductible and a doughnut hole. Average CCIs under the two deductible modalities are simulated for the full sample, and separately, for a group of high-risk individuals and the complementary group of low-risk individuals. Morbidity information is used to determine to which risk-group an individual belongs: those individuals with (without) a DCG, PCG, HCG and/or MHC are considered as a high-risk individual (low-risk individual). In this sample 72% is considered as a low-risk individual and 28% as a high-risk individual.

It is important to mention that – next to the assumption on rational behavior – our concept is based on some other (implicit) assumptions. For example, we assume a linear relationship between the probability that healthcare expenses end up in the deductible range and the CCI. Furthermore, we focus on the CCI regarding total healthcare utilization that is subject to the deductible and neglect the composition of the care that is used. The implications of these and other assumptions, will be discussed in section 8.

### 6.1 Estimate an expenditure model

First, to predict expected healthcare expenses E(Y) for each individual, an expenditure model is estimated with actual expenses in 2011 as dependent variable and age-gender classes, DCGs, PCGs, HCGs and MHCs as explanatory variables. We opted for a Generalized Linear Model

(GLM) with a gamma distribution and a log-link function, which is considered to be an appropriate statistical method for modelling healthcare expenses in many studies (e.g., Beeuwkes-Buntin 2004; Blough et al. 1999; Duan et al. 1983; Manning & Mullahy 2001; Van Kleef et al. 2009). Basically, all risk characteristics are statistically significant at the conventional level (given the large sample size). On average the expected healthcare expenses were €2537 with a standard deviation of €7762, and the $R^2$ of the model is 0.39. In the subsequent tables we show that our estimation approach provides an acceptable fit between the actual and predicted parameters of the CCI.[10]

## 6.2 Approximate the probabilities that healthcare expenses end up in the deductible range

After estimating an expenditure model, the probability P that healthcare expenses Y remain below deductible amount d, starting point s and endpoint s+d is approximated. We follow the procedure as described by van Kleef and colleagues (2009), who have identified the relevant parameters given the use of a gamma distribution with a log-link. The probabilities that we are interested in can be derived by equations (3) to (5).

$$P(Y < d) = \Gamma(c_d,k) \quad (3)$$

$$P(Y < s) = \Gamma(c_s,k) \quad (4)$$

$$P(Y < s+d) = \Gamma(c_{s+d},k) \quad (5)$$

where c(.) is the cumulative density function of the gamma distribution, the scale parameter k is 0.4969, and:

$$\lambda = k/E(Y) \quad (6)$$

$$c_d = d^*\lambda \quad (7)$$

$$c_s = s^*\lambda \quad (8)$$

$$c_{s+d} = (s+d)^*\lambda \quad (9)$$

Given the assumptions made and given our dataset, we check whether the results based on equations (3) to (9) are in line with the actual figures in the sample; the proportion $\varrho$ and probability P that healthcare expenses Y remain the deductible amount d under a first-euro deductible are compared. Table 2.1 shows that $\varrho$ (Y<d) and P(Y<d) follow the same pattern, specifically in case of a relatively high deductible amount.

10 We also took into consideration other specifications of the model varying in terms of distribution and link-function. We opted for a GLM with a gamma distribution and a log-link function based on a comparison of mean, standard deviation, minimum, maximum and mean absolute predicted error of actual and expected expenses in the sample and per expenditure quintile.

**Table 2.1.** Proportions ϱ and probabilities P in the sample that healthcare expenses Y remain below various deductible amounts d

| d | ϱ (Y<d) | P(Y<d) |
|---|---|---|
| 500 | 0.47 | 0.43 |
| 1,000 | 0.61 | 0.57 |
| 2,000 | 0.75 | 0.73 |
| 3,000 | 0.82 | 0.81 |

### 6.3 Approximate the expected expenses given that they end up in the deductible range

Given expected expenses E(Y) and the parameters calculated in the previous step, expected expenses given that expenses end up in the interval [0, d], [0, s], respectively [0, s+d] can be calculated by equations (10), (11), and (12) (Van Kleef et al. 2019).

$$E(Y|Y < d) = E(Y) * \Gamma(c_d,k + 1) / \Gamma(c_d,k) \quad (10)$$

$$E(Y|Y < s) = E(Y) * \Gamma(c_s,k + 1) / \Gamma(c_s,k) \quad (11)$$

$$E(Y|Y < s+d) = E(Y) * \Gamma(c_{s+d},k + 1) / \Gamma(c_{s+d},k) \quad (12)$$

Table 2.2 shows the actual expenses and expected expenses given that expenses remain below first-euro deductible amount d. Our approach somewhat underestimates these expenses for the relatively small first-euro deductibles and somewhat overestimates them for the higher ones, but these deviations do not seem important.

**Table 2.2.** Mean of actual expenses Y and expected expenses E(Y) in the sample given that expenses Y remain below various deductible amounts d

| d | Y\|Y<d | E(Y\|Y<d) |
|---|---|---|
| 500 | 186 | 158 |
| 1,000 | 314 | 302 |
| 2,000 | 517 | 551 |
| 3,000 | 688 | 755 |

Based on the results presented in Tables 2.1 and 2.2, there seems to be no reason to believe that the overestimations of the mean and the standard deviation of expected healthcare expenses compared to the actual healthcare expenses have unacceptable effects on the key parameters of interest in this paper.

### 6.4 Simulate the CCI

As discussed in section 4, the CCI is conceptualized as a product of the probability that individual healthcare expenses end up in the deductible range and the expected expenses given that they end up in the deductible range. Therefore, parameters obtained in step 2 and step 3 are combined

in order to determine the CCI for each individual. The CCI under a first-euro deductible with deductible amount d is calculated by equation (1) The CCI under a doughnut hole with starting point s and deductible amount d is approximated by equation (2). The CCI is presented in Euros and can be interpreted as the marginal amount of healthcare expenses for which a consumer is fully price sensitive. Hypothetically speaking, the CCI will be zero for a consumer who knows for sure his spending will exceed the deductible amount. For a consumer who knows for sure his spending will not exceed the deductible amount, the CCI will equal his expected spending.

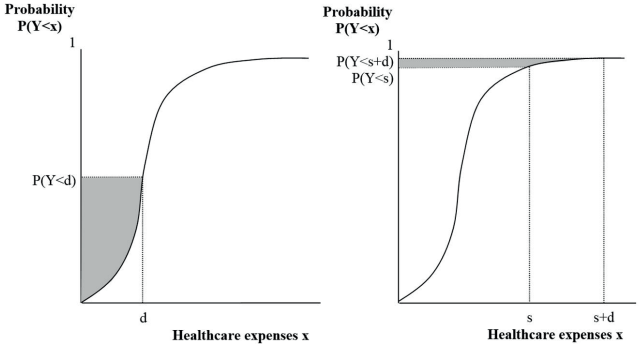
### 6.5 Implications

At least three implications arise from the conceptual framework as described in section 4. These hypotheses are to be addressed in section 7 where the simulation results are presented. First, the CCIs under a deductible increase when the deductible amount increases. If, ceteris paribus, the deductible amount increases (i.e., point d and, accordingly, point s+d is shifted to the right), the deductible range is broadened. As a result, both the probability that expenses end up in the deductible range and the expected expenses in the deductible range once they ended up in the deductible interval are expected to increase. This will result in a stronger CCI.

Second, we expect that different deductible modalities lead to different CCIs. Shifting the deductible influences the CCI. The direction of the effect is an interesting empirical question. On the one hand, a shift of the deductible to higher expenditure levels reduces the probability to reach the deductible range, which negatively affects the CCI. On the other hand, such a shift increases the expected expenses given that they end up in this range, which positively affects the CCI.

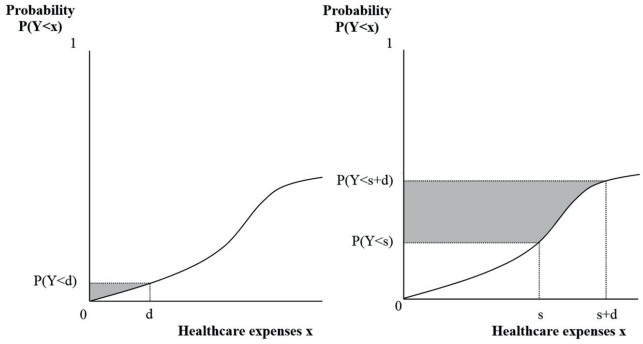
Third, we hypothesize that the CCI under a first-euro deductible and a doughnut hole will differ across risk groups. Figure 2.6 shows P(Y<x) of a relatively low-risk individual under a first-euro deductible (left panel) and under a doughnut hole with a starting point at the mean of actual healthcare expenses in the population (right panel). E(Y) for this healthy individual are relatively low, but there is always a certain level of uncertainty whether this individual needs care. This implies that, under a first-euro deductible, there is a low probability that healthcare expenses exceed the deductible amount. In contrast, under a doughnut hole with a starting point at the mean of healthcare expenses, it is not very likely that this low-risk individual ends up in the doughnut hole. P(Y<s) and P(Y<s+d) both approximate 1. As a result of the relatively high P(Y<d) under a first-euro deductible compared to P(s<Y<s+d) under a doughnut hole, the CCI for this low-risk individual is relatively strong in case of a first-euro deductible in comparison to a doughnut hole.

**Figure 2.6.** The CCI for a relatively low-risk individual under a first-euro deductible (left panel) and under a doughnut hole with a starting point at the mean of actual healthcare expenses in the population (right panel)




Now consider a relatively high-risk individual, such as a chronically ill patient. P(Y<x) is depicted in Figure 2.7. E(Y) for this relatively unhealthy individual are above average. Accordingly, under a first-euro deductible, P(Y<d) is low (Figure 2.7, left panel). In contrast, P(s<Y<s+d) is relatively high when the starting point of the doughnut hole is set at the mean of actual healthcare expenses (Figure 2.7, right panel). Consequently, for this high-risk individual the CCI is relatively strong in case of a doughnut hole in comparison to a first-euro deductible.

**Figure 2.7.** The CCI for a relatively high-risk individual under a first-euro deductible (left panel) and under a doughnut hole with a starting point at the mean of actual healthcare expenses (right panel)




The previous consideration implies that, at the population level, it is not obvious whether a first-euro deductible leads to a stronger or weaker CCI than a doughnut hole. On the one hand, a shift of the starting point of the deductible to a higher expenditure level than €0 may increase the CCI for the high-risk individuals (a relatively small group with relatively high savings potential). On the other hand, such a shift may decrease the CCI for the low-risk individuals (a relatively large group with relatively low savings potential). In our empirical illustration we aim to simulate the net outcome of these two effects.

## 7. RESULTS

As an illustration of the method developed, this section presents the empirical results for a first-euro deductible and a doughnut hole. Results are shown for the full sample and also separately for a group of high-risk individuals and the complementary group of low-risk individuals.

### 7.1 Full sample

In Table 2.3 the results are presented for a first-euro deductible of various deductible amounts for the total sample. The mean probability that healthcare expenses remain below the deductible amount, the expected expenses given that they remain below the deductible amount, and the product of these two parameters are shown. As hypothesized in section 6.5, Table 2.3 reveals that an increase in the deductible amount indeed leads to a higher P(Y<d) and higher E(Y|Y<d). Thus, the higher the deductible amount is, the stronger the CCI will be. Note that this conclusion also holds for a doughnut hole, as P(s<Y<s+d) and E(Y|s<Y<s+d) increase with a higher deductible amount.

**Table 2.3.** The CCI under a first-euro deductible of various deductible amounts d for the full sample

| d | P(Y<d) [a] | E(Y\|Y<d) [b] | CCI |
|---|---|---|---|
| 500 | 0.43 | 158 | 68 |
| 1,000 | 0.57 | 302 | 171 |
| 2,000 | 0.73 | 551 | 393 |
| 3,000 | 0.81 | 755 | 598 |
| 4,000 | 0.86 | 921 | 773 |
| 5,000 | 0.89 | 1,059 | 920 |
| 10,000 | 0.96 | 1,475 | 1,371 |

[a] The probability that healthcare expenses remain below the deductible amount.
[b] The expected expenses given that they remain below the deductible amount.

Table 2.4 shows (the relevant parameters for determining) the CCI under a doughnut hole of €1000 with various starting points (the CCI under a doughnut hole assuming other deductible amounts is shown in the Appendix). The mean probability that healthcare expenses remain below the starting point, respectively the endpoint of the deductible, the expected expenses given that they end up in the interval [0, s], respectively [0, s+d] and the CCI are shown. Table 2.4 shows that P(Y<s) is lower compared to P(Y<s+d). Similarly, E(Y|Y<s) are lower compared to E(Y|Y<s+d). Second, results suggest that the CCI under a doughnut hole with deductible amount €1000 increases when the starting point of the doughnut hole is shifted to the right until a starting point of €1000 is used. On average, a stronger CCI is realized under a doughnut hole with a starting point at €1000 compared to a starting point at the mean of actual healthcare expenses in the sample (i.e., €2257). These results imply that, given the dataset and the assumptions made, the 'sweet spot' of the starting point is located somewhere around €1000. This finding might

suggest that the starting point of the doughnut hole should be located below the overall mean of actual healthcare expenses, implying that the starting point of the doughnut hole in the Medicare drug coverage system should be lowered, since it is currently set at the overall mean of actual healthcare expenses.

**Table 2.4.** The CCI under a doughnut hole with deductible amount d €1,000 with various starting points s for the full sample

| s | P(Y<s) [a] | P(Y<s+d) [b] | E(Y\|Y<s) [c] | E(Y\|Y<s+d) [d] | CCI |
|---|---|---|---|---|---|
| 0 [e] | 0 | 0.57 | 0 | 302 | 171 |
| 500 | 0.43 | 0.66 | 158 | 433 | 215 |
| 1,000 | 0.57 | 0.73 | 302 | 551 | 222 |
| 2,000 | 0.73 | 0.81 | 551 | 755 | 204 |
| 2,257 | 0.75 | 0.82 | 607 | 801 | 197 |
| 3,000 | 0.81 | 0.86 | 755 | 921 | 175 |
| 4,000 | 0.86 | 0.89 | 921 | 1,059 | 147 |
| 5,000 | 0.89 | 0.92 | 1,058 | 1,173 | 123 |

[a] The probability that healthcare expenses remain below the starting point of the deductible.
[b] The probability that healthcare expenses remain below the endpoint of the deductible.
[c] The expected expenses given that they end up in the interval [0, s].
[d] The expected expenses given that they end up in the interval [0, s+d].
[e] A doughnut hole with a starting point of €0 is effectively a first-euro deductible; the CCI and related probabilities and expected expenses are identical (see Table 3).

A comparison of the results under a first-euro deductible with those under a doughnut hole suggests that different deductible modalities lead to differences in CCIs. Assuming a deductible amount of €1000, a doughnut hole with a relatively low starting point leads on average to a stronger CCI compared to a first-euro deductible. For example, a first-euro deductible of €1000 leads to a CCI of €171 while a doughnut hole of €1000 with a starting point at €1000, respectively at the mean of actual healthcare expenses leads to a CCI of €222, respectively €197. Results suggest that this pattern in favor of a doughnut hole reverses (and the CCI will be stronger in case of a first-euro deductible) when the starting point of the doughnut hole is located somewhere between €3000 and €4000.

## 7.2 Low-risk individuals and high-risk individuals

Table 2.5 provides the CCI under a first-euro deductible specifically for the low-risk individuals and the high-risk individuals. For the high-risk individuals P(Y<d) is lower while E(Y|Y<d) are higher in comparison to the low-risk individuals. Under a first-euro deductible, the CCI is strongest for the low-risk individuals compared to the high-risk individuals, as long as the deductible amount is relatively low; when the deductible amount is set somewhere between €4000 and €5000, this pattern is reversed.

**Table 2.5.** The CCI under a first-euro deductible of various deductible amounts d for the low-risk individuals and the high-risk individuals

| | d | P(Y<d) [a] | E(Y\|Y<d) [b] | CCI |
|---|---|---|---|---|
| Low-risk individuals | 500 | 0.48 | 157 | 75 |
| | 1,000 | 0.63 | 296 | 187 |
| | 2,000 | 0.79 | 529 | 418 |
| | 3,000 | 0.88 | 709 | 617 |
| | 4,000 | 0.92 | 846 | 776 |
| | 5,000 | 0.95 | 952 | 899 |
| | 10,000 | 0.99 | 1,199 | 1,188 |
| High-risk individuals | 500 | 0.30 | 162 | 48 |
| | 1,000 | 0.41 | 318 | 130 |
| | 2,000 | 0.55 | 609 | 329 |
| | 3,000 | 0.64 | 875 | 547 |
| | 4,000 | 0.70 | 1,119 | 765 |
| | 5,000 | 0.75 | 1,341 | 976 |
| | 10,000 | 0.87 | 2,203 | 1,853 |

[a] The probability that healthcare expenses remain below the deductible amount.

[b] The expected expenses given that they remain below the deductible amount.

The CCI under a doughnut hole of €1000 with various starting points is shown in Table 2.6 for the two risk-groups. The CCI under a doughnut hole is stronger for the high-risk individuals than for the low-risk individuals, as long as the starting point of the deductible is shifted to the right considerably. If the starting point is set at a relatively low point (i.e., at €500 or at €1000), the CCI under a doughnut hole is stronger for the low-risk individuals. For the low-risk individuals, the 'sweet spot' of the starting point seems to be located somewhere around €1000 while for the high-risk individuals this is somewhere around the overall mean of actual healthcare expenses.

**Table 2.6.** The CCI under a doughnut hole with deductible amount d €1,000 with various starting points s for the low-risk individuals and the high-risk individuals

| | s | P(Y<s) [a] | P(Y<s+d) [b] | E(Y\|Y<s) [c] | E(Y\|Y<s+d) [d] | CCI |
|---|---|---|---|---|---|---|
| Low-risk individuals | 0 [e] | 0 | 0.63 | 0 | 296 | 187 |
| | 500 | 0.48 | 0.73 | 157 | 420 | 230 |
| | 1,000 | 0.63 | 0.79 | 296 | 529 | 231 |
| | 2,000 | 0.79 | 0.88 | 529 | 709 | 199 |
| | 2,257 | 0.82 | 0.89 | 580 | 748 | 189 |
| | 3,000 | 0.88 | 0.92 | 709 | 846 | 159 |
| | 4,000 | 0.92 | 0.95 | 846 | 952 | 123 |
| | 5,000 | 0.95 | 0.97 | 952 | 1,031 | 94 |
| High-risk individuals | 0 [e] | 0 | 0.41 | 0 | 318 | 130 |
| | 500 | 0.30 | 0.49 | 162 | 467 | 177 |
| | 1,000 | 0.41 | 0.55 | 318 | 609 | 200 |
| | 2,000 | 0.55 | 0.64 | 609 | 875 | 218 |
| | 2,257 | 0.57 | 0.65 | 680 | 940 | 219 |
| | 3,000 | 0.64 | 0.70 | 875 | 1,119 | 218 |
| | 4,000 | 0.70 | 0.75 | 1,119 | 1,341 | 211 |
| | 5,000 | 0.75 | 0.78 | 1,341 | 1,545 | 200 |

[a] The probability that healthcare expenses remain below the starting point of the deductible.

[b] The probability that healthcare expenses remain below the endpoint of the deductible.

[c] The expected expenses given that they end up in the interval [0, s].

[d] The expected expenses given that they end up in the interval [0, s+d].

[e] A doughnut hole with a starting point of €0 is effectively a first-euro deductible; the CCI and related probabilities and expected expenses are identical (see Table 5).

A comparison of the CCI under the two deductible modalities shows that, given our dataset and under the assumptions made in this research, for the low-risk individuals, a doughnut hole on average leads to a stronger CCI compared to a first-euro deductible until a starting point of €3000 or more is chosen. For example, the CCI under a doughnut hole with a starting point at €1000 is €231 compared to the CCI of €187 under a first-euro deductible. Nevertheless, only small differences exist when comparing a first-euro deductible to a doughnut hole with a starting point at the mean of actual healthcare expenses; the CCI equals €187 compared to €189. For the high-risk individuals the CCI is noticeably stronger under a doughnut hole compared to a first-euro deductible, even if the starting point is shifted to the right only moderately. The CCI is, for instance, €177 under a doughnut hole with a starting point at €500 compared to €130 under a first-euro deductible. Results suggest that for the high-risk individuals, a doughnut hole with a starting point at the mean of actual expenditures leads to a stronger CCI compared to a first-euro deductible (€219 compared to €130).

## 8. CONCLUSION AND DISCUSSION

Starting from the traditional economic theory that consumers act like a homo economicus, this paper has developed a method to simulate Cost Containment Incentives (CCI) under different deductible modalities. For a homo economicus the CCI depends on two parameters: (1) the probability that individual healthcare expenses end up in the deductible range and (2) the total expected healthcare expenses given that they end up in the deductible range. We have empirically illustrated the method for two modalities applied in practice, i.e., a first-euro deductible and a doughnut hole. Given our dataset and under the assumptions made, our findings lead to four conclusions.

First, not surprisingly, the CCI increases with the deductible amount, ceteris paribus. The developed method can be used to simulate the impact of a higher deductible on the CCI. Second, the CCI differs between deductible modalities. Which deductible modality is opted for by policymakers seems to have consequences in terms of the CCI and it can thus be valuable to take the CCI into consideration when comparing the effectiveness of these different deductible designs. In our sample, a doughnut hole with a well-chosen starting point (i.e., below €4000) on average provides a stronger CCI than a first-euro deductible. This would imply that, to realize a strong CCI, the starting point of the deductible should be higher than zero for all insured. This finding is in line with the conclusion of van Kleef and coauthors (2009). Third, the CCI differs across risk-groups. We have found that under a first-euro deductible the CCI is strongest for the low-risk individuals, as long as the deductible amount is relatively low (i.e., until the deductible amount is set somewhere between €4000 and €5000). Under a doughnut hole, the CCI is strongest for the high-risk individuals, as long as the starting point is higher than €1000. Our findings suggest that the CCI is stronger under a doughnut hole than under a first-euro deductible for both the low-risk individuals – at least when a starting point below €3000 is chosen – and for the high-risk individuals. Fourth, our results suggest that, in order to provide a stronger CCI, the starting point of the doughnut hole should not be located at the mean of actual healthcare expenses in the sample, but somewhere below that mean. This finding suggests that the CCI under the doughnut hole in the Medicare drug coverage system could be increased by lowering the starting point.

It is important to note that our empirical findings depend on several assumptions which deserve further elaboration. In addition, many important topics remain for future research. Six of these issues are discussed below. First, a note of caution should be raised against the assumption of individuals behaving completely rationally, since in practice, insured might actually act differently than the classical theory suggests. There is empirical evidence that individuals tend to overestimate small probabilities and underestimate large probabilities (Kahneman & Tversky 1979:279; Van Winssen et al. 2015). This may have consequences for the first parameter in our framework (i.e., the probability that healthcare expenses fall in the deductible range). For example, if a low-risk individual under a first-euro deductible would overestimate the probability of becoming ill, this individual's perceived probability that healthcare expenses remain below

the deductible amount decreases, leading to a weaker CCI. In addition, Brot-Goldberg et al. (2015) show that, in practice, consumer behavior departures from fully rational behavior in that sense that individuals seem to act in a myopic way. In particular, they show that in the decision of using healthcare, individuals are not responsive to the expected marginal end-of-year price but often respond to easier to understand prices such as spot prices or their prior end-of- year marginal price. This evidence suggests that the second parameter of our framework (i.e., the total expected expenses in the deductible range)might be influenced. Although there is growing empirical evidence on alternative assumptions concerning consumer behavior, there is limited research on how these 'new' assumptions should be incorporated in economic simulation studies. It is yet unclear how these insights exactly translate into our simulation framework. For instance, it would be interesting to study how our framework could be extended with weights or additional parameters to incorporate new insights.

Second, in this paper a linear relation between the probability of exceeding the deductible and the CCI is assumed. If there are reasons to believe that an alternative relationship is more realistic, it is possible to interchange the assumption of linearity and plug-in any other relationship in the conceptual framework.

Third, the expected healthcare expenses are an important parameter in the approximation of the CCI. The expenditure model based on age-gender classes, DCGs, PCGs, HCGs and MHCs probably predicts expenses less than perfectly. Therefore, obtained results cannot expected to be perfect either. Overestimated expected expenses might explain why—in contrast to what we hypothesized—a doughnut hole instead of a first-euro deductible leads to the strongest CCI for the low-risk individuals. Further research is needed to simulate the CCI with better prediction models. Significantly better predictions can be expected if expenses in previous years are added to the model, since previous expenses proved to be a strong predictor for future expenses, even when the abovementioned predictors are already included (Bertsimas et al. 2008; Van Veen et al. 2015a). A better prediction model will likely lead to a larger variance in expected expenses and larger differences in the CCI across risk groups.

Fourth, for reasons of simplicity we did not incorporate a correction for the moral hazard effect. In our empirical illustration we apply a substantially higher deductible amount (i.e., €1000) than the amount originally applied in our data (i.e., €170). If the higher deductible amount was implemented in practice this would have led to less moral hazard and thus lower healthcare expenses. An interesting question is whether or not consumers include the 'moral hazard effect' in their expectations about future healthcare expenses. If they do (e.g., by expecting lower healthcare expenses in case of a higher deductible amount) this effect should ideally be incorporated in the type of simulations applied in this paper. This would be possible by modifying the healthcare expenses on which the expenditure model is based.

Fifth, different cost sharing designs are expected to have different implications in terms of solidarity. For example, for the high-risk individuals, a first-euro deductible can be considered as socially inequitable (assuming insufficient financial compensation), because these individuals

incur, on average, higher out-of-pocket expenses than their healthy counterparts. In addition, for these high-risk individuals, a first-euro deductible can be considered as ineffective in reducing moral hazard, because these individuals know ex-ante that their yearly healthcare expenses will exceed the deductible amount. The relation between different cost sharing designs and solidarity and to what extent a stronger CCI has an effect on moral hazard reduction might benefit from future research.

Sixth, in this paper only two deductible modalities are empirically illustrated. The method developed allows approximation of the CCI under other deductible modalities as well. Examples of other modalities are a doughnut hole with a risk-adjusted starting point and an income-related deductible. Under a doughnut hole with a risk-adjusted starting point (as proposed in the literature by van Kleef et al. 2009), the location of the doughnut hole depends on specific individual risk characteristics of the insured, such as demographics, diagnostics or prior healthcare utilization. The starting point could be, for example, based on maximized uncertainty in out-of-pocket expenses or on a maximized CCI. It is expected that a doughnut hole with a risk-adjusted starting point leads to a stronger CCI than a first-euro deductible and a uniform doughnut hole. In addition to the possibility to simulate the CCI under other deductible modalities, the method provides the opportunity to determine the CCI under other forms of cost sharing than deductibles, such as co-insurance (i.e., insured are obliged to pay a percentage of the healthcare expenses per service out-of-pocket) or co-payments (i.e., insured are required to pay a predefined amount per service out-of-pocket). This might be an interesting topic for future research.

Last, we acknowledge that the CCI may be regarded as one of the multiple criteria that can be taken into consideration by policymakers when deciding on the design of effective consumer cost sharing in health insurance. Other criteria, such as the practical and political-ideological aspects of different deductible modalities could be relevant as well. For example, an important aspect in the deductible design decision would be the trade-off between a stronger CCI versus transparency and simplicity. Specifically, in a system with a doughnut hole where the starting point of the deductible depends on individual risk-characteristics, the average CCI might be higher compared to a first-euro deductible, but transparency may be worse when the majority of insured does not understand how and why certain starting points are assigned to them. Consequently, acceptance of the deductible system might be in danger. Another issue would be how policymakers will try to level the government's cash flow. Switching to a deductible system where a relatively strong CCI can be realized, might lead to a reduction in revenues from deductibles due to more cost-conscious behavior. An option to overcome this reduction in revenues would be to increase the deductible amount (Rosenthal 2004).

Though the results of our empirical illustration should be interpreted with caution, we believe the method developed in this paper to simulate the CCI can be useful to researchers, insurers and policymakers who want to indicate the relative effects of different cost sharing designs on the incentives for cost-conscious behavior.

## APPENDIX

**Table A.2.1.** The CCI under a doughnut hole with various deductible amounts d and various starting points s for the full sample

| d | s | P(Y<s) [a] | P(Y<s+d) [b] | E(Y\|Y<s) [c] | E(Y\|Y<s+d) [d] | CCI |
|---|---|---|---|---|---|---|
| | 0 | 0 | 0.43 | 0 | 158 | 68 |
| | 500 | 0.43 | 0.57 | 158 | 302 | 104 |
| 500 | 1,000 | 0.57 | 0.66 | 302 | 433 | 112 |
| | 2,257 | 0.75 | 0.79 | 607 | 709 | 102 |
| | 5,000 | 0.89 | 0.91 | 1059 | 1118 | 64 |
| | 0 | 0 | 0.73 | 0 | 551 | 393 |
| | 500 | 0.43 | 0.77 | 158 | 658 | 431 |
| 2000 | 1,000 | 0.57 | 0.81 | 302 | 755 | 426 |
| | 2,257 | 0.75 | 0.87 | 607 | 959 | 365 |
| | 5,000 | 0.89 | 0.93 | 1059 | 1269 | 226 |
| | 0 | 0 | 0.81 | 0 | 755 | 598 |
| | 500 | 0.43 | 0.84 | 158 | 842 | 621 |
| 3000 | 1,000 | 0.57 | 0.86 | 302 | 921 | 602 |
| | 2,257 | 0.75 | 0.90 | 607 | 1090 | 505 |
| | 5,000 | 0.89 | 0.94 | 1059 | 1348 | 313 |
| | 0 | 0 | 0.89 | 0 | 1059 | 920 |
| | 500 | 0.43 | 0.91 | 158 | 1118 | 917 |
| 5000 | 1,000 | 0.57 | 0.92 | 302 | 1173 | 872 |
| | 2,257 | 0.75 | 0.93 | 607 | 1290 | 722 |
| | 5,000 | 0.89 | 0.96 | 1059 | 1475 | 451 |

[a] The probability that healthcare expenses remain below the starting point of the deductible.

[b] The probability that healthcare expenses remain below the endpoint of the deductible.

[c] The expected expenses given that they end up in the interval [0, s].

[d] The expected expenses given that they end up in the interval [0, s+d].

# Part II

## Value-based payment incentives for providers

# Chapter 3

## Value-based provider payment: Towards a theoretically preferred design

With Frank Eijkenaar and Erik Schut

Health Economics, Policy and Law, 2020, 15(1): 94-112

## ABSTRACT

Worldwide, policymakers and purchasers are exploring innovative provider payment strategies promoting value in health care, known as value-based payments (VBP). What is meant by 'value', however, is often unclear and the relationship between value and the payment design is not explicated. This paper aims at: (1) identifying value dimensions that are ideally stimulated by VBP and (2) constructing a framework of a theoretically preferred VBP design. Based on a synthesis of both theoretical and empirical studies on payment incentives, we conclude that VBP should consist of two components: a relatively large base payment that implicitly stimulates value and a relatively small payment that explicitly rewards measurable aspects of value (pay-for-performance). Being the largest component, the base payment design is essential, but often neglected when it comes to VBP reform. We explain that this base payment ideally (1) is paid to a multidisciplinary provider group (2) for a cohesive set of care activities for a predefined population, (3) is fixed, (4) is adjusted for the population's risk profile and (5) includes risk-mitigating measures. Finally, some important trade-offs in the practical operationalization of VBP are discussed.

## 1. INTRODUCTION

Worldwide, there is dissatisfaction with current, input-oriented, and supply-led health care systems. These systems are characterized by monodisciplinary and segmented care and result in fragmented care processes, suboptimal quality and waste of resources (Porter & Teisberg 2006; Berwick 2011; De Bakker et al. 2012; Mechanic & Tompkins 2012; Pronovost 2013; Tsiachristas 2015). There is consensus that flawed provider payment methods contribute to this problem (McGuire 2000; Porter & Teisberg 2006; McGuire 2011). In particular, predominant payment methods generate perverse incentives for health care providers regarding the delivery of services. For example, fee-for-service (FFS) – in which providers are paid retrospectively for each service delivered – is still a very common payment method in health care (especially in the United States) because it is relatively easy to administer and encourages productivity (Jegers et al. 2002; Marmor et al. 2011). However, this payment method may generate a 'more-is-better culture' and therefore tends to overprovision. In addition, providers who promote population health and successfully prevent treatment are financially penalized for that (Jegers et al. 2002; Ellis & Miller 2008). Another widespread payment method (especially in Europe) is capitation, in which providers receive a fixed amount per person per period. This payment method also has important drawbacks, such as encouraging underprovision and risk selection (Porter & Kaplan 2016). Furthermore, both FFS and capitation (as well as other predominant payment methods) do not reward the provision of high-quality care and innovation. Finally, because these methods traditionally remunerate single, monodisciplinary providers instead of multidisciplinary groups of providers, they preserve fragmentation and thwart cooperation and coordination across the continuum of care (Epping-Jordan et al. 2004; Van Exel et al. 2005). In short, predominant payment methods are not fully aligned with 'value'.

In order to tackle the problems related to current payment methods, worldwide, policymakers and purchasers of care are exploring alternative payment strategies to help steering health care systems towards value (Conrad et al. 2014; Burwell 2015). A well-known endeavor in this regard is pay-for-performance (P4P), in which providers are explicitly rewarded for 'doing a better job'. Although P4P is an appealing idea, explicit financial incentives for value should in principle be used only modestly in provider payment methods because of the multitasking problem (Holmstrom & Milgrom 1991; section 3.2).

Therefore, it is not surprising that in practice, the majority of provider revenues (typically referred to as the base payment) is not explicitly linked to value. This base payment, however, does create implicit (dis)incentives for value, because each payment method influences providers' behavior through incentives (Jensen & Meckling 1976; Enthoven 1988; Prendergast 1999; McGuire 2000; Gaynor et al. 2004; Berenson 2010; Christianson & Conrad 2011; McGuire 2011). In this paper, we underline the importance of carefully considering the design of particularly these implicit financial incentives, in such a manner that desired behavior is fostered and value is incentivized. We discuss a theoretically preferred design of a payment method that both

implicitly and explicitly stimulates value in a broad sense, henceforth referred to as value-based payment (VBP).

There is substantial literature on the theory and implementation of payment incentives (for an overview, see McGuire 2000; McGuire 2011; Conrad et al. 2014; Conrad 2015; Conrad et al. 2016). However, the theoretical basis of VBP design is fragmented and in the available work, the ters 'value' and 'VBP' are often implicitly used for different dimensions of value. In addition, the relationship between what a health care system ideally pursues in terms of value and what is required in terms of the VBP design to achieve this has not been explicated. Therefore, this paper aims at: (1) identifying key-value dimensions that are ideally stimulated by VBP and (2) constructing a conceptual framework of a theoretically preferred VBP design according to these dimensions. We achieve these goals based on a synthesis of findings of key theoretical and empirical studies conducted in the field of health services research, health economics, contract theory and the general economic theory on incentive design. Throughout, we relate our findings to VBP initiatives from practice, and end with illustrating some important trade-offs in the practical operationalization of VBP. The insights from this paper are of practical relevance for policymakers and purchasers who are responsible for (re)designing existing and future VBP initiatives.

The structure of this paper is as follows. In the next section, key-value dimensions are discussed, followed by a section containing a concise theoretical background on payment methods. The fourth section focusses on a theoretically preferred VBP design. Section 5 illustrates several important trade-offs in the practical operationalization of VBP, followed by some concluding remarks.

## 2. KEY-VALUE DIMENSIONS IN HEALTH CARE

In previous work, the term 'value' in health care has been defined in different ways. According to the Institute of Medicine (IOM 2001), health care needs to be safe, effective, patient-centered, timely, efficient, and equitable. Berwick et al. (2008) state that a health care system should pursue a Triple Aim of limiting per capita cost of care, improving individual patient experience, and improving population health. Porter (2009; 2010) provides a more global description of health care system goals, namely maximal value, defined as the best health outcomes achieved per dollar spent. Value encompasses efficiency and the central focus is on multidimensional outcomes, rather than inputs and processes. Conrad (2015) defines value as maximum health benefit (i.e., health outcomes, processes of care and patient experience) at minimum cost.

Based on these descriptions as well as arguments derived from the societal debate on what stakeholders in health care should ideally aim for (Eijkenaar & Schut 2015), five key-value dimensions can be distinguished:

1. High-quality care. Care is safe, effective, patient-centered, and timely. High quality comprises 'technical' or clinical quality as well as patient-reported measures regarding individual care

paths and outcomes (e.g., PROMs). Technical quality can be operationalized in structures (e.g., having an up-to-date patient registry for diabetes patients affiliated with the primary care practice), processes (e.g., regularly checking the blood glucose levels of diabetes patients) and (intermediate) outcomes (e.g., acceptable blood sugar levels for diabetes patients or absence of diabetes-related complications) (Donabedian 1988).

2. Cost-conscious behavior. Scarce resources are efficiently used, so there is no misuse or overuse.
3. Well-coordinated care. Multidisciplinary providers communicate and cooperate well in order to realize integrated, well-orchestrated care across the continuum of care. This dimension mainly regards coordination between providers of different disciplines and sites. A team-based approach in which multidisciplinary providers work side-by-side is of great importance, particularly given the increase in the number of individuals with multiple (chronic) diseases.
4. Cost-effective innovation. Cost-saving services result in equal or better health and health-promoting innovations are worth the additional costs.
5. Cost-effective prevention. Deteriorations of health problems are prevented in a cost-effective way. This dimension entails primary, secondary and tertiary prevention.

In this paper, a payment method is considered 'value-based' if it simultaneously provides incentives for all dimensions. Clearly, these dimensions are interrelated. For instance, well-coordinated care can be considered an element of high-quality care. However, for the purpose of describing a theoretically preferred VBP design, it is necessary to explicitly distinguish the different dimensions of value. Note, however, that it is not the goal of this paper to develop indicators for measuring value. As we will argue below (section 3.2), the measurement of all aspects of value and calculating payments only based on indicator scores is neither feasible nor desirable.

## 3. THEORETICAL BACKGROUND ON PROVIDER PAYMENT METHODS

### 3.1 Financial incentives to counterbalance agency problems

Agency theory, as part of contract theory, studies the relation between two contracting parties: the principal and the agent (Spence & Zeckhauser 1971; Ross 1973). In this paper, the focus is on the health care provider acting as a double agent, interacting with both the patient and the purchaser (Blomqvist 1991). Information asymmetry between providers as the relatively well informed party relative to patients and purchasers is not a problem, as long as the interests of all involved parties are aligned (Laffont & Martimort 2002). However, in case of conflicting interests, agency problems may evolve and providers may exploit their information surplus for their own (financial) benefit (Jensen & Meckling 1976; Richardson 1981).

An important strategy to counterbalance agency problems entails 'controlling' the agent by means of (financial) incentives (Vermaas 2006). The goal of controlling is to align providers' interests with those of patients' and 'purchasers' and is based on the assumption that providers

are in the position to improve value if they are motivated to do so. Providers' responsiveness to financial incentives has been well documented in the literature, implying that the (design of the) payment method is an important factor influencing providers' behavior and can thus be used to help steering health care systems towards value (Jensen & Meckling 1976; Enthoven 1988; Prendergast 1999; McGuire 2000; Gaynor et al. 2004; Berenson 2010; McGuire 2011).

### 3.2 The need for a base payment

Ideally, providers who are 'doing a good job' in terms of key-value dimensions are explicitly rewarded for this. A prerequisite of a payment method based fully on providers' performance with respect to value is that all aspects of value can be captured in the payment contract (i.e., for each aspect an indicator is available on which providers can be 'scored'). Complete contracts are, however, unfeasible in health care since the outcomes of some of the multiple tasks that providers perform, are more difficult (or even impossible) to measure objectively than others. For instance, for some medical interventions reliable and valid outcome indicators are available, whereas for other care activities – e.g., good communication and coordination of care – the added value is difficult to measure and appropriate registries are lacking. This problem has been referred to as the multitasking problem (Holmstrom and Milgrom 1991; Eggleston 2005; Frølich et al. 2007) and is defined as the challenge of designing incentives to motivate appropriate effort across multiple tasks when the desired outcomes for some tasks are more difficult to measure than others (Eggleston 2005). An important potential consequence of this challenge is that explicitly rewarding providers for some specific aspects of value may result in undesired behavior. Specifically, providers may focus disproportionately on those tasks that are measured and rewarded and neglect unincentivized tasks. This phenomenon has been referred to as 'teaching to the test' (Holmstrom & Milgrom 1991) and has actually been observed in practice (Steel et al. 2007; Glickman et al. 2007; Campbell et al. 2009; Mullen et al. 2010).

Due to the multitasking problem and the associated risk of teaching to the test, explicit financial incentives for value can and should be used only modestly in provider payment methods. As a consequence, the majority of providers' revenues can and should not be explicitly related to value. This part of providers' revenue is commonly referred to as the base payment. This base payment will typically comprise the largest part of total provider payment, whereas the payment component explicitly related to performance indicators (P4P) is likely to be relatively small. Indeed in practice, base payments currently comprise at least 90% of total provider payment (Eijkenaar 2013a; Ryan et al. 2015; Milstein & Schreyögg 2016). So far, papers investigating VBP reform have focused mainly on the design of the relatively small P4P component. Being the largest payment component, however, the design of the base payment is at least equally and arguably more important.

## 3.3 Shortcomings of predominant and alternative base payment methods

The four most frequently applied base payment methods in practice are payment per item-of service (FFS), payment per case (e.g., DRG's), payment per person (capitation) and payment per period (salary for individual providers and fixed budget for organizations). In Table 3.1, the incentives generated by these methods in relation to the key-value dimensions are summarized, based on Jegers et al. (2002), Ellis & Miller (2008) and Christianson & Conrad (2011). This table shows that, although each payment method to some extent stimulates at least one key-value dimension, other dimensions are not incentivized or even discouraged.

As none of the predominant base payment methods adequately promotes all key-value dimensions, alternative base payment methods have been developed. One example is combining predominant methods with opposing incentives in order to sustain the favorable elements of each method, while neutralizing the drawbacks (Ellis & McGuire 1986; Robinson 2001a; Christianson & Conrad 2011; McGuire 2011). Unfortunately, it is still unlikely that all value dimensions are stimulated under these mixtures (see Table 3.1, for a mixed payment method of 50% FFS and 50% capitation). Another recent example of an alternative base payment method is bundled or episode-based payment (De Brantes et al. 2009; Mechanic & Altman 2009; De Bakker et al. 2012; Ridgely et al. 2014). Although bundling stimulates cost-conscious behavior and well-coordinated care, value is only stimulated to some extent and only for those services inside the bundle (Wilensky 2014; Table 3.1).

**Table 3.1.** Base payment methods and their incentives for key value dimensions [a]

| | High-quality care | Cost-conscious behavior | Well-coordinated care [b] | Cost-effective innovation | Cost-effective prevention |
|---|---|---|---|---|---|
| Payment per item-of-service | +/- | - | - | - | - |
| Payment per case | +/- | +/- | - | +/- | - |
| Payment per person | - | + | - | - | + |
| Payment per period | - | +/- | - | - | +/- |
| Mixed payment method of 50% FFS and 50% capitation | +/- | + | - [b] | - | +/- |
| Bundled or episode-based payment | +/- | +/- | +/- | +/- | - |

[a] Authors' own analysis.

[b] By definition, no incentives for well-coordinated care exist because in these examples the payment is assumed to apply to a single, monodisciplinary provider.

## 4. A THEORETICALLY PREFERRED VBP DESIGN

### 4.1 Core components of a theoretically preferred VBP

Building on the theory as discussed in section 3, a theoretically preferred VBP should consist of two core components: (1) a substantial base payment that implicitly stimulates key-value dimensions and (2) a relatively small variable payment that explicitly rewards some measurable aspects of value dimensions (P4P). A base payment is a vital component of a theoretically preferred VBP because of the multitasking problem and the risk of teaching to the test when using high-powered explicit incentives (section 3.2). Nevertheless, relatively small explicit rewards are a crucial component of a theoretically preferred VBP. This payment component is required to ensure that value aspects that are not or cannot be implicitly incentivized by the base payment, are given sufficient attention by providers. The variable payment is particularly suitable for stimulating aspects of value that can be relatively easily and objectively measured and that are difficult to incentivize implicitly (Vlaanderen et al. 2019). Typically, these aspects are related to the value dimension 'high-quality care' since a broad spectrum of indicators has already been developed and is increasingly becoming available as a result of an increasing number of P4Pexperiments and initiatives employed by the International Consortium for Health Outcomes Measurement (ICHOM). Other measurable aspects of other value dimensions, however, can be part of the variable payment as well (e.g., smoking cessation counselling as an element of cost-effective prevention; Lindenauer et al. 2007; Mendelson et al. 2017).

The two components should be well tailored to ensure every value dimensions is implicitly and/or explicitly incentivized by VBP. The variable payment can be either designed as an 'add-on' to the base payment or as an integral part. The first modality is similar to most current P4P-programs, while in the latter modality receiving (part of) the base payment is conditional on meeting specific value targets. Note that the relative shares of the two components may vary over time and may depend heavily on the specific context (section 5). For instance, if better performance indicators become available, the share of the variable component that explicitly rewards high quality may increase relative to the base component.

In practice, there are several payment methods that come close to the theoretically preferred VBP design as described above. Box 1 provides a description of three prominent examples. In the remainder of this paper, we relate our findings to these examples.

**Box 1.** VBP practice initiatives

**Medicare Accountable Care Organisations (ACOs)**
ACOs are networks of healthcare providers that are jointly accountable for a share of the financial and clinical outcomes of a defined population during a predetermined period. Examples of public sector ACO models are the Medicare Pioneer ACO model and the Medicare Shared Savings Program (MSSP). Under the MSSP, a global budget based on the historical expenses of an assigned population of Medicare FFS beneficiaries is calculated. This 'benchmark' is corrected for national growth and is adjusted for population risk. Shared savings (and losses) are determined by comparing the benchmark to the ACO's actual expenditures and are conditional on meeting a minimum savings rate and quality standard. Assignment of the population to ACOs is mainly done retrospectively (Rose et al. 2016; Song 2014; Pham et al. 2010; McWilliams et al. 2015; Lewis et al. 2013).

**The Alternative Quality Contract (AQC)**
The Alternative Quality Contract is a five year ACO agreement in the private sector introduced by Blue Cross Blue Shield of Massachusetts (BCBS). Under the AQC, an annual fixed payment is provided, based on a per member per month amount. Providers are responsible for the total continuum of care for a defined population of enrolees that is prospectively attributed to a provider group by means of the affiliation of their designated primary care physician. The base payment is set using historical expenses and is adjusted periodically for (changes in) health risk. The base payment and future increases thereof (i.e., annual growth rates) are negotiated between provider groups and BCBS. Providers share both financial savings and losses. In addition to the global budget, providers who meet quality benchmarks are explicitly rewarded via the P4P-program (a bonus of maximal 10 per cent of the global budget). Shared savings and losses directly depend on the quality score as well; as quality improves (declines), the share of providers' deficit decreases (increases) while the share of providers' surplus increases (decreases). The base payment and the variable payment are thus highly integrated (Chernew et al. 2011; Song et al. 2012; Song 2014; Mechanic & Altman 2009).

**Gesundes Kinzigtal**
Gesundes Kinzigtal is a population-based integrated care approach in the Kinzigtal region, Germany. Providers are (financially) accountable for care across all health service sectors and indications (e.g., active health promotion for the elderly, disease management programs for chronic conditions, and patient university programs). The target population consists of all individuals who are insured by one of the two sickness funds in the region. Key to this initiative is the shared health gain approach by means of a shared savings contract (i.e., financial goals are realised if actual costs in the region increase at a lower rate than the German norm costs). The base payment is a global budget and equals the costs of the German risk-adjusted standard (i.e., the norm costs within the context of the risk-equalisation system). Quality is stimulated by means of a P4P-program (Hildebrandt et al. 2010; Hildebrandt et al. 2012; Busse & Stahl 2014).

Henceforth, we focus on the first component of a theoretically preferred VBP – the base payment – for two reasons. First, the design of the second component – the variable payment – has already been extensively discussed in the literature (for an overview, see Eijkenaar 2013a; Milstein & Schreyögg 2016). Second, as argued above, the base payment typically comprises the majority of providers' revenues, underlining the importance of carefully designing the implicit incentives generated by this component.

## 4.2 Five key features of a theoretically preferred base payment

Below, we explain which key features of a theoretically preferred base payment are required to stimulate value in a broad sense. Based on a synthesis of the findings of key studies conducted in the field of health services research, health economics, contract theory and the general economic theory on incentive design, we conclude that the base payment should preferably be paid (1) to a

multidisciplinary provider group for delivering (2) a cohesive set of care activities to a predefined population. In addition, the base payment should (3) be fixed, (4) be adjusted for the population's risk profile and (5) include risk-mitigating measures. We acknowledge that these five key features are interrelated (e.g., for the provision of a comprehensive set of care activities a multidisciplinary provider group is required).

### *4.2.1 Multidisciplinary provider group*

To encourage well-coordinated care, the base payment should jointly remunerate multidisciplinary groups of providers who have agreed to work together as an 'accountable group' for the delivery of a cohesive set of care activities. Depending on the exact nature of the care activities, these groups may consist of different types of physicians (e.g., primary care physicians or medical specialists), other health care professionals (e.g., nurses or physiotherapists) and various care facilities (e.g., specialty hospitals or rehabilitation centers).

Financial barriers between separately paid providers are removed once a single, integrated payment for a provider group is introduced. Such an integrated payment to a provider team is expected to encourage multidisciplinary cooperation and collaboration, fostering greater (cross-specialty) coordination and increasing active provider engagement in improvements across the whole care path (Anderson & Weller 1999; Berenson 2010; Burwell 2015; Mehrotra & Hussey 2015). This is of relevance particularly for the increasing number of individuals with multiple coexisting (chronic) health problems who will likely benefit from well-coordinated, integrated care (DeGruy & Etz 2010; Pollack et al. 2012; Leijten et al. 2017). In addition, paying a provider group instead of individual providers is likely to result in more flexibility in the use of resources (Mechanic & Altman 2009; Miller 2009; Cutler & Ghosh 2012; Tsiachristas et al. 2013). Another advantage is that the financial risk that is associated with VBP is pooled. This may prevent individual providers from being confronted with excessive financial risk and may reduce incentives for undesired behavior (Anderson & Weller 1999; Gaynor et al. 2004; Vermaas 2006; Frakt & Mayes 2012).

Group-based incentives require a certain entity that contracts with the purchaser and receives the payment on behalf of the provider group. This 'main contracting entity' administers the payment and is responsible for the organization, coordination and (possibly) the delivery of care activities and employs or subcontracts other providers (Anderson & Weller 1999). The main contractor thus initially bears the financial risk and has to divide the pain and gain among the group members. Entities such as ACOs, health maintenance organizations (HMOs) and hospitals might qualify for this role because of their size and level of professionalism. To pass the incentives along from the group to the individual providers, a transparent payment distribution mechanism needs to be developed; it has to be decided 'who is getting paid, how much, for doing what' (Frølich et al. 2007). For instance, distribution can be in proportion to the provider's share of the target population or the provider's contribution to the group's performance (Olson 1965; Gaynor et al. 2004; Conrad 2015).

### *4.2.2 Cohesive set of care activities for a predefined population*

To encourage cost-effective prevention, the base payment should remunerate a provider group for the provision of a cohesive set of (preventive) care activities to a predefined population of individuals. From a theoretical perspective, VBP ideally involves 'whole-person accountability'. Key to such an approach is that the payment is not disease-specific but person-centered and holistic. The payment covers all relevant health services given a person's needs. An evident set of care activities that is covered by the payment is (virtually) the full continuum of services included in the relevant benefit package in place. For instance, if a provider group accepts whole-person accountability for a target population of diabetics, the provider group is not only responsible for all diabetes-related care but for all care services that the diabetics in the target population might need, within limitations of the relevant benefit package covered by the health plan or other third party payer. The target population may consist of any defined set of individuals, including those not currently in need of care (Kindig 2007).

Whole-person accountability triggers incentives for health promotion and prevention because prevention is often more effective and cheaper than cure. The more a provider group improves the health of the population, the greater the financial gain (Sharfstein 2016). Stimulating preventive efforts is of great importance, since the causes of many health problems lie in individual behavior (e.g., smoking and unhealthy diet) and the current system does not effectively promote healthy behavior (Berwick et al. 2008; Casalino et al. 2015). Another advantage of a whole person accountability approach is that effective long-term management of chronic diseases (e.g., delaying the progression of diseases and preventing exacerbations) is stimulated (Berenson 2010; McClellan et al. 2013; Conrad 2015). In addition, cost-shifting becomes more difficult once the payment applies to a broad set of care activities and is even impossible if the payment applies to all care services (Sood 2011; Busse & Stahl 2014). The provision of unnecessary services is expected to gradually be phased out (Gaynor et al. 2004; McClellan et al. 2013). Finally, the risk of double payment for the same services decreases. Double payment is plausible in particular for patients with comorbidity and if services are paid for through different systems (Hussey et al. 2011; EIB 2012; Ridgely et al. 2014).

Four characteristics can be used to delineate the target population: (1) individual characteristics (e.g., age or diagnoses), (2) geographical catchment areas (e.g., region or ZIP-codes), (3) provider affiliation (e.g., enrolment in a GP practice or retrospective assignment to a provider based on actual utilization) and (4) purchaser affiliation (e.g., having an insurance policy with a specific insurer). The characteristics are not mutually exclusive. Under the AQC (Box 1), the target population consists of individuals who are below 65 years of age, live in Massachusetts, are registered with a primary health care provider, and have an HMO or preferred provider organization (PPO) insurance policy from BCBS (Mechanic & Altman 2009; Chernew et al. 2011; Song et al. 2012; Song 2014). Assignment of the target population to the provider group for the coming year can be done prospectively (e.g., based on enrolment with affiliated primary care physicians, or on health care utilization in the prior year) or retrospectively (e.g., based on the plurality of utiliza-

tion in the completed year). In case of prospective assignment, provider groups know beforehand for whom they are responsible in the coming year, enabling providers to proactively reach out to and improve care for their target population (Lewis et al. 2013). A potential advantage of retrospective assignment is that it stimulates providers to manage costs and quality for all of their patients, instead of just the assigned population. However, professional ethics may effectively prevent that – under prospective assignment – providers will actually distinguish between assigned and unassigned patients in terms of (the quality of) provided services. Under the AQC, assignment is done prospectively, while under the MSSP a retrospective form is used (Box 1).

### *4.2.3 Fixed payment for a defined period of time*

To encourage cost-conscious behavior and cost-effective innovation, the base payment should be fixed for a defined period of time, implying that there is no link with actual costs (Anderson & Weller 1999; Jegers et al. 2002). Such a method implies that (some of) the financial risk is transferred from the purchaser to the provider. The financial result is retrospectively determined by the difference between actual expenses and the prospectively defined, fixed payment ('reconciliation').

A fixed payment for a defined period of time is theoretically preferred over a variable payment because of the high potential for cost-conscious behavior and cost-effective innovation. Because marginal benefits are zero, providers are stimulated to reduce costs and to reconsider the full care process (Jegers et al. 2002; Miller 2009; Cutler & Gosh 2012; Conrad et al. 2014; Conrad 2015). Critically assessing care processes might also uncover room for substitution of relatively expensive for relatively inexpensive services or providers (Casalino 2001). In addition, because the payment can be flexibly deployed, more attention can be paid to cost-effective, creative management of care (Anderson & Weller 1999; McConnell et al. 2014). Note, however, that a fixed payment for a defined period of time also is a main feature of traditional capitation that was heavily criticized in the past for, amongst other things, triggering care rationing and threatening patient choice (Porter & Kaplan 2016). These drawbacks from traditional capitation can be addressed by adding a variable payment component guaranteeing high-quality care (section 4.1), by adopting adequate risk adjustment (section 4.2.4.) and by including arrangements to mitigate excessive financial risk (section 4.2.5).

Below, three design issues of a fixed payment for a defined period of time are discussed: setting the payment level, multiyear contracts and risk transfer.

*Setting the payment level*

In general, three methods for setting the fixed payment level can be discerned. A first method is based on historical expenses (Douven et al. 2015; Rose et al. 2016). An advantage of this approach is that calculation is relatively straightforward. However, because the payment level is based on prior expenses, past inefficiencies are 'buried' in the payment (Newhouse et al. 1997; Berenson 2010). Moreover, providers have a perverse incentive to increase expenses in the years

prior to the onset of the contract, in order to build up the historical expenses that lie at the basis of the payment level (Berenson 2010; Chernew et al. 2011; Douven et al. 2015). A second approach is basing the payment on average expenses, for instance per relevant peer group or region (Newhouse et al. 1997; Ellis & McGuire 1988). An advantage is that the payment is relatively easy to calculate and providers with higher than average expenses due to inefficiency are stimulated to reassess their delivery processes. However, providers with higher than average expenses as a result of a disproportionate amount of high-risk individuals in the target population are disadvantaged (Rose et al. 2016). In this case, the payment level can be considered as unfair and inaccurate, calling for appropriate risk adjustment (section 4.2.4). A third option is to base the payment on acceptable expenses (Newhouse et al. 1997). In this case, the payment is set at a level that is sufficient to cover only those expenses generated in delivering medically necessary, cost-effective care (Van de Ven & Ellis 2000). Although this approach seems theoretically preferred, it is difficult to implement in practice, since selecting the 'right' care activities and putting a price upon each service is disputable or likely to be unfeasible. Regardless of the chosen method for setting the payment level, the absolute price is clearly of relevance too. The payment should at least be sufficient to cover (potential) resource costs and to make the provision of high-value care worthwhile for providers.

*Multiyear contracts*

Contracts in which the fixed payment level is specified can be expected to be incomplete on a range of variables due to the multitasking problem (Maskin & Tirole 1999; Hart 2003). In the case of incomplete contracts, a certain level of mutual trust between the purchaser and the provider group is vital. Multiyear contracts are a sign of mutual trust and prevent costly effort on 'overwriting' complex, short-term contracts (Marques & Berg 2011). Microeconomic theory suggests that long-term contracts produce more favorable effects as compared to short-term contracts. A multiyear contract is likely to stimulate innovation and prevention because, over the longer term, providers are more likely to reap the financial benefits of their investments (Silberberg 1990; Christianson & Conrad 2011; Shortell 2013). On the other hand, providers and purchasers may also be hesitant to conclude multiyear payment contracts because of the concern about being locked into the contract. This calls for a certain level of flexibility in the contract to be able to adjust for inflation and unforeseen events (Chernew et al. 2011; Rose et al. 2016). In practice, multiyear VBP contracts have evolved, such as the five-year AQC contracts (Box 1).

*Risk transfer*

An important consequence of a fixed base payment for a defined period of time is that (some of) the financial risk is transferred from the purchaser to the accountable provider group. Two types of risk may be transferred: insurance risk and performance risk. Insurance risk is the risk that is typically borne by the purchaser and concerns the random variation around the mean health care expenses. Performance risk is the systematic variation around the mean expected health care

expenses due to providers acting as imperfect agents. This risk can be influenced by providers, as it directly relates to the clinical skills and the choices made by the provider (Vermaas 2006; De Brantes & Rastogi 2008; Miller 2009; Berenson 2010).

Ideally, only performance risk is transferred to the provider group, whereas insurance risk remains with the purchaser (Porter & Kaplan 2016). After all, it is the typical function of a purchaser to deal with random variation by pooling risks, and transforming providers into insurers is not the goal of VBP. Because the target population of a provider group is likely to be smaller than the total number of individuals the purchaser is responsible for, the conditions of the law of the large numbers for effective risk pooling might not be sufficiently fulfilled. Therefore, the provider group might face substantial financial risk due to large random variation from the statistically expected result (Christianson & Conrad 2011; Van de Ven 2014). In comparison to purchasers, providers have limited financial means at their disposal to compensate for this random variation. Transferring insurance risk to providers could encourage risk selection (section 4.2.4) and, in extremis, providers might go bankrupt (Anderson & Weller 1999; Vermaas 2006).

Unfortunately, it is practically unfeasible to split insurance risk and performance risk (Vermaas 2006). Often, unravelling the extent to which health outcomes are the result of chance or of providers acting as (im)perfect agents is virtually impossible. For instance, a lower incidence of diabetes-related health problems in the target population could be the result of a decrease of the number of individuals with obesity due to a successful government campaign to improve lifestyle but could also stem from a provider's successful effort in monitoring blood glucose levels. The first explanation is not necessarily linked to the provider's performance, while the second cause refers to the provider acting as a good agent. While risk-splitting is thus not possible, distributing the financial risk among providers and the purchaser in such a way that providers bear some, but not all, of the risk may be a viable option (Frakt & Mayes 2012).

### *4.2.4 Risk adjustment*

To prevent undesired behavior that may thwart key-value dimensions, the base payment should be risk adjusted. If the payment is not corrected for systematic variation in expenses due to differences in risk characteristics of the target population, incentives for risk selection evolve because then the financial result is partly determined by the risk composition of the population, rather than a mark of achievement. Providers would be unfairly penalized financially if they are responsible for a disproportionate amount of high-risk individuals rather than low-risk individuals. In this case, providers have a financial incentive for risk selection which is the practice of attracting low-risk individuals for which the payment exceeds expected expenses and/or avoiding high-risk individuals for which the opposite holds (Iezzoni 2003; Sood 2011; Rose et al. 2016). Risk selection is undesired because it may jeopardize quality, equal access and efficiency (Welch 1999; Jegers et al. 2002; Barros 2003). Several empirical studies provide evidence of risk selection by capitated provider groups (Newhouse & Byrne 1988; Frank & Lave 1989; Newhouse 1989;

Cutler & Zeckhauser 1998; Altman et al. 2000; Dranove et al. 2003; Chang et al. 2012; Hsieh et al. 2016).

In case of fixed payments, provider groups may experience incentives for risk selection. Because of the relatively small size of target populations a small number of high-risk individuals may have a large impact on the global budget. Providers are in the position to be successful in risk selection. First, providers are particularly well equipped to effectively identify low-risk and high-risk individuals because they have information about the health status of their target population, and they are professionally trained to assess this type of information. Second, providers have subtle tools for risk selection. For instance, a provider might advise a high-risk patient to switch to a different provider by suggesting that he or she would be better served elsewhere (Folland et al. 2013). Non-financial restraints, such as peer review and professional ethics, may however counteract incentives for providers to engage in risk selection (Eggleston 2000). With risk adjustment predictable, systematic variation in expenses as a result of differences in risk characteristics of the population is recognized and accounted for. In this way, risk adjustment contributes to a fair allocation of payments and ensures that providers are willing to accept and serve high-risk individuals. Ideally, risk adjustment creates a level playing field for providers (Anderson & Weller 1999; Iezzoni 2003; McGuire 2011; Ash & Ellis 2012; Omachi et al. 2013; Brilleman et al. 2014; Rose et al. 2016). In Gesundes Kinzigtal, the base payment equals the normative cost-level calculated using the German risk-adjustment model for health insurers, and ACO and AQC models use population risk-score software to adjust for differences in risk characteristics of the target population (Box 1). It is an interesting question to what extent existing risk-adjustment models – most of which were originally developed to adjust capitation payments for insurers – can be (adequately) used to adjust provider payments, taking account of differences between provider and insurer payment regarding incentives and tools for risk selection.

#### *4.2.5 Arrangements to limit excessive financial risk*

To prevent undesired behavior that may thwart key-value dimensions, the base payment should include arrangements that effectively mitigate excessive financial risk for providers. As discussed before, providers accepting VBP share financial risk with the purchaser. Risk adjustment accounts for systematic, predictable variation in expenses. However, the majority of between-person variance is random and unpredictable (Van Vliet 1992; Newhouse 1996; Ellis & McGuire 2007). This implies that, even in the unlikely case of perfect risk adjustment, providers still face significant residual financial risk. To protect providers against excessive financial risk, additional approaches to mitigate this risk are likely to be required. In principle, these arrangements are focused on protecting providers against large, unpredictable, random losses (i.e., insurance risk). However, such arrangements could also include protection against predictable and systematic risk that is, for whatever reason, not corrected for by a risk-adjustment model. Note that risk-mitigating arrangements could be used not only to limit but also to (gradually) expand the financial risk a

provider runs. Below, we elaborate on two main parameters that can be simultaneously used to bring the financial risk to appropriate levels.

*Type of risk sharing*

Two main types of risk contracts can be distinguished. Under a one-sided risk contract, providers that keep expenses below the global payment share in the savings with the purchaser. An advantage is that providers can get familiar to accepting financial 'risk' without sharing in the losses and, keeping all else constant, have less incentives for undesired behavior such as risk selection (Berwick 2011). Under a two-sided risk contract, providers share in the savings, but also in the losses if expenses exceed the global budget. Providers accepting two-sided risk qualify for higher shared savings rates (Berwick 2011; Rose et al. 2016). Theoretical and empirical evidence from the field of behavioral economics has shown that individuals tend to prefer avoiding losses to achieving equivalent gains (Kahneman & Tversky 1979; McNeil et al. 1982), suggesting that a two-sided risk contract provides stronger incentives for value than a contract that includes rewards only (Berenson 2010). However, incentives for undesired behavior increase under a two-sided risk contract (assuming imperfect risk adjustment). In the MSSP, ACOs can opt for a one-sided or a two-sided risk contract while they are in their first two contract periods. After this period, they can only accept a two-sided risk contract (Berwick 2011; Rose et al. 2016). In addition to one- and two-sided risk contracts, risk corridors and reinsurance can be used to bring financial risk to the appropriate level. Risk corridors protect against cumulative losses, because losses and gains are limited beyond a predefined acceptable range (Layton et al. 2016). Reinsurance can be defined as "the insurance of contractual liabilities incurred under contracts of direct insurance or reinsurance" (Carter 1983:4). In the case of VBP, reinsurance would imply that providers are retrospectively reimbursed by the purchaser for some or all of the expenses of specific individuals from their population, based on prospectively determined conditions. Under the AQC, for example providers can buy reinsurance from BCBS or an external entity (Chernew et al. 2011). A variety of non-mutually exclusive reinsurance techniques exist, such as stop-loss contracts, proportional risk sharing and outlier risk sharing (e.g., Carter 1983; Von Eije 1989; Bovbjerg 1992; Van Barneveld et al. 1998; Anderson & Weller 1999; Vermaas 2006; Miller 2009).

*Extent of risk sharing*

Under VBP, the main contractor (i.e., provider group) shares the financial risk with the purchaser. Thus, the provider group is typically liable for less than 100% of the financial result (Vermaas 2006; Frakt & Mayes 2012). The risk rate (i.e., the share of savings/ losses the provider group is accountable for) should not be set too high in order to keep the risk manageable for the provider group and to prevent (strong) incentives for risk selection in the case of imperfect risk adjustment. However, this rate should not be set too low either, because then the incentives lack power to actually affect provider's behavior (Laffont & Tirole 1993; Gaynor et al. 2004).

The risk rate ideally depends on several variables. A first factor concerns the size of the target population. Ceteris paribus, if the size of the population increases, the payment is expected to gain in stability due to the law of the large numbers, allowing higher risk rates. Second, it seems natural to increase the risk rate for primary relative to secondary care if a primary care group acts as main contractor, while the opposite may be preferred if a hospital accepts this role. Third, the diminishing marginal utility of income might be taken into account (Conrad & Perry 2009). Under the AQC, the risk borne by the different groups of providers ranges from 50% to 100% and is periodically (re)negotiated between the provider group and BCBS (Chernew et al. 2011).

In addition to the risk rate, carve-outs can be used to influence the extent of risk sharing. Carve-outs mitigate the financial risk for providers by placing a portion of the risk outside the payment and contracting separately for this risk (Frank & McGuire 1998). For VBP, this would imply that certain services, medical conditions, or populations are excluded from the contract and are paid for on a separate basis, such as FFS. Consequently, providers are protected against the associated high expenses of these services, conditions or populations and the high costs that are associated with acquiring the needed expertise (Frank & McGuire 1998). Examples of possible carve-outs are intensive care, organ transplantation, mental health or cancer care. Carve-outs may also be required if whole-person accountability is not (instantly) feasible from a practical point of view or for those care services for which risk adjustment is not or insufficiently attainable; carve-outs can be used as an interim measure to (temporarily) exclude certain care types from the payment.

## 5. TRADE-OFFS IN THE OPERATIONALIZATION OF THE BASE PAYMENT

In section 4, a theoretically preferred VBP design was discussed. We explained how the largest component of VBP – the base payment – should preferably be designed to incentivize value. When it comes to the practical operationalization of the base payment, several inherent trade-offs arise, implying that no 'one size fits all' design exists that can optimally incentivize all key-value dimensions simultaneously. The practical operationalization of the base payment and the extent to which the different value dimensions are incentivized, depend on three determinants: (1) compatibility of incentives, (2) preferences and (3) context. Below, these determinants are briefly discussed and illustrated.

### 5.1 Compatibility of incentives

Theory predicts that several key features of the base payment are likely to conflict to a certain extent. For instance, regarding the optimal composition and size of the provider group, stronger incentives for well-coordinated care must be traded-off against weaker incentives for cost-conscious behavior. In order to be able to deliver (virtually) the full continuum of care and realize well-coordinated care, the provider group will have to be composed of many different types of providers. But, as the composition becomes more diverse, the size of the provider group is likely

to increase as well. Consequently, the financial risk that is associated with VBP is necessarily spread over more providers within the group, reducing the financial incentives for individual providers and increasing incentives for free-rider behavior (Gaynor & Gertler 1995; Gaynor et al. 2004; Town et al. 2004; Conrad 2015).

Another example of a practical decision involving trade-offs is about the comprehensiveness of the set of activities a provider group is responsible for. If the payment covers a broader set of care activities, the payment moves towards 'whole-person accountability' and incentives for health promotion and (primary) prevention become stronger. However, given that perfect risk adjustment is practically unfeasible, a more comprehensive set of activities will also increase the incentives for risk selection. Hence, stronger incentives for cost-effective prevention should be weighed against stronger incentives for risk selection.

### 5.2 Preferences

In trading-off the different value dimensions, decision-makers should carefully weigh preferences for each dimension, taking full account of relevant (societal) interests. For instance, if in a country health care expenses are considered to be at an acceptable level, while quality is considered to be suboptimal, decision-makers may attach greater importance to incentives for high-quality care and compromise on the incentives for cost-conscious behavior (under the assumption that higher quality is associated with higher expenses). In this case, the share of the variable payment may be expanded, whereas the financial risk for providers may be reduced. Alternatively, a country with escalating health care expenses and an inefficient health care system may choose to intensify incentives for cost-conscious behavior by expanding the financial risk for providers, while accepting the possible negative consequences in terms of stronger incentives for risk selection.

### 5.3 Context

The context of implementation can have a major impact on the practical operationalization of the base payment, implying that VBP should be structured in relation to the circumstances of time and place (Conrad et al. 2016). The following four examples illustrate this. First, if limited individual-level data on population risk characteristics are routinely available, a base payment that requires sophisticated risk adjustment is unlikely to be practically feasible. Second, in a setting where providers still predominantly work in monodisciplinary 'silos', it might be problematic to find provider groups that are willing and able to accept whole-person accountability. Third, in a setting in which the IT-infrastructure is underdeveloped, it is unlikely that a multidisciplinary provider group is effectively able to share the information required to realize well-orchestrated, integrated care for the target population (Miller 2009; Berwick 2011). Fourth, expanding the size and scope of providers groups covered by VBP may also affect market concentration and therefore may reduce consumer choice and workable competition. Therefore, in countries with a competitive health care system, the optimal size and scope of provider groups covered by VBP may be smaller than in countries with a more centralized health care system.

## 6. CONCLUDING REMARKS

This paper has provided a conceptual framework of key components and design features of a theoretically preferred VBP method. We consider a provider payment method 'value-based' if it stimulates value in a broad sense, that is if it offers incentives for: (1) high-quality care, (2) cost-conscious behavior, (3) well-coordinated care, (4) cost-effective innovation and (5) cost-effective prevention.

To our knowledge, this is the first paper in the provider payment literature with a prime focus on the design of such a VBP method, and in particular of arguably the most important component thereof: the base payment. Based on a synthesis of existing literature from a variety of fields, this paper provides insight in the contours of a theoretically preferred VBP method.

The main contribution of this paper is twofold. Inspired by the societal debate on what stakeholders in health care should ideally strive for, as well as by existing definitions of value, we first described and further specified the concept of value, facilitating the specification of requirements in the design of VBP. We conclude that, in this respect, value is ideally conceptualized as a multifaceted concept, comprising not only high quality of care at the lowest possible costs but also efficient cooperation, innovation and health promotion. Second, starting from these value dimensions, we derived various design features of a theoretically preferred VBP model. We conclude that in order to stimulate value in a broad sense, the payment should consist of two main components that must be carefully designed. The first component is a risk-adjusted global base payment with risk-sharing elements paid to a multidisciplinary provider group for the provision of (virtually) the full continuum of care to a certain population. The second component is a relatively low-powered variable payment that explicitly rewards aspects of value that can be adequately measured.

Although a well-designed VBP is clearly a necessary condition for realizing value-based health care, we acknowledge that it is unlikely to be a sufficient condition. Non-financial mechanisms as well as organizational structures may be at least as important (Robinson 2001a; Christianson & Conrad 2011; Phipps-Taylor & Shortell 2016). Furthermore, as explained above, the practical operationalization and implementation of a well-designed VBP model should not be underestimated and be well tailored to the specific context. Nevertheless, several innovative payment initiatives in practice already come quite close to the theoretically preferred VBP-design described in this paper, indicating that this design can actually be implemented in various contexts. An interesting direction for future research would be gaining more insight in how a two-component model as described in this paper can be practically operationalized and successfully implemented given the relevant context, as well as in the short- and long-term effects of introducing such a model on different value dimensions.

# Chapter 4

## Value-based provider payment initiatives combining global payments with explicit quality incentives: A systematic review

With Frank Eijkenaar

Medical Care Research and Review, 2020, 77(6): 511-5237

## ABSTRACT

An essential element in the pursuit of value-based health care is provider payment reform. This article aims to identify and analyze payment initiatives comprising a specific manifestation of value-based payment reform that can be expected to contribute to value in a broad sense: (a) global base payments combined with (b) explicit quality incentives. We conducted a systematic review of the literature, consulting four scientific bibliographic databases, reference lists, the Internet, and experts. We included and compared 18 initiatives described in 111 articles/documents on key design features and impact on value. The initiatives are heterogeneous regarding the operationalization of the two payment components and associated design features. Main commonalities between initiatives are a strong emphasis on primary care, the use of 'virtual' spending targets, and the application of risk adjustment and other risk-mitigating measures. Evaluated initiatives generally show promising results in terms of lower spending growth with equal or improved quality.

# 1. INTRODUCTION

## 1.1 Background

Worldwide, the interest in value-based health care (VBHC) is growing rapidly. In many developed countries there is public recognition that waste and inefficiency can be reduced, while quality and health outcomes can be improved (Berwick & Hackbarth 2012). Encouraging health care providers to deliver high-value care is thus a focal point in health policy.

An essential element in the pursuit of VBHC is provider payment reform. The reason for this is twofold. First, financial incentives in general, convincingly show to substantially influence provider behavior (Gaynor et al. 2004; McGuire 2000 and 2011; Robinson 2001a). For example, physicians paid on a fee-for-service (FFS) basis, tend to provide more care compared with capitated and salaried physicians (Gosden et al. 2000). Second, predominant payment methods – in particular FFS – are not well aligned with value (Christianson & Conrad 2011; Ellis & Miller 2008; Jegers et al. 2002; Robinson 2001a). Specifically, paying providers separately and per activity encourages overprovision, maintains fragmentation, discourages prevention, and does not stimulate high-quality care. Since working toward VBHC, while leaving financial incentives for low-value care intact would clearly be counterproductive, there is consensus that VBHC and payment reform should go hand-in-hand.

Over the past decade, there has been much experimentation with various types of value-based payment (VBP) models. In this regard, both 'value' and 'VBP' are defined and operationalized in different ways. According to Berwick et al. (2008), high-value care requires pursuit of the 'triple aim': limiting per capita cost of care, improving individual patient experience, and improving population health. Porter (2009 and 2010) provides a more general description of value, namely, the best health outcomes achieved per dollar spent. Conrad (2015) defines value as maximum health benefit (operationalized as health outcomes, processes of care, and patient experience) at minimum cost. A commonality in these definitions is that value is considered a multidimensional concept, comprising not only high quality and integration of care but also cost consciousness and good health outcomes, which in turn require prevention.

Regarding VBP reform, emphasis is primarily on developing and implementing bundled-payment models for specific conditions or treatments as well as pay-for-performance (P4P) models that explicitly reward specific, measurable aspects of value (Chee et al. 2016; Roland & Campbell 2014; Ryan et al. 2017). Examples of the former are the Bundled Payment for Care Improvement Initiative and the Acute Care Episode Demonstration, both implemented in U.S. Medicare. Examples of the latter are the Hospital Value-Based Purchasing Program in U.S. acute care hospitals and the Quality and Outcomes Framework in the U.K. primary care sector. Although bundled payment and P4P could contribute to improvement of specific value dimensions, other important dimensions are unlikely to be strongly affected. Bundled payment mainly stimulates cost-conscious behavior and coordination, regarding the services pertaining to the condition or treatment in question (Stokes et al. 2018). P4P, by design, only focuses on

aspects of value that can be explicitly measured using indicators, which are typically aspects of clinical quality. In other words, both types of VBP adopt a relatively narrow definition of value and are not well-suited for simultaneously incentivizing the multiple value dimensions as defined in the literature.

If payment reform is to substantially contribute to value in a broad sense, more profound reform of current payment models is likely to be required. Indeed, there is growing recognition in the literature as well as in practice that VBP models be designed in such a manner that incentives for high-value care stretch beyond the level of conditions or treatments. In addition, these models should not only stimulate measurable aspects of high-quality care but also cost-conscious behavior, well-coordinated care, and prevention (Peikes et al. 2018; Quentin et al. 2018; Scott et al. 2018). Arguably, this can be realized by combining two payment components: (a) global base payments and (b) explicit quality incentives (Cattel et al. 2020a; see section 2.1 for a justification). Over the past years, payment reform initiatives adopting these two components have been gaining ground, for example, in the shape of accountable care organizations (ACOs). To date, however, these initiatives have not been systematically identified and described.

### 1.2 New contribution

Prior literature reviews investigating VBP reform mainly focused on bundled payment and P4P initiatives, which adopt a relatively narrow definition of value (Conrad et al. 2014; Mendelson et al. 2017; Milstein & Schreyögg 2016; Scott et al. 2018). A comprehensive overview of VBP initiatives aiming at improving value in a broad sense via global base payments combined with explicit quality incentives is lacking. Currently, it is unclear how these initiatives are being designed and to what extent they are effective in improving value. In this article, we aim to fill this gap by systematically identifying and analyzing VBP initiatives comprising these two payment components. Specifically, we (a) describe the design features of these initiatives and (b) assess the extent to which initiatives have been successful in improving value. In doing so, we aim to provide policy makers, payers, and health care providers insight in promising and practically feasible modalities of VBP reform. In turn, this could support additional innovation, facilitate future model comparison, and ultimately contribute to VBHC. The integration of non-U.S. initiatives is especially valuable to stimulate international comparisons and shared learning.

This article proceeds as follows. The next section presents a framework of a VBP model comprising global base payments and explicit quality incentives, which will be used to systematically describe and compare identified initiatives. Section 3 elaborates on the strategy followed while conducting this systematic literature review, and section 4 presents the results. The last section reflects on the main findings and provides an overall conclusion.

## 2. CONCEPTUAL FRAMEWORK

Recent papers have attempted to explicate the relationship between what a health care system ideally pursues in terms of value and what is required in terms of the design of provider payment systems (e.g., Cattel et al. 2020a; Eijkenaar 2013a; Scott et al. 2018). After reviewing existing descriptions of value and arguments used in the societal debate on what stakeholders in health care ideally aim for, we conclude that value is a multidimensional concept. The commonality in all descriptions is that value encompasses not only high-quality care, but also multidisciplinary coordination, cost-conscious behavior, and prevention (Berwick et al. 2008; Conrad 2015; Donabedian 1988; Eijkenaar & Schut 2015; IOM 2001; Porter 2009 and 2010; Stokes et al. 2018). Based on a comprehensive synthesis of the payment incentive literature, Cattel et al. (2020a) conclude that a combination of global base payments with explicit quality incentives seems well-suited to stimulate all these value dimensions simultaneously. The next section briefly elaborates on the rationale of such a two-component model.

### 2.1 The rationale of global base payments in combination with explicit quality incentives

The first component of a VBP model that stimulates value in a broad sense is a substantial global base payment. In essence, global payments are a form of bundled payment, with the bundle being constructed at a higher level than at the level of conditions or treatments. This addresses the shortcomings of lower level forms of bundled payment mentioned in the Introduction. The second component is a relatively low-powered P4P payment that explicitly rewards some measurable aspects of value.

Any provider payment system will at least consist of a base component that is not directly linked to providers' measured performance. The reason is that many aspects of value, such as well-coordinated care and many health outcomes, are difficult or impossible to measure and attribute. While important, these aspects can thus not 'explicitly' be accounted for in the payment contract (Eggleston 2005; Holmstrom & Milgrom 1991). The base payment can be designed in such a manner that it 'implicitly' incentivizes aspects of value that cannot be adequately measured and thus not stimulated through explicit incentives (section 2.2). Designing the base payment as a global payment facilitates cost-consciousness and well-coordinated care across the full continuum of care, with a focus on whole persons instead of on separate conditions or treatments.

Global base payments transfer financial risk from payer to provider. A possible danger is that providers become exposed to too much financial risk. As a result, they may be inclined to skimp on quality or act too aggressively in attempts to reduce spending by underproviding necessary but expensive services. These concerns, which are not just theoretical (Frakt & Mayes 2012; Robinson 2001a), can be mitigated by supplementing the global base payment with risk-sharing arrangements and explicit quality incentives. Risk sharing results in a situation in which providers are being held accountable for only a share of savings/losses realized under the global base

payment. Explicit quality incentives may trigger providers to give sufficient attention to value aspects that are unlikely to be incentivized by the global base payment but may be prone to quality skimping or underprovision (Eijkenaar 2013b). These incentives should be relatively low-powered to prevent a disproportionate focus on rewarded tasks (Campbell et al. 2009; Mullen et al. 2010; Steel et al. 2007). In addition, high-powered explicit incentives may have a negative effect on physicians' intrinsic motivation (Eijkenaar 2013b; Wynia 2009).

Empirical work supports the theoretical rationale of a two-component VBP model. Vlaanderen et al. (2019), for example, conclude that using explicit incentives for (outcome) quality paired with global base payments seems preferred over using explicit quality incentives alone.

## 2.2 Design of global base payments and explicit quality incentives

In this review, we analyze VBP initiatives combining global base payments with explicit quality incentives in terms of design and impact on value. For this purpose, we use two existing conceptual frameworks: one for the global base payment (Cattel et al. 2020a) and one for the explicit quality incentives (Eijkenaar 2013a). Although other frameworks made important contributions to the VBP literature, they are not suited for thoroughly describing and comparing key design features of payment models adopting the two-component structure described above. Shortell et al. (2014), for example, established a taxonomy to classify and understand early ACOs using eight general attributes that are not all related to payment design. In another article, Stokes et al. (2018) proposed a typology of payment models for integrated care. Since the focus of that article is specifically on incentives and facilitators for integrated care, it is also not suitable for the purpose of our review.

Table 4.1 summarizes design features and issues regarding both payment components, which we briefly discuss below. First, providing the global base payment to a multidisciplinary provider group fosters coordination across the continuum of care (Anderson & Weller 1999; Berenson 2010; Burwell 2015; Mehrotra & Hussey 2015). Financial barriers between providers and sites are removed, resulting in more flexibility in the resource deployment (Cutler & Ghosh 2012; Mechanic & Altman 2009; Miller 2009). Generally, a main contractor is responsible for administering and distributing the payment and employing and/or subcontracting individual providers (Anderson & Weller 1999).

**Table 4.1.** Core components and associated design features of a value-based payment model combining global base payments with explicit quality incentives

| | |
|---|---|
| **Core component 1: Global base payment** | |
| To a multidisciplinary group | Which provider type included?<br>Who is main contractor?<br>Group members employed or subcontracted? |
| For a cohesive set of care activities to a predefined population | What care services to include?<br>How to delineate the population?<br>How to attribute patients to provider group? |
| Fixed for a defined period of time | Is payment real of virtual?<br>How to set the payment/target?<br>What is the contract duration? |
| Risk adjusted | Is risk adjustment applied?<br>Which risk adjustors to use? |
| Risk-mitigating measures | One-sided or two-sided risk?<br>What is the risk-sharing rate?<br>Include reinsurance provisions?<br>What care to carve-out? |
| **Core component 2: Explicit quality incentives** | |
| Method of linking the payment to quality | Shared savings/losses conditional on quality?<br>Add-on for quality? |
| Quality measurement | Which indicators to use?<br>What measurement level (individual/group)? |
| Quality incentive structure | Rewards and/or penalties?<br>Maximum payment size relative to total payment?<br>Absolute, relative and/or improvement targets?<br>Payment frequency? |

Note. Based on Cattel et al. (2020a) and Eijkenaar (2013a).

Second, a global base payment pertains to a comprehensive set of care services for a predefined population of individuals. By adopting a person-based rather than a condition-based approach, incentives for prevention and cost-conscious behavior are strengthened. Another advantage is that cost-shifting becomes more difficult and is even impossible if the payment applies to the full continuum of care (Busse & Stahl 2014; Hussey et al. 2011; Ridgely et al. 2014). The population can be delineated in various ways, for example, based on provider and/ or payer affiliation. Attribution of this population to the provider group can be done prospectively or retrospectively.

Third, providing a payment that is fixed for a defined period of time stimulates cost-conscious behavior because it transfers financial risk to providers (Conrad 2015; Frakt & Mayes 2012; Jegers et al. 2002; Miller 2009; Robinson 2001a). The payment can be determined in various ways, including based on historical spending and on average per capita spending in the region. The payment can be implemented as a 'real' payment that actually replaces existing payment systems or as a 'virtual' spending target with end-of-period reconciliation with claims. Regarding the contract period, in principle multiyear contracts seem preferable over short-term contracts because they provide room for earning back investments in value improvement. In addition,

multiyear contracts signal mutual trust and prevent costly effort on 'overwriting' complex, short-term contracts (Christianson & Conrad 2011; Marques & Berg 2011; Shortell 2013; Silberberg 1990). In practice, however, multiyear contracting could be difficult, especially in settings with high rates of beneficiary 'churn'.

Finally, to realize better effects on the different value dimensions, theory recommends risk adjusting the base payment and applying risk-mitigating measures. Risk adjustment prevents providers from being unfairly penalized for caring for a disproportionate share of high-risk individuals and from being incentivized to select favorable risks (Iezzoni 2003; Rose et al. 2016). Adopting risk-mitigating measures protects providers against excessive financial risk due to large random shocks in spending. Several options are available to bring financial risk to appropriate levels, including using one- or two-sided risk contracts (i.e., sharing upside risk only or also downside risk), varying the risk-sharing rate, adding reinsurance provisions, and carving out specific high-cost services from the contract.

The second component of a two-component VBP model is a payment explicitly linked to quality. Three main design features are of relevance: the method used to link payment to quality, quality measurement, and the quality incentive structure (Eijkenaar 2013a). Regarding the method for linking payment to quality, the payment can either be applied as 'add-on' to the global base payment or the provider share of realized savings/losses can be made conditional on aggregated quality scores. Regarding quality measurement, indicators could reflect 'technical' quality (structures, processes, and outcomes) and/or patient-reported quality. Finally, the incentive structure concerns choices with regard to rewards versus penalties, incentive size relative to the total payment, type of quality targets, and payment frequency. Although each choice has advantages and disadvantages, prior literature suggests that using relatively low-powered rewards (Deci et al. 1999; Eijkenaar 2013a; Holmstrom & Milgrom 1991; Moscucci et al. 2005; Shen 2003), limiting the time lag between care delivery and payment (Conrad & Perry 2009; Frederick et al. 2002; Thaler 1981), and using absolute quality targets (Conrad & Perry 2009; Rosenthal & Dudley 2007; Young et al. 2007) is most likely to be effective in stimulating desired behavior.[11]

11 The VBP model as described in this section shows similarities with the global capitation payment model traditionally used by Health Maintenance Organizations (HMOs). In both models, provider groups receive a fixed payment for the provision of a comprehensive set of care activities for a predefined population, with the goal to increase efficiency by shifting financial risk to providers (Frakt & Mayes 2012). However, both models differ in two important respects, specifically meant to address the concerns that were often raised against HMOs and global capitation: underprovision and quality skimping (section 'The rationale of global base payments in combination with explicit quality incentives'; Frakt & Mayes 2012). First, under VBP, providers and payer share financial risk, while HMOs typically use full capitation models that involve much more financial risk for providers. Second, under VBP, total compensation is partly dependent on quality performance, while in HMOs this was often not the case or only to a relatively limited extent (Frakt & Mayes 2012). Thus, the VBP model takes advantage of the benefits of traditional capitation, while trying to avert its main disadvantages.

## 3. METHOD

### 3.1 Search strategy and selection procedure

Complying with the Cochrane Handbook for systematic reviews (Higgins & Green 2011), we conducted a systematic review of the literature on VBP initiatives written in English or Dutch and published between January 2000 and April 2017. We included articles/documents describing VBP initiatives that

1. have been implemented in developed countries;
2. combine global base payments with explicit quality incentives;
3. involve payments to multidisciplinary provider groups; and
4. involve payment for the provision of cohesive sets of care activities to predefined populations.

Consequently, we excluded initiatives that have not been implemented as well as initiatives that have adopted payment models without clearly discernable global base payments and/or explicit quality incentives, that are targeted at individual providers, and/or that are organized around specific conditions or treatments.

We mainly focused on articles published in peer-reviewed scientific journals. However, we did not exclude unpublished studies, reports, or policy briefs beforehand, because they may still describe initiatives meeting our inclusion criteria. Our main focus was on articles/documents *describing* VBP initiatives; the absence of a quantitative evaluation was not an exclusion criterion. Insofar available, however, we included studies describing quantitative effects on value, but only if published in peer-reviewed scientific journals and if the research approach corresponds to a difference-in-differences, interrupted-time series, randomized controlled trial, or systematic review design.

In identifying eligible VBP initiatives, we consulted four sources: (a) scientific bibliographic databases, (b) reference lists, (c) the Internet, and (d) experts publishing in the field of VBHC and/or VBP. We started our review by searching four bibliographic databases on April 12, 2017: Embase, Medline, Web of Science, and Cochrane Central. We used the same search terms for each database, while taking into account database-specific requirements (see Appendix A). In consultation with an information specialist of the library of the Erasmus Medical Center in Rotterdam, we developed the search strings using a combination of the terms *value-based payment* and *care provider*. After removal of duplicates, we independently screened the titles and abstracts of all articles yielded by the search and assessed their potential eligibility for inclusion. We compared initially included articles and resolved discrepancies by discussion. In a second round of screening, the first author retrieved full texts and assessed each article on eligibility.

Next, we examined reference lists of included articles/documents resulting from the database search and used forward citation tracking to identify additional VBP initiatives. Together with the database search, this resulted in a preliminary list of initiatives. To gather additional information on these initiatives and identify potentially relevant other initiatives, we searched Google

and websites of relevant organizations, including the Centers for Medicare and Medicaid Services (CMS) and health insurers. Last, we consulted experts (see Appendix B) to validate our preliminary list of initiatives and to suggest additional initiatives, if any. Importantly, we consulted the four sources in an iterative process. For example, if we encountered an initiative via reference screening that was not identified based on the database search using the original search string, we used initiative-specific key words to search the databases again and obtain additional articles/documents.

### 3.2 Analysis and synthesis

For each identified VBP initiative, we extracted data on (a) general characteristics, (b) key design features with regard to the global base payment and the explicit quality incentives, and (c) effects on value. Regarding the general characteristics, we recorded the name of the initiative, setting, year of implementation, main contracting entities, and availability of a quantitative evaluation. We analyzed the results concerning the two payment components according to the design features shown in Table 4.1. Finally, for initiatives that were evaluated, we recorded the design of quantitative studies, the effects on the applicable value dimensions, and information on the magnitude and statistical significance of effects. Because of heterogeneity in study design and outcome measures used, formal meta-analysis was not possible. Therefore, we present the results narratively.

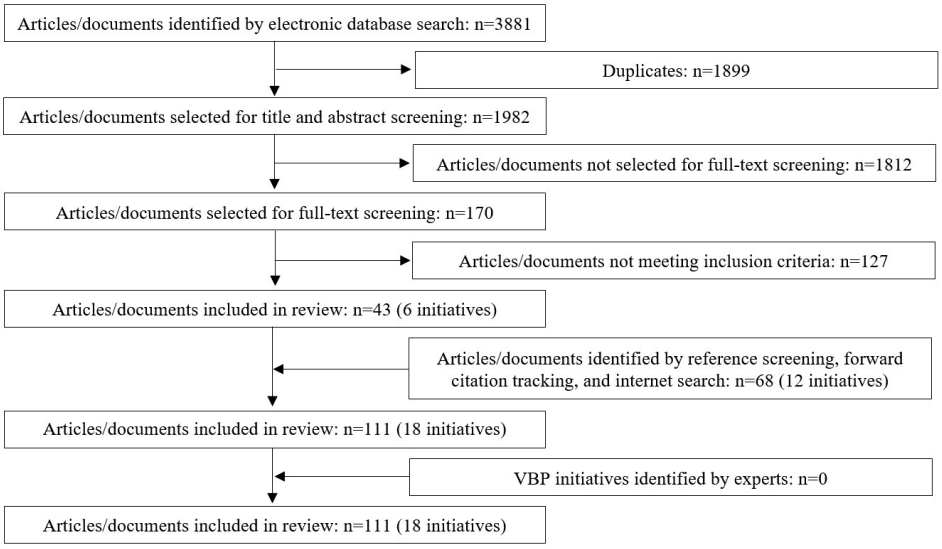
We extracted relevant information using three standardized extraction forms. In case of inconsistencies among articles/documents describing the same initiative, we used information from the article/document with the most recent publication date. After completion of the extraction forms, we summarized the information in three compressed tables with key results only.

## 4. RESULTS

### 4.1 Search results

Applying our search string in the four databases resulted in 3,881 hits (Embase = 1,215; Medline Ovid = 1,403; Web of Science = 1,160; Cochrane Central = 103). After removing duplicates and examining titles and abstracts, we retrieved full texts of 170 potentially relevant articles/documents, which were screened in detail by the first author. Of these, we included 43 articles/documents describing six VBP initiatives. Based on reference screening, forward citation tracking, and searching the Internet, we added 68 articles/documents describing another 12 VBP initiatives. Since expert consultation did not result additional initiatives or articles/documents, we included a total of 111 articles/documents in the review (see Appendix C), representing 18 VBP initiatives (Figure 4.1).

**Figure 4.1.** Flow diagram of steps taken in the systematic review




The 18 included initiatives represent approximately 15% of all payment reform initiatives that we identified in our search ($N = 126$). More than 40% of all identified initiatives pertain to payment models comprising only one of the two components. Generally, these models are 'traditional' P4P initiatives without global base payments. Examples are the hospital Value-Based Purchasing Program and the Programs for All-inclusive Care for the Elderly. In almost 25% of the cases, we excluded initiatives because they use alternative payment models that do not fit our inclusion criteria. Examples are models where providers receive a case rate for an episode of care related to a specific condition or treatment or separate fees for coordinating patient care (e.g., the Acute Care Episode Demonstration and the Cigna Collaborative Accountable Care Model).

Despite fitting our inclusion criteria, we excluded two initiatives—the Physician Group Practice Demonstration and the Pioneer ACO Model—because they are precursors of current models that are included (#14, 15). Experiences and lessons learned in these 'early versions' were used to (re)design current models and in that sense, we still indirectly incorporated these two initiatives in our review (CMS 2018). For the remaining excluded cases, insufficient information was available to determine whether the payment model consisted of the two payment components and/or to describe the design of these components. Examples are the Medica Patient Choice Model, the Rhode Island Health System Transformation Model, and the Medicaid ACO Learning Collaborative in New York, Vermont, and Washington, respectively.

## 4.2 Description of general characteristics

Table 4.2 summarizes the general characteristics of the 18 identified VBP initiatives. The initiatives were implemented in four different countries: 15 in the United States, 1 in Spain, 1 in Germany,

and 1 in the Netherlands. Most VBP models are regional initiatives, with four initiatives having been implemented nationally (#3, 14, 15, 18). Seven initiatives were initiated by public payers, nine by private payers, and two by public–private partnerships. Of the seven public initiatives, three are U.S. Medicare programs (#14, 15, 18), and four are U.S. Medicaid programs (#1, 8, 12, 16). Five initiatives have been formally evaluated on their impact on spending and/or quality.

**Table 4.2.** General characteristics of identified VBP initiatives

| Name initiative | Country | Setting | Year of implementation | Contracting entities | Evaluated on impact on value |
|---|---|---|---|---|---|
| Accountable Care Collaborative Program | US, Colorado | Public (Medicaid) | 2011 | CMS + the State of Colorado + Regional Accountable Entities | No |
| Advocate Care | US, Greater Chicago area | Private | 2011 | Private health insurer + private group of physicians | No |
| Aetna's Shared Savings Model | US, nationwide | Private | 2011 | Private health insurer + integrated health systems | No |
| Alternative Quality Contract | US, Massachusetts | Private | 2009 | Private health insurer + ACOs | Yes, spending and quality |
| Alzira Model | Spain, Valencia | Public-private partnership | 2003 | The regional Health Ministry + private contractor that owns a hospital | No |
| Anthem WellPoint ACO Arrangement | US, California | Private | 2011 | Private health insurer + health care delivery systems | No |
| CalPERS Sacramento ACO Program | US, California | Public-private partnership | 2010 | Private health insurer + public pension fund + large, independent physician association + hospital system | No |
| Coordinated Care Organizations | US, Oregon | Public (Medicaid) | 2012 | CMS + the State of Oregon + Coordinated Care Organizations | No |
| Dutch Shared Savings Program | The Netherlands, Twente region | Private | 2014 | Private health insurer + multispecialty primary care provider groups | No |
| Gesundes Kinzigtal | Germany, Kinzigtal region | Private | 2005 | Two statutory private health insurers + physician network that concluded a contract with health management company | Yes, only quality |
| Horizon BCBS New Jersey ACO Pilot | US, North of Atlantic City, New Jersey | Private | 2010 | Private health insurer + large, multispecialty medical group | No |
| Integrated Health Partnership Demonstration Project | US, Minnesota | Public (Medicaid) | 2013 | CMS + the State of Minnesota + health care delivery systems | No |
| Medica Shared Savings Model | US, Minnesota | Private | 2009 | Private health insurer + integrated health systems and physician clinics | No |

**Table 4.2.** General characteristics of identified VBP initiatives (continued)

| Name initiative | Country | Setting | Year of implementation | Contracting entities | Evaluated on impact on value |
|---|---|---|---|---|---|
| Medicare Shared Savings Program | US, nationwide | Public (Medicare) | 2012 | CMS + ACOs | Yes, spending and quality |
| Next Generation ACO Model | US, nationwide | Public (Medicare) | 2016 | CMS + ACOs | No |
| Partners for Kids Program | US, Ohio | Public (Medicaid) | 2012 | CMS + five Medicaid Managed Care Plans + large pediatric ACO | Yes, spending and quality |
| ProvenHealth Navigator | US, Pennsylvania | Private | 2006 | Private health insurer + Patient-Centered Medical Homes | Yes, only spending |
| Independence at Home | US, nationwide | Public (Medicare) | 2012 | CMS + primary care practices | No |

Note. ACO = accountable care organization; BCBS = Blue Cross Blue Shield; CalPERS = The California Public Employees' Retirement System; CMS = Centers for Medicare and Medicaid Services; US = United States.

## 4.3 Key design features of identified VBP initiatives

Table 4.3 summarizes the initiatives' key design features. In sections 4.3.1 and 4.3.2 these findings are discussed and synthesized for the global base payment and the explicit quality incentives, respectively. The structure of these sections mirror Table 4.1.

### *4.3.1 Key design features of the global base payment*

*Multidisciplinary provider group*

In most initiatives, large, multispecialty provider groups act as main contractor. Typically, these groups comprise different types of physicians, other health care professionals (e.g., nurses, nurse practitioners, physician assistants, case managers, and social workers), and facilities such as hospitals, labs, and outpatient clinics. Although generally a broad range of provider types is involved, all initiatives have a particularly strong focus on substitution to primary care, which becomes evident from the explicit and central role of primary care physicians (PCPs) in all initiatives. We were unable to determine whether individual providers are being employed or subcontracted by the main contractor.

Within each group, providers are jointly accountable for the care for the attributed population with regard to quality and spending. Often, the groups are referred to as ACOs (#4, 10, 14, 15, 16), although terminology varies. Across the 18 initiatives, different types of provider groups take on the role of main contractor. Examples are groups of independent practices that have united themselves into organized networks (e.g., #9), multispecialty group practices that usually have a strong link with hospitals (e.g., #7), and integrated delivery systems including hospitals and a range of other care services like home health care, skilled nursing care, and physician services (e.g., #8). Note that within the same initiative, multiple group types may take on the role of main contractor (e.g., #6).

*Cohesive set of care activities to a predefined population*

Typically, the payment covers virtually the full continuum of primary and specialized medical services and prescription drugs, covered by the relevant benefit package. Information was lacking for #17. In some initiatives (e.g., #1, 8), the payment even covers a broader scope than medical care services only, including behavioral health care and long-term care. In case of the Medicare Shared Savings Program (#14), the Next Generation ACO Model (#15), and the Independence at Home Demonstration (#18), the payment covers the full set of services furnished under Medicare Parts A and B, including, among other services, inpatient care, physician care, outpatient care, skilled nursing facility care, home health agency care, hospice care, and durable medical equipment. Prescription drugs covered under Medicare Part D are not included in the payment of these initiatives.

Commercial initiatives (#2, 3, 4, 6, 9, 10, 11, 13, 17) often use payer affiliation, geographical catchment areas, or a combination of both as a ground for delineating the population. For example, the Alternative Quality Contract (AQC) (#4) only includes Blue Cross Blue Shield of

Massachusetts' members with a health maintenance organization (HMO) or point-of-service policy. The four Medicaid initiatives (#1, 8, 12, 16), automatically enroll all Medicaid beneficiaries in the region in the program. For the three Medicare initiatives (#14, 15, 18) the population consists of Medicare FFS beneficiaries (i.e., age 65 years and older), with the Independence at Home Demonstration (#18) focusing on the most expensive and frailest elders. One initiative (#16) delineates the population based on age, since the focus is on children only. Six of the 18 initiatives (#4, 5, 12, 13, 14, 18), impose a minimum population size per provider group to reduce the influence of stochastic variation (e.g., 5,000 in #4).

Information on the method used to attribute the population to provider groups was not available for five initiatives (#7, 8, 10, 16, 17). Of the other 13 initiatives, 6 use prospective attribution based on prior utilization (#1, 2, 6), affiliation with a provider group or PCP practice (#4, #9, #18), or region (#5). In contrast, three initiatives (#11, 12, 13) retrospectively attribute populations based on the plurality of utilization in the completed year. The three remaining initiatives (#3, 14, 15) use a mixture of assignment methods, depending on, for example, the specific financial risk 'tracks' provider groups may opt for.

*Fixed payment for a defined period of time*

Fourteen initiatives incorporate 'virtual' spending targets by building risk-sharing arrangements on the existing payment modality, most often a FFS-chassis. Three initiatives (#5, 8, 12) actually replaced existing payment systems with 'real' global base payments in the shape of per-member-per-month (PMPM) payments. The remaining initiative (#15) uses both modalities; depending on the 'track' chosen, providers are confronted with a 'virtual' spending target or a 'real' PMPM payment.

Information on the method for setting the payment/target was unavailable for eight initiatives (#1, 3, 5, 6, 7, 8, 11, 16). In 6 of the 10 other initiatives, historical spending in the prior year(s) is the basis for the payment/target. Advocate Care (#2) and the Medica Shared Savings Model (#13) use relative cost benchmarks as targets, that is, the average medical cost trend in the relevant market and the total cost of care of a peer group, respectively. The Independence at Home Demonstration (#18) uses Medicare FS Part A and B expenditures that would have been incurred by beneficiaries in the absence of the initiative as the spending target. Gesundes Kinzigtal (#10) uses a combination of the German 'standardized norm cost' (i.e., the average cost across all insurers, risk adjusted using the German risk-equalization formula) for the specific provider group and spending during a reference period prior to the start of the initiative as a spending target. In nine initiatives, spending targets are trended forward using annual growth rates (#4, 5, 8, 9, 12, 14, 15, 16, 18).

Most initiatives rely on multiyear contracts, although information was missing for six initiatives. One initiative (#7) assumes a multiyear contract but does not specify the exact duration. Nine initiatives apply a contract of 2 to 5 years (#1, 2, 4, 6, 11, 12, 14, 15, 18), one initiative

administers a 15-year contract that is extendable to 20 years (#5), and one initiative even applies an unlimited contract (#10), although the precise content of this contract is unclear.

*Risk adjustment*

In 14 initiatives, the payment/target is adjusted to the risk profile of the attributed population. For the other four initiatives (#1, 3, 5, 6), it was unclear whether or not risk adjustment is being applied. Among the initiatives using risk adjustment, information on the specific variables used is available for 11 initiatives. In one of these (#16), the risk-adjustment model includes only demographic information, while 10 other initiatives (#2, 4, 7, 9, 10, 12, 13, 14, 15, 18) use rather sophisticated models including demographic, socioeconomic, and diagnoses-based morbidity information. Typically, initiatives adopt existing 'off-the-shelf' algorithms, originally developed in the context of risk adjustment for health plan payment. For example, the Medicare Shared Savings Program (#14) uses the CMS Hierarchical Condition Category (HCC) risk-adjustment model (Pope et al. 2004). This model funnels diagnostic codes into diagnoses and ranks them into condition categories, representing conditions with similar cost patterns.

*Risk-mitigating measures*

In eight initiatives providers accept upside risk only (#1, 3, 6, 9, 10, 13, 17, 18), while in eight other initiatives providers also assume downside risk (#2, 4, 5, 7, 8, 11, 15, 16). In the remaining two initiatives, provider groups are free to choose either a one-sided or two-sided contract (#14), or groups are accountable for upside risk only in the first year, and downside risk as well from the second year onward (#12). In initiatives in which providers also assume downside risk, the provider share of savings is larger compared with initiatives in which providers assume upside risk only. For example, in the Medicare Shared Savings Program (#14), providers assuming only upside risk receive 50% of accrued savings, while providers assuming both upside and downside risk receive 60% of savings.

With regard to the risk-sharing rate, information is available for 14 initiatives; for the other 4 initiatives, rates are not available/confidential (#1, 9, 10, 11). Risk-sharing rates for providers exceed 50% in six initiatives (#4, 8, 14, 15, 16, 18), while all other initiatives use a rate of maximally 50%. For example, in the Alzira Model (#5) the risk rate is maximally 7.5%, whereas this rate is 50% in the Anthem WellPoint ACO Arrangement (#6). One initiative (#7) adjusts the risk-sharing rate according to provider groups' ability to influence cost in a particular category. For example, if a provider group is considered not to have any influence over mental health care utilization, the financial risk for this group in this particular domain is zero. For initiatives #12, 14, and 15, the risk-sharing rate increases over time. Typically, in two-sided contracts, the sharing rates for savings are higher than for losses.

The majority of identified VBP contracts include reinsurance provisions, although information is lacking for seven initiatives (#1, 2, 3, 5, 10, 16, 17). The AQC (#4), for example, applies overall cost trend corridors to protect provider groups against significant trends that affect the complete

market. Another example is the Dutch Shared Savings Program (#9), in which providers are protected against high-cost cases by means of a cap of €22.500 (about $25.500) per patient per year. Finally, in all but one (#1) of the 10 initiatives for which information is available, some specific high-cost services are carved-out from the payment contract. Examples are dental care services (#9, 10, 12, 13), transplants (#2, 6, 12), behavioral health services and drugs (#4, 8, 12, 13), and long-term care (#8, 12). The Medicare initiatives (#14, 15, 18) exclude prescription drugs furnished under Medicare Part D from the payment.

### *4.3.2 Key design features of the explicit quality incentives*

*Method of linking payment to quality*

Across the 18 initiatives, we observe three main modalities of linking payment to quality. The most common modality (#1, 2, 4, 7, 8, 10, 11, 13, 15, 17) applies quality incentives as add-on payment in combination with a system in which the provider share of realized savings/losses depends on quality. In the AQC (#4), for example, providers passing higher 'quality gates' receive both a higher bonus and a larger share of savings (or a smaller share of losses). In the second modality, savings/losses also depend on quality but there is no direct add-on payment for high quality scores (#3, 6, 9, 12, 14, 18). The last modality only involves add-on payments (#5, 16).

*Quality measurement*

The initiatives use a broad range of indicators. Clinical quality indicators are adopted most frequently (e.g., #16), although many initiatives incorporate other domains such as patient experience (e.g., #14), patient safety (e.g., #12), and avoidable hospital admissions (e.g., #3). Most initiatives predominantly use measures of process quality, with few initiatives also using outcome measures (e.g., #2). Often, the indicator set is based on a selection of nationally accepted measures (e.g., HEDIS [Healthcare Effectiveness Data and Information Set] measures in #11). For 10 initiatives (#2, 3, 5, 6, 7, 11, 12, 13, 14, 15), we were unable to determine the level of measurement or payment. The remaining initiatives measure quality at the level of individual providers (#10) or provider groups (#1, 4, 8, 9, 16, 18). One initiative splits the savings between individual providers and the relevant group practice (#17).

*Quality incentive structure*

Among the 12 initiatives that implemented add-on payments for quality, eight initiatives only use rewards (#1, 4, 5, 7, 8, 11, 16, 17), while three also use penalties (#2, 13, 15). Information for #10 is missing. The maximum size of the add-on payment relative to the total payment is 10% (#2, 4, 10), but typically lower (e.g., 2% to 3% for #8 and 2% to 8% for #13). An exception is the Alzira Model in Spain (#5) in which the maximum payment size is 20%, although this percentage also includes on-call payments for providers. For initiatives #7, 11, and 16, information on payment size is lacking.

Across the 15 initiatives for which information is available, providers are typically rewarded for both achieving absolute targets and improving over time or relative to other providers. For example, in the Medicare Shared Savings Program (#14), providers share in realized savings only if they attain certain quality levels and show improvement relative to national Medicare FFS and Medicare Advantage. With regard to payment frequency, five initiatives pay on an annual basis (#4, 12, 13, 14, 17) and two on a quarterly basis (#1, 2). Information is lacking for other initiatives.

**Table 4.3.** Key design features of identified VBP initiatives

| **Name initiative** | **Multidisciplinary provider group**<br>a) Main contractor<br>b) Providers in the group<br>c) Employed or subcontracted | **Cohesive set of care activities for a predefined population**<br>a) Healthcare services<br>b) Population<br>c) Attribution method |
|---|---|---|
| 1. Accountable Care Collaborative Program | a) Regional accountable care entity (e.g., community partnerships and insurers), responsible for developing provider networks.<br>b) Formal networks of PCPs and informal networks of specialists, hospitals, and social services.<br>c) N/A. | a) Regular Health First Colorado benefit package: medical care, long-term care, and behavioral health.<br>b) All Medicaid beneficiaries in the region are automatically enrolled.<br>c) Attribution to PCP and corresponding regional accountable care entity based on prior utilization. If a patient did not use care, they are to select a PCP. |
| 2. Advocate Care | a) Private physician group that partners with not-for-profit multi-hospital integrated health system.<br>b) Numerous care sites, including integrated children's hospitals, acute care hospitals, and home care providers. Provider groups consists of solo, group, single- and multi-specialty practices.<br>c) Both (employed and independent). | a) Full continuum of care.<br>b) Fully insured and self-insured commercial PPO members receiving care from the provider group at least 2 times during 2 years. No minimum size.<br>c) Prospective attribution based on prior utilization (claims from previous 2 years). |
| 3. Aetna's Shared Savings Model | a) Variety of health systems (e.g., independent physician associations, multispecialty physician groups, and multispecialty physician groups with contracted hospitals).<br>b) N/A.<br>c) N/A. | a) Full continuum of care.<br>b) Varies by health system.<br>c) In some cases prospective attribution based on enrolment with an ACO. In other cases retrospective attribution based on the plurality of utilization in the completed year. |

| | **Fixed payment for a defined period of time**<br>a) Virtual or real, current main payment system<br>b) Setting the payment or target<br>c) Contract duration | **Risk adjustment & risk-mitigating measures**<br>a) Risk adjustment<br>b) One-sided or two-sided risk<br>c) Risk-sharing rate<br>d) Reinsurance provisions<br>e) Carve-outs | **Explicit quality incentives**<br>a) Link payment and quality<br>b) Quality measures<br>c) Level of measurement/payment<br>d) Rewards or penalties<br>e) Maximum payment size relative to total payment/target.<br>f) Absolute or relative targets<br>g) Payment frequency |
|---|---|---|---|
| | a) Virtual, FFS and PMPM payment for coordination and case management.<br>b) N/A.<br>c) One-year contract, with possibility to renew contract annually for up to 4 years. | a) N/A.<br>b) One-sided risk.<br>c) N/A.<br>d) N/A.<br>e) No carve-outs. | a) P4P and savings conditional on achieving quality thresholds.<br>b) Eight key performance indicators: total cost of care, emergency department visits for conditions that could be prevented with primary care, wellness visits, members receiving behavioral health services/prenatal care/dental care services, rates of overweight/obesity, use of electronic consultations, and agreements with specialists.<br>c) Payments to regional accountable care entity and PCPs.<br>d) Rewards.<br>e) 5% of behavioral health capitation.<br>f) Improvement and meeting criteria.<br>g) Quarterly. |
| | a) Virtual, FFS.<br>b) Benchmark is the projected average medical cost trend in the market (i.e. BCBS Illinois' PPO network)<br>c) Three-year contract. | a) Yes, using DxCG software.<br>b) Two-sided risk.<br>c) Up to 50%.<br>d) N/A. Cost are not truncated.<br>e) Some high-cost services such as transplantation. | a) P4P and savings conditional on achieving quality thresholds.<br>b) 116 measures of clinical quality (i.e. preventive care, acute care processes, and outcomes), patient safety, and patient satisfaction.<br>c) N/A.<br>d) Rewards and penalties (i.e. lower unit price in next year if quality has declined).<br>e) 10%.<br>f) Maintain quality baseline during year 1; thereafter negotiated improvements.<br>g) Quarterly, with annual reconciliation. |
| | a) Virtual, payment system varies by health systems.<br>b) N/A.<br>c) N/A. | a) N/A.<br>b) One-sided risk.<br>c) Up to 50%.<br>d) N/A.<br>e) N/A. | a) Savings conditional on meeting efficiency thresholds and set of clinical quality measures. Whether P4P as add-on is used, is unclear.<br>b) Clinical quality measures and thresholds related to other domains (e.g., avoidable inpatient admissions and ER visits).<br>c) N/A.<br>d) N/A.<br>e) N/A.<br>f) N/A. |

**Table 4.3.** Key design features of identified VBP initiatives (continued)

| **Name initiative** | **Multidisciplinary provider group**<br>a) Main contractor<br>b) Providers in the group<br>c) Employed or subcontracted | **Cohesive set of care activities for a predefined population**<br>a) Healthcare services<br>b) Population<br>c) Attribution method |
|---|---|---|
| 4. Alternative Quality Contract | a) ACOs.<br>b) Variety of primary and specialty providers (e.g., physicians, hospitals, post-acute care facilities). Each ACO is required to include a PCP.<br>c) Physicians are either employed or independent; for other providers information N/A. | a) All medical services BCBS pays for; full continuum of care.<br>b) BCBS members with a HMO/POS policy. Minimum population size of 5,000.<br>c) Prospective attribution based on affiliation with PCP whom enrollees designate each year. |
| 5. Alzira Model | a) Private contractor that owns a hospital, consisting of health insurer, 3 regional savings banks, and 2 construction companies.<br>b) Numerous care sites (e.g., health centers, outpatient clinics, and a hospital).<br>c) Hospital physicians and about half of the PCPs are employed and paid salary. Others are public employees or civil servants. | a) Primary and specialty care.<br>b) Health zones of Alzira.<br>c) Prospective attribution to primary health center based on geographical catchment area. |
| 6. Anthem WellPoint ACO Arrangement | a) Health care delivery systems (e.g., integrated health systems and independent practice associations in private practice).<br>b) Multiple care sites for a broad spectrum of care services (e.g., primary and specialty care, laboratory, physical therapy, radiology, pharmacy, and urgent care).<br>c) N/A. | a) The full continuum of medical services.<br>b) Minimum population size of 15,000.<br>c) Attribution is prospective and based on prior utilization in the past 2 years. To be attributed to a provider group, a patient should have received at least 50 per cent of their care with this group. |

| | **Fixed payment for a defined period of time**<br>a) Virtual or real, current main payment system<br>b) Setting the payment or target<br>c) Contract duration | **Risk adjustment & risk-mitigating measures**<br>a) Risk adjustment<br>b) One-sided or two-sided risk<br>c) Risk-sharing rate<br>d) Reinsurance provisions<br>e) Carve-outs | **Explicit quality incentives**<br>a) Link payment and quality<br>b) Quality measures<br>c) Level of measurement/payment<br>d) Rewards or penalties<br>e) Maximum payment size relative to total payment/target.<br>f) Absolute or relative targets<br>g) Payment frequency |
|---|---|---|---|
| | a) Virtual, FFS.<br>b) Spending target is negotiable. Historical PMPM spending in the population of the group's PCP serves as a starting point and spending is trended forward using a negotiated annual growth rate.<br>c) Five-year contract. | a) Yes, using DxCG software. Health status is measured concurrently.<br>b) Two-sided risk.<br>c) Negotiated, 50-100%.<br>d) Mandatory reinsurance, unit cost corridor, and in some cases overall cost trend corridor.<br>e) Behavioral health services. | a) P4P and risk-sharing rates depend on passing quality gates.<br>b) 64 measures: 32 in ambulatory setting (i.e. HEDIS clinical process and intermediate outcome measures, and patient experience measures) and 32 in hospital setting (i.e. process measures for specific diseases/treatments, patient safety indicators, and patient experience measures). In total, 47 process, 5 outcomes for diabetes, hypertension, and cardiovascular disease, and 12 patient experience measures.<br>c) Payment to ACO.<br>d) Rewards.<br>e) 10%.<br>f) Passing predefined 'gates' and year-to-year performance.<br>g) Annually. |
| | a) Real, annual capitation paid to main contractor.<br>b) N/A, updated according to the yearly growth rate in the Valencian health budget.<br>c) 15-year contract, extendable to 20 years. | a) N/A.<br>b) Two-sided risk.<br>c) Up to 7.5%.<br>d) N/A.<br>e) N/A. | a) P4P, no link between quality and savings.<br>b) Quality and safety targets, including indicators for processes, clinical outcomes, and patient experience.<br>c) N/A.<br>d) Rewards.<br>e) Negotiated, up to 20% between €6.000 and €24.000 per year. Percentage and amount also include on-call payments.<br>f) N/A.<br>g) N/A. |
| | a) Virtual, FFS and care management fee.<br>b) N/A.<br>c) Five-year contract. | a) N/A.<br>b) One-sided risk.<br>c) 50%.<br>d) Caps on high-cost cases and stop-loss reinsurance.<br>e) Transplants. | a) Savings conditional on meeting quality thresholds and efficiency criteria.<br>b) Clinical quality measures and measures related to other domains (e.g., avoidable ER visits or all-cause readmissions), specific to physician care and hospital care.<br>c) N/A.<br>d) Not applicable.<br>e) Not applicable.<br>f) Improvement and attainment.<br>g) N/A. |

**Table 4.3.** Key design features of identified VBP initiatives (continued)

| **Name initiative** | **Multidisciplinary provider group**<br>a) Main contractor<br>b) Providers in the group<br>c) Employed or subcontracted | **Cohesive set of care activities for a predefined population**<br>a) Healthcare services<br>b) Population<br>c) Attribution method |
|---|---|---|
| 7. CalPERS Sacramento ACO Program | a) Large, independent physician association for primary and specialized care and a not-for-profit hospital system.<br>b) Multiple care sites for primary and specialty care.<br>c) N/A. | a) The full continuum of care.<br>b) Blue Shield HMO members in the Sacramento area.<br>c) N/A. |
| 8. Coordinated Care Organizations | a) CCOs i.e. networks of physical, mental, and dental care providers linked to publicly funded health programs.<br>b) A broad range of primary and specialty providers.<br>c) N/A. Each CCO must decide how to contract providers. PCPs usually paid capitation; specialty care providers receive less frequently capitated budget. | a) Full continuum of care, including services for physical health, behavioral health, oral health, mental health, and addiction.<br>b) All Medicaid beneficiaries in the region are automatically enrolled.<br>c) N/A. |
| 9. Dutch Shared Savings Program | a) A multidisciplinary primary care provider group.<br>b) Provider group is led by primary care physicians and comprises nurse practitioners, physician assist, pharmacists, and physiotherapists.<br>c) N/A. | a) All medical services for which health insurer provides coverage under both mandatory and supplementary benefits packages.<br>b) Individuals who take up health insurance from the pilot insurer.<br>c) Attribution based on enrolment with PCP. |

| **Fixed payment for a defined period of time**<br>a) Virtual or real, current main payment system<br>b) Setting the payment or target<br>c) Contract duration | **Risk adjustment & risk-mitigating measures**<br>a) Risk adjustment<br>b) One-sided or two-sided risk<br>c) Risk-sharing rate<br>d) Reinsurance provisions<br>e) Carve-outs | **Explicit quality incentives**<br>a) Link payment and quality<br>b) Quality measures<br>c) Level of measurement/payment<br>d) Rewards or penalties<br>e) Maximum payment size relative to total payment/target.<br>f) Absolute or relative targets<br>g) Payment frequency |
|---|---|---|
| a) Virtual, hospital receives FFS payment and physician group receives capitation budget and pays individual providers FFS.<br>b) PMPM cost target for specific cost categories. Information on how targets are set N/A.<br>c) Multi-year contract, information on exact duration N/A. | a) Yes, based on 'case complexity'.<br>b) Two-sided risk.<br>c) Depends on partner's ability to influence particular costs category. Hospital system: up to 50%. Independent physician association: up to 33.3%.<br>d) Stop-loss reinsurance.<br>e) N/A. | a) P4P and savings conditional on maintaining or improving quality.<br>b) Quality, utilization, and patient satisfaction measures.<br>c) N/A.<br>d) Rewards.<br>e) Unclear, but top-performing physicians have earning potential of 150% of Medicare rates.<br>f) N/A.<br>g) N/A. |
| a) Real, CCOs receive PMPM payment.<br>b) Unclear, adjusted according to historical growth rate.<br>c) N/A. | a) Yes, information on which variables are used N/A<br>b) Two-sided risk.<br>c) Full financial risk: 100%.<br>d) Mandatory reinsurance.<br>e) Mental health drugs, long-term care, case management, and public health. | a) P4P and savings conditional on quality metrics.<br>b) 17 measures on preventive care, access, patient satisfaction, chronic illness management, behavioral health, maternal care, overuse, and electronic health record adoption and use.<br>c) Payment to CCOs.<br>d) Rewards.<br>e) Approximately 2-3%.<br>f) Achievement of benchmark metric or improving performance relative to the State's benchmark.<br>g) N/A. |
| a) Virtual, PCPs are paid salary or combination of capitation, FFS, bundled payment, and P4P.<br>b) Historical spending in the past 3 years (with larger weights attached to more recent years), updated using a growth rate based on spending in a control group of randomly sampled nonparticipating providers in the region, and adjusted for periodic effects (e.g., inflation).<br>c) N/A. | a) Yes, adjusted for demographics and socioeconomic status (concurrently) and morbidity (prospectively).<br>b) One-sided risk.<br>c) Confidential risk rate.<br>d) Cost cap at €22.500 ($25.376) per patient per year.<br>e) Dental care services. | a) Savings conditional on overall quality score. In case performance has declined more than 5% during the year, the overall quality score is insufficient to be eligible for sharing any savings.<br>b) 41 measures in 4 domains: patient satisfaction, chronic care, drug prescription behavior, and practice management.<br>c) Measurement at provider group level.<br>d) Not applicable.<br>e) Not applicable.<br>f) Absolute performance and improvement relative to prior year.<br>g) N/A. |

**Table 4.3.** Key design features of identified VBP initiatives (continued)

| Name initiative | Multidisciplinary provider group<br>a) Main contractor<br>b) Providers in the group<br>c) Employed or subcontracted | Cohesive set of care activities for a predefined population<br>a) Healthcare services<br>b) Population<br>c) Attribution method |
|---|---|---|
| 10. Gesundes Kinzigtal | a) Physician network (including local independent primary care physicians, specialists, and hospitalists) that concluded a contract with a health management company specialized in the management of integrated care.<br>b) Multidisciplinary teams including PCPs, specialists, hospitals, nursing homes, ambulatory agencies, psychotherapists, physiotherapists, and social workers.<br>c) N/A. | a) Care across all health service sectors and indications. Noticeable focus on preventive programs and health promotion.<br>b) Individuals living in the Kinzigtal region who have an insurance policy with 1 of the 2 insurers.<br>c) N/A. |
| 11. Horizon BCBS New Jersey ACO Pilot | a) Multispecialty medical group.<br>b) Primary care, specialty care, ancillary services, and some ambulatory and surgery services.<br>c) N/A. | a) Full continuum of care.<br>b) Patients with a commercial self-insured PPO policy.<br>c) Retrospective attribution based on percentage of total visits. |

| | **Fixed payment for a defined period of time**<br>a) Virtual or real, current main payment system<br>b) Setting the payment or target<br>c) Contract duration | **Risk adjustment & risk-mitigating measures**<br>a) Risk adjustment<br>b) One-sided or two-sided risk<br>c) Risk-sharing rate<br>d) Reinsurance provisions<br>e) Carve-outs | **Explicit quality incentives**<br>a) Link payment and quality<br>b) Quality measures<br>c) Level of measurement/payment<br>d) Rewards or penalties<br>e) Maximum payment size relative to total payment/target.<br>f) Absolute or relative targets<br>g) Payment frequency |
|---|---|---|---|
| | a) Virtual, FFS.<br>b) Spending target determined by combining the German 'standardized norm costs' and spending during a reference period prior to the start of the initiative.<br>c) Unlimited contract. | a) Yes, age, sex, and morbidity, based on German risk-equalization model.<br>b) One-sided risk.<br>c) N/A.<br>d) N/A.<br>e) Dental care services. | a) Payment similar to P4P and savings depending on quality.<br>b) Information on specific measures N/A, but clinical outcome measures and patient satisfaction included.<br>c) Measurement at individual provider level.<br>d) Variable performance-related rewards (i.e. an add-on payment to encourage coordination, rewards for activities such as participating in the electronic health record, and hourly rates for participating in certain project groups).<br>e) 10%.<br>f) N/A.<br>g) N/A. |
| | a) Virtual, FFS.<br>b) N/A.<br>c) Two-year contract. | a) Yes, information on which variables are used N/A.<br>b) Two-sided risk.<br>c) Negotiated, but specific percentages N/A.<br>d) Outliers are eliminated.<br>e) N/A. | a) P4P and savings conditional on meeting quality threshold.<br>b) Variety of HEDIS measures regarding quality of care, diabetes, cardiovascular disease, oncology, and (over)weight assessment.<br>c) N/A.<br>d) Rewards.<br>e) N/A.<br>f) Reward if provider is in top-10% of best performers.<br>g) N/A. |

**Table 4.3.** Key design features of identified VBP initiatives (continued)

| Name initiative | Multidisciplinary provider group<br>a) Main contractor<br>b) Providers in the group<br>c) Employed or subcontracted | Cohesive set of care activities for a predefined population<br>a) Healthcare services<br>b) Population<br>c) Attribution method |
|---|---|---|
| 12. Integrated Health Partnership Demonstration Project | a) Integrated delivery systems (e.g., multispecialty provider network or not-for-profit medical practice group).<br>b) Provider groups deliver full scope of primary care services, coordinate with specialty providers and hospitals, and partner with community organizations and social service agencies.<br>c) N/A. | a) All Medicaid services.<br>b) Medicaid enrollees in Minnesota (children and adults). Minimum population size applies to Track 2 participants (i.e. 2,000 patients).<br>c) Retrospective attribution based on plurality of utilization (>1 visit with provider affiliated with the program), using a 24-month look-back period. |
| 13. Medica Shared Savings Model | a) Integrated health systems and physician clinics.<br>b) A broad range of primary and specialty care (e.g., primary care clinics, inpatient care providers, and home care providers).<br>c) N/A. | a) Full continuum of care.<br>b) Medica's members enrolled in fully insured and self-insured PPOs and some members enrolled in commercially insured HMOs. Minimum population size of 15,000 to 20,000 member-months or 1,250 to 1,667 patients.<br>c) Retrospective attribution based on claims (attribution in case of receiving >50% of primary care services from the group) with 1 year look-back. |

| **Fixed payment for a defined period of time**<br>a) Virtual or real, current main payment system<br>b) Setting the payment or target<br>c) Contract duration | **Risk adjustment & risk-mitigating measures**<br>a) Risk adjustment<br>b) One-sided or two-sided risk<br>c) Risk-sharing rate<br>d) Reinsurance provisions<br>e) Carve-outs | **Explicit quality incentives**<br>a) Link payment and quality<br>b) Quality measures<br>c) Level of measurement/payment<br>d) Rewards or penalties<br>e) Maximum payment size relative to total payment/target.<br>f) Absolute or relative targets<br>g) Payment frequency |
|---|---|---|
| a) Real, population-based payment.<br>b) Negotiable. Prior year's spending is starting point and trended forward using an expected trend rate.<br>c) One-year contract that renews annually during 3 years. | a) Yes, age, sex, and diagnostic information using Johns Hopkins Adjusted Clinical Groups tool.<br>b) One-sided risk in year 1, thereafter two-sided risk.<br>c) 25% in year 1 and 2, thereafter 50%. Up to an agreed maximum savings/losses threshold.<br>d) Cost cap at $200.000 per patient per year.<br>e) Dental care services, transportation, personal care services in home care, long-term care, and residential mental health. | a) Savings conditional on total quality score; losses do not depend on quality.<br>b) Measures of care quality (nationally accepted indicators for e.g., screening and patient safety; weight 70%), health information technology (weight 20%), and pilot measures (based on populations served; weight 10%).<br>c) N/A.<br>d) Not applicable.<br>e) Not applicable.<br>f) In year 1 only reporting. Thereafter, relative thresholds (i.e. being at least in 30th percentile for State or Medicaid average rates) and improvement during the years.<br>g) Annually. |
| a) Virtual, FFS with withholds or prospective adjustments for the risk and reward pool.<br>b) Spending target in comparison to a peer group.<br>c) N/A. | a) Yes, age, sex, and diagnostic information using Johns Hopkins Adjusted Clinical Groups tool.<br>b) One-sided risk.<br>c) Up to 50%.<br>d) Cost cap at $250.000 or $500.000 per patient per year.<br>e) Behavioral health and dental care services. | a) P4P and savings conditional on quality.<br>b) Measures of quality, patient experience, provider collaboration, and utilization among practices, according to Minnesota Community Measurement Program focusing on prevention, chronic care, and utilization.<br>c) N/A.<br>d) Rewards and penalties.<br>e) 2-8%.<br>f) Attainment and improvement.<br>g) Annually. |

**Table 4.3.** Key design features of identified VBP initiatives (continued)

| Name initiative | Multidisciplinary provider group<br>a) Main contractor<br>b) Providers in the group<br>c) Employed or subcontracted | Cohesive set of care activities for a predefined population<br>a) Healthcare services<br>b) Population<br>c) Attribution method |
|---|---|---|
| 14. Medicare Shared Savings Program | a) Medicare ACOs.<br>b) ACO professionals (i.e. physicians and certain non-physician practitioners). Involvement of PCP is mandatory.<br>c) N/A. | a) The full set of services furnished under Medicare Parts A and B.<br>b) Medicare FFS beneficiaries. Minimum population size of 5,000.<br>c) Attribution is based on where patients have received the plurality of primary care services in that year. Track 1 and 2: prospective attribution, with retrospective reconciliation. Track 3: prospective attribution. |
| 15. Next Generation ACO Model | a) ACOs that are experienced in coordination care for defined populations.<br>b) Participants (i.e. PCPs aligned with ACO), preferred providers (e.g., specialists, hospitals, home health facilities), and all other Medicare providers (no formal link between these providers and the model).<br>c) N/A. | a) All services covered by Medicare Part A or Part B.<br>b) Medicare FFS beneficiaries.<br>c) Prospective attribution based on claims using provider lists, supplemented with possibility for beneficiaries to confirm a care relationship with an ACO. |

| **Fixed payment for a defined period of time**<br>a) Virtual or real, current main payment system<br>b) Setting the payment or target<br>c) Contract duration | **Risk adjustment & risk-mitigating measures**<br>a) Risk adjustment<br>b) One-sided or two-sided risk<br>c) Risk-sharing rate<br>d) Reinsurance provisions<br>e) Carve-outs | **Explicit quality incentives**<br>a) Link payment and quality<br>b) Quality measures<br>c) Level of measurement/payment<br>d) Rewards or penalties<br>e) Maximum payment size relative to total payment/target.<br>f) Absolute or relative targets<br>g) Payment frequency |
|---|---|---|
| a) Virtual, FFS.<br>b) Historical spending in the past 3 years (with larger weights attached to more recent years), trended forward by the national growth rate.<br>c) At least three-year contract. | a) Yes, using the CMS-HCC model. Initially prospectively, but retrospectively adjusted.<br>b) ACOs can choose to accept one-sided risk (track 1) or two-sided risk (track 2 and 3).<br>c) Track 1 (50% of savings), track 2 (60% of savings and 40-60% of losses), track 3 (70% of savings and 40-75% of losses). Maximum share of savings payment capped at 10% (track 1), 15% (track 2), and 20% (track 3) of spending target.<br>d) Expenditures capped at 99th percentile of expenditure distribution.<br>e) N/A. | a) Savings depend on overall quality score. Minimum savings rate and minimum losses rate that must at least be met to qualify for shared savings or repay shared losses.<br>b) Four quality domains: Patient/caregiver experience, care coordination/patient safety, preventive health, and at risk population<br>c) N/A.<br>d) Not applicable.<br>e) Not applicable.<br>f) Attainment and improvement, relative to national Medicare FFS and Medicare Advantage percentiles.<br>g) Annually. |
| a) Both possible. Virtual, FFS or FFS and PMPM payment. Real, PMPM payment equal to percentage FFS reduction or capitation.<br>b) Historical spending trended forward by the national growth rate and Medicare geographic pricing factors.<br>c) Three-year contract, extendable to five-year contract. | a) Yes, using the CMS-HCC model. Initially prospectively, but retrospectively adjusted.<br>b) Two-sided risk.<br>c) Type A: performance year 1-3 80% and performance year 4 and 5 85%. Type B: 100%. Maximum share of savings payment is capped at 15% of spending target.<br>d) Expenditures capped at 99th percentile of expenditure distribution.<br>e) N/A. | a) Share of savings is conditional on quality; losses are independent. In addition, the quality score is used in determining the discount applied to the spending target.<br>b) 31 measures on 4 domains with equal weights: patient/caregiver experience, care coordination/patient safety, preventive health, and population at-risk of chronic diseases.<br>c) N/A.<br>d) Not applicable.<br>e) Not applicable.<br>f) Attainment and improvement, relative to national Medicare FFS and Medicare Advantage percentiles.<br>g) N/A. |

**Table 4.3.** Key design features of identified VBP initiatives (continued)

| Name initiative | Multidisciplinary provider group<br>a) Main contractor<br>b) Providers in the group<br>c) Employed or subcontracted | Cohesive set of care activities for a predefined population<br>a) Healthcare services<br>b) Population<br>c) Attribution method |
|---|---|---|
| 16. Partners for Kids Program | a) Pediatric ACO.<br>b) Academic medical center with multiple facilities (NCH), primary and specialty physician practice groups and advanced practice professionals.<br>c) N/A. | a) All Medicaid care.<br>b) All Medicaid beneficiaries aged 0-18 years in central and southeastern Ohio.<br>c) N/A. |
| 17. ProvenHealth Navigator | a) Patient-centered medical homes (i.e. reengineered primary care practices) owned by private health insurer or private independent physician practices.<br>b) Medical home teams composed of PCPs, teams of specialists, physician's assistants, nurses, case managers, pharmacists, social workers, and community health assistants.<br>c) N/A. | a) N/A.<br>b) Adult commercial population.<br>c) N/A. |

| **Fixed payment for a defined period of time**<br>a) Virtual or real, current main payment system<br>b) Setting the payment or target<br>c) Contract duration | **Risk adjustment & risk-mitigating measures**<br>a) Risk adjustment<br>b) One-sided or two-sided risk<br>c) Risk-sharing rate<br>d) Reinsurance provisions<br>e) Carve-outs | **Explicit quality incentives**<br>a) Link payment and quality<br>b) Quality measures<br>c) Level of measurement/payment<br>d) Rewards or penalties<br>e) Maximum payment size relative to total payment/target.<br>f) Absolute or relative targets<br>g) Payment frequency |
|---|---|---|
| a) Virtual. Three payment mechanisms: (1) FFS + P4P for independent providers contracted as member, (2) FFS for community providers not contracted as member, and (3) capitation for the academic personal from NCH.<br>b) N/A.<br>c) N/A. | a) Yes, age and sex.<br>b) Two-sided risk.<br>c) Full financial risk: 100%.<br>d) N/A.<br>e) N/A. | a) P4P, no link between quality and savings. P4P for contracted providers, not for non-members and hospital physicians.<br>b) Selection of HEDIS measures (n=14), number of Medicaid members accepted per physician, completion of Maintenance of Certification program, and being recognized as PCMH.<br>c) Payment at provider group level.<br>d) Rewards.<br>e) N/A.<br>f) N/A.<br>g) N/A. |
| a) Virtual, FFS.<br>b) Spending in the past 2 years, adjusted for medical cost inflation.<br>c) N/A. | a) Yes, information on which variables are used N/A.<br>b) One-sided risk.<br>c) 50%.<br>d) N/A.<br>e) N/A. | a) P4P and savings conditional on meeting quality targets.<br>b) Shared savings conditional on 10 measures regarding chronic illnesses, preventive care, care transition, patient/professional experience, and continuous improvement. For P4P, a more comprehensive set of HEDIS-measures is used.<br>c) Measurement at primary care practices level. Payments split between providers and practice.<br>d) Rewards.<br>e) 9%.<br>f) Improve and maintain quality.<br>g) Annually. |

**Table 4.3.** Key design features of identified VBP initiatives (continued)

| Name initiative | Multidisciplinary provider group<br>a) Main contractor<br>b) Providers in the group<br>c) Employed or subcontracted | Cohesive set of care activities for a predefined population<br>a) Healthcare services<br>b) Population<br>c) Attribution method |
|---|---|---|
| 18. Independence at Home | a) Single primary care practices, other multidisciplinary teams or consortia (multiple primary care within a region) that are led by physicians or nurse practitioners (in total 14).<br>b) Physicians, nurses, physician assistants, pharmacists, social workers, and other staff required to deliver complete range of primary care services in home setting.<br>c) N/A. | a) Care across all settings.<br>b) High-cost, frail Medicare beneficiaries with multiple chronic conditions and functional dependencies (e.g., feeding and walking). Minimum population size of 200.<br>c) Attribution based on enrolment with PCP. |

Note. ACO = accountable care organization; BCBS = Blue Cross Blue Shield; CCO = coordinated care organization; CMS-HCC = Centers for Medicare and Medicaid Services' hierarchical condition category risk-adjustment model; ER = emergency room; FFS = fee-for-service; HEDIS = healthcare effectiveness data and information set; HMO = health maintenance organization; N/A = not available; NCH = Nationwide Children's Hospital; PCP = primary care provider/physician; PMPM = per member per month; POS = point-of-service; PPO = preferred provider organization; P4P = pay-for-performance.

| **Fixed payment for a defined period of time**<br>a) Virtual or real, current main payment system<br>b) Setting the payment or target<br>c) Contract duration | **Risk adjustment & risk-mitigating measures**<br>a) Risk adjustment<br>b) One-sided or two-sided risk<br>c) Risk-sharing rate<br>d) Reinsurance provisions<br>e) Carve-outs | **Explicit quality incentives**<br>a) Link payment and quality<br>b) Quality measures<br>c) Level of measurement/payment<br>d) Rewards or penalties<br>e) Maximum payment size relative to total payment/target.<br>f) Absolute or relative targets<br>g) Payment frequency |
|---|---|---|
| a) Virtual, FFS.<br>b) Medicare FFS Part A and B expenditures that would have been incurred by beneficiaries in the absence of the initiative, trended forward using set annual growth rate.<br>c) Five-year contract. | a) Yes, using the CMS-HCC and CMS ESRD model. To reflect functional impairment, frailty factors are used.<br>b) One-sided risk.<br>c) Ranging from 50 to 80%, with higher shares with higher quality.<br>d) Expenditures capped at 99th percentile of expenditure distribution.<br>e) Claims associated with hurricane Sandy were not included. Indirect and graduate medical education and disproportionate share hospital payments excluded. | a) Savings conditional on meeting at least 3 of the 6 quality targets and surpassing savings threshold of 5%.<br>b) Shared savings depending upon proportion of 6 quality measures met: rates of emergency department and inpatient admissions for ambulatory care–sensitive conditions, 30-day readmission rate, contact with and visits to beneficiaries within 48 hours of hospital admission and discharge, completed medication reconciliation, and documentation of patient preferences.<br>c) Practice / consortium level.<br>d) Not applicable.<br>e) Not applicable.<br>f) N/A.<br>g) N/A. |

## 4.4 Effects on value

Table 4.4 presents information on the effects on value of the five VBP initiatives that have been evaluated. For these initiatives, only effects on quality and spending are available (yet). In total, we included 24 studies, 20 of which pertain to either the AQC (#4) or the Medicare Shared Savings Program (#14). Partners for Kids (#16) was evaluated in two studies, while both Gesundes Kinzigtal (#10) and ProvenHealth Navigator (#17) were each evaluated in one study.

Typically, studies adopted a difference-in-differences design investigating the effects of the initiative on both spending/resource use and quality of care. Initiative #10 has only been evaluated on its impact on quality and #17 only on its impact on spending. Usually, studies compared enrollees attributed to providers participating in the initiative with comparable enrollees attributed to providers not participating in the initiative, using pre- and post-intervention longitudinal data. Below, we summarize the main findings of the evaluation studies separately for the AQC, the Medicare Shared Savings Program, and the three other initiatives.

### *4.4.1 Alternative Quality Contract*

Using 3 years of pre-intervention data and 4 years of postintervention data, Song et al. (2014) investigated the impact of the AQC on medical spending growth and quality of care for the general population of Massachusetts AQC enrollees. The authors found that spending growth was significantly lower in the first 4 years of the contract for the four cohorts under study (2009-2012) compared with control states. For the 2009 cohort, for example, 6.8% savings were realized over the 4-year period ($p < .001$), mainly as a result of lower prices and volumes in the outpatient facility setting. Similar results were found for the other three cohorts. For the 2009 cohort, savings first exceeded quality incentive payments and investments in, for example, information technology in 2012. Regarding quality, Song et al. compared scores on 18 measures of ambulatory care processes and five outcome measures for chronic diseases to New England and national HEDIS averages. Quality improvements were generally significantly larger for the AQC cohorts. Two earlier studies conducted by largely the same researchers (Song et al. 2011; Song et al. 2012) found similar results regarding both spending and quality.

Nine other studies explored the effects on spending on and utilization of specific services and the effects in specific populations. McWilliams et al. (2013) found significant reductions in spending for FFS Medicare beneficiaries served by provider organizations in the AQC compared with beneficiaries served by providers not in the contract, suggesting a positive spillover effect. Uptake of tobacco cessation treatment slightly increased in the AQC population (Huskamp et al. 2016). Song, Fendrick et al. (2013) provide evidence that providers participating in the contract used lower priced facilities and services more often than providers outside the contract. Barry et al. (2015), however, show that mental health care delivery was not meaningfully affected in the first years of the AQC. In addition, other studies did not find significant differences in pharmaceutical spending and utilization, pediatric health care spending or utilization, emergency department use, and substance use disorder treatment between intervention and control groups

(Afendulis et al. 2014; Chien et al. 2014; Sharp et al. 2013; Stuart et al. 2017). Finally, Song et al. (2017) found no significant changes in spending between enrollees in the AQC in areas with lower and higher socioeconomic status.

With regard to quality, one study (Chien et al. 2014) found small but significant positive effects on pediatric preventive care measures, but no effects for diabetes, cardiovascular disease, and HEDIS measures related to substance use (Barry et al. 2015; Stuart et al 2017). Two other studies (McWilliams et al. 2013; Song et al. 2017) observed a positive change for some measures—such as annual rates of low-density lipoprotein cholesterol and adult preventive care—but not for others.

#### *4.4.2 Medicare Shared Savings Program*

Eight studies evaluated the effect of the Medicare Shared Savings Program on spending/utilization and/or quality. Of the four studies evaluating the impact on spending/utilization, three found significant reductions relative to the control groups. Specifically, McWilliams et al. (2016) and Colla et al. (2016) found reductions in total spending of approximately 1% compared with beneficiaries served by providers not participating in the program. McWilliams et al. (2017) show a 9% reduction in post-acute spending and Colla et al. (2016) found a decrease of hospitalizations and emergency department visits of 1.3 and 3 events per 1,000 beneficiaries per quarter, respectively. One study (Busch et al. 2016) found no significant changes in spending and utilization of mental health care.

Of the six studies reporting on the impact on quality, three studies found insignificant effects (Busch et al. 2016; McWilliams et al. 2016; Winblad et al. 2017). The three remaining studies found small but significant reductions of hospital readmissions after common surgical procedures (Borza et al. 2019) and significant improvements of some patient experience measures (McWilliams et al. 2014). Finally, Winblad et al. (2017) demonstrate a significant reduction of 1% in rehospitalization rates from skilled nursing facilities compared with the control group.

#### *4.4.3 Other initiatives*

Four different studies evaluated Gesundes Kinzigtal, Partners for Kids Program, and ProvenHealth Navigator. Kelleher et al. (2015) demonstrate lower PMPM spending in the Partners for Kids Program compared with Ohio Medicaid FFS ($p < .001$) and Ohio Managed Care ($p = .121$) populations. A study investigating the effects of the ProvenHealth Navigator (Gilfillan et al. 2010) found that the number of hospital admissions and readmissions reduced by 18% ($p < .01$) and 36% ($p = .02$), respectively, although total cost of care did not change.

Regarding quality, three studies mainly found positive or null effects as a result of participation in the particular program relative to the control group. For example, Pimperl et al. (2017) show improvements for Gesundes Kinzigtal enrollees in potential years of life lost and estimated survival time, but found no significant effect in average age at time of death. In contrast, one study (Kelleher et al. 2015) provides evidence of significant declines in quality for 2 of the 15 measures used in the Partners for Kids Program: diabetes short-term admission rates and perioperative hemorrhage or hematoma rates.

**Table 4.4.** Effects of five identified VBP initiatives that have been formally evaluated

| Name initiative | References | Study design |
|---|---|---|
| Alternative Quality Contract | (1) Afendulis et al. 2014<br>(2) Barry et al. 2015<br>(3) Chien et al. 2014<br>(4) Huskamp et al. 2016<br>(5) McWilliams et al. 2013<br>(6) Sharp et al. 2013<br>(7) Song et al. 2011<br>(8) Song et al. 2012<br>(9) Song et al. 2013<br>(10) Song et al. 2014<br>(11) Song et al. 2017<br>(12) Stuart et al. 2017 | (1) DiD analyses of drug spending and utilization between 2006 and 2010.<br>(2) DiD analyses of probability of mental health service use, spending, HEDIS metrics for diabetes and cardiovascular conditions using 2006-2011 data.<br>(3) DiD analyses of quality and spending between 2006 and 2010 for children aged 0 to 21 years, including children with special health care needs (CSHCN).<br>(4) DiD analyses of tobacco cessation service use using 2006-2011 data.<br>(5) DiD analyses of spending and quality between 2007 and 2010 for elderly FFS Medicare beneficiaries in Massachusetts served by 11 provider organizations entering the AQC in 2009 or 2010 versus beneficiaries served by other providers.<br>(6) DiD analyses of emergency department (ED) visits using 2006-2009 data.<br>(7) DiD analyses of spending and quality using 2006-2009 data.<br>(8) DiD analyses of spending using 2006-2010 data for the 2009 and 2010 intervention cohort.<br>(9) DiD analyses of spending and utilization of several categories of medical technologies and quality using 2006-2010 data.<br>(10) DiD analyses of spending and unadjusted DiD analyses for ambulatory process quality and outcome measures during the first 4 years (2009-2012) of the initiative for the 2009, 2010, 2011, and 2012 cohorts using 2006-2012 data.<br>(11) DiD analyses of spending and quality using 2006-2012 data for enrollees in areas with lower and higher socioeconomic status. Outcome measures were measured only after the intervention.<br>(12) DiD analyses of substance use disorder service use, spending, and three HEDIS-based performance measures related to substance use disorder using 2006-2011 data. |
| Gesundes Kinzigtal | Pimperl et al. 2017 | (1) Quasi-experimental design using propensity score matched control to evaluate the effect on population health using 2005-2013 data. Control group is a random sample of all members of the two insurers in the region Baden-Wurttemberg of 18 years and older. |

| Effects on resource use/spending | Effects on quality |
|---|---|
| (1) No significant effect on drugs utilization.<br>(2) Intervention group is slightly less likely (–1.41%; P<0.05) to use mental health services. No significant change in mental health spending, but a 1% annual decline in total health care spending for mental health services users.<br>(3) No significant effect on spending trends.<br>(4) Significant increases rates of tobacco cessation treatment use for the overall population (+0.13%; P<0.0001).<br>(5) Significant reductions in spending for Medicare beneficiaries in intervention (change of -$99 or -3.4% relative to an expected quarterly mean of $2,895; P=0.02).<br>(6) No significant effect on ED use.<br>(7) Smaller spending increase for intervention group, i.e., $15.51 less per quarter (-1.9%; P = 0.007).<br>(8) Savings of $22.58 over 2 years (-2.8%; P=0.04).<br>(9) Higher use of colonoscopies for the intervention group in the first 2 years of the contract (+5.2%; P=0.04). Decreases in spending on cardiovascular services in the first 2 years (-7.4%; P=0.02), and on imaging services (-6.1%; P<0.001). No effect in orthopedics.<br>(10) Over the 4-year period lower spending growth for the intervention group (6.8% for the 2009 cohort; P<0.001). The 2010/2011/2012 cohorts had savings of 8.8% (P<0.001), 9.1% (P<0.001), and 5.8% (P=0.04).<br>(11) No significant differences in spending between areas with lower versus higher socioeconomic status.<br>(12) No sizeable changes. | (2) No significant improvements for diabetes or cardiovascular disease among enrollees with co-occurring mental healthcare use. For two measures (nephropathy monitoring and retinal exams) non-mental health users appear to have benefited more than mental health care users (annual change in probability of –2.90%; P<0.01 and –2.57%; P<0.05).<br>(3) Significant, positive effect on pediatric preventive care quality measures tied to P4P (+1,8% for CSHCN and +1.2% for non-CSHCN; P<0.001). No significant changes for measures not tied to P4P.<br>(5) Significant improvements of some measures (e.g., 3.1% for low-density lipoprotein cholesterol testing [P<0.001] and 2.5% for cardiovascular disease [P<0.001]), but no differential change for others.<br>(7) Improved quality for chronic conditions in adults (P<0.001) and pediatric care (P=0.001) after 1 year, but not for adult preventive care.<br>(8) Improvements in measures for chronic care management (+3.7%; P<0.001), adult preventive care (+0.3%; P=0.008), and pediatric care (+0.3%; P<0.001).<br>(10) Measures of chronic disease management increased by 3.9%, and unadjusted performance in adult preventive care and pediatric care increased by 2.7% and 2.4% (P-values are unavailable) compared to the HEDIS national average. The five outcome measures for patients with diabetes, patients with coronary artery disease, and patients with hypertension improved compared to the national and regional HEDIS scores (size of the effect and P-values unavailable).<br>(11) Process measures improved +1.2% per year more among individuals living in areas with lower versus higher socioeconomic status (P<0.001). No significant differences in outcome measures.<br>(12) No sizeable changes. |
| Not available. | (1) For the ACO intervention group age at time of death is on average 1.4 years higher compared to the control group but not significant, 639 fewer years of potential life were lost compared to the control group (P<0.05), and the estimated survival time is approximately 7 days higher for beneficiaries participating in the program (significant; p-value unavailable). |

**Table 4.4.** Effects of five identified VBP initiatives that have been formally evaluated (continued)

| Name initiative | References | Study design |
|---|---|---|
| 14. Medicare Shared Savings Program | (1) Borza et al. 2019<br>(2) Busch et al. 2016<br>(3) Colla et al. 2016<br>(4) Herrel et al. 2016<br>(5) McWilliams et al. 2014<br>(6) McWilliams et al. 2016<br>(7) McWilliams et al. 2017<br>(8) Winblad et al. 2017 | (1) DiD analyses of hospital readmission after common surgical procedures using 2010-2014 data.<br>(2) DiD analyses of mental health care spending, utilization, and quality using 2008-2013 data.<br>(3) DiD analyses of spending and high-cost institutional use using 2009-2013 data.<br>(4) DiD analyses of 30-day mortality, complications, readmissions, and length of stay for patients undergoing a major surgical resection for various types of cancer using 2011-2013 data.<br>(5) DiD analyses of patient experience using 2010-2013 data.<br>(6) DiD analyses of spending and quality using 2009-2013 data.<br>(7) DiD analyses of post-acute spending and utilization using 2009-2014 data.<br>(8) DiD analyses of all-cause rehospitalizations from skilled nursing facilities using 2007-2013 data. |
| 16. Partners for Kids Program | (1) Gleeson et al. 2016<br>(2) Kelleher et al. 2015 | (1) DiD analyses of pediatric performance of primary care physicians using 2010-2013 data.<br>(2) Observational study of spending, growth rates, and quality using 2008-2013 data. Results for the PFK group is compared to Ohio Medicaid FFS and Ohio managed care (MC). |
| 17. ProvenHealth Navigator | (1) Gilfillan et al. 2010 | (1) DiD analyses of hospital admissions, readmission rates, and the total cost of care using 2005-2008 data for Medicare Advantage patients at 11 intervention sites and 75 control groups. |

| Effects on resource use/spending | Effects on quality |
|---|---|
| (2) No significant changes in mental health care spending and utilization.<br>(3) Modest reductions in total spending (-1.3%; P<0.001). Hospital and ED use reduced significantly by 1.3 (P<0.05) and 3.0 (P<0.01) events per 1000 beneficiaries per quarter.<br>(6) Significant reductions in spending for the 2012 cohort (-1.4%; P=0.02), but not for the 2013 cohort.<br>(7) Significant reductions in post-acute spending (-9.0%; P=0.003 for 2012 ACO cohort and smaller for the 2013 and 2014 cohort). | (1) Significant reduction in readmissions for hospitals in the program (-0.52%; P=0.021).<br>(2) No significant changes in quality metrics.<br>(4) No significant effect on perioperative outcome measures.<br>(5) Improvements in some patients experience measures (e.g., effect size for reports of timely access to care is 2.1 standard deviation of the ACO-level distribution, adjusted for trends; P=0.02), but not (significantly) in others (e.g., overall ratings of care and physicians).<br>(6) No significant differences in quality or use of low-value services for the majority of measures.<br>(8) Significant reduction in re-hospitalization rate (-0.994%; P<0.01). |
| (2) Compared to both control groups, PMPM spending was significantly lower in 2008, and grew at a rate of $2.40 per year compared to $16.15 per year in the FFS group (P<0.001) and $6.47 per year (P<0.121) in the MC group. | (1) Significant improvements in 8 of the 14 HEDIS measures for preventive care, chronic care, and acute care primary care services for the group of Nationwide Children Hospital physicians compared to incentivized physicians ('traditional' P4P). ORs favoured the intervention group mainly in the immunization measures (range of OR of 0.34 with CI of 0.31-0.37 for hepatitis vaccine to 0.86 with CI of 0.78-0.95 for meningococcal vaccine).<br>(2) Significant improvement for gastroenteritis admission rate (-0.05 events/1000; P=0.000), pediatric quality acute composite (-0.03 events/1,000; P=0.018), and pediatric quality overall composite (-0.05 events/1,000; P=0.046). Significant declines in quality regarding diabetes short-term admission rates (+0.02 events/1,000; P=0.027) and perioperative haemorrhage or hematoma rates (+3.99 events/1,000; P=0.048). No significant differences on 10 other measures. |
| (1) Significant reduction in hospital admissions (-18%; P<0.01) and readmissions (-36%; P=0.02). Total cost of care decreased 7% (not significant). | Not available. |

## 5. DISCUSSION

### 5.1 Summary and discussion of main findings

In this article, we systematically identified and analyzed 18 VBP initiatives aiming at improving value in a broad sense. Specifically, our focus was on initiatives combining global base payments with payments explicitly linked to quality. Our analysis has resulted in a comprehensive overview of the possibilities in terms of operationalization of the two payment components and associated design features. Six main findings merit further discussion.

First, although all identified initiatives share the same two payment components, they differ considerably in the exact operationalization thereof. Specifically, we observed heterogeneity in the degree of risk sharing, the method of attributing populations to provider groups, the sophistication of the risk-adjustment methodology, and the way in which payment is linked to quality. Reasonable explanations for this heterogeneity are local preferences and contextual differences among settings. For example, in a setting in which providers lack experience with bearing downside risk, payers may choose to start with transferring upside risk only, allowing providers to gain this experience. After an adaption period, incentives for cost-conscious behavior can be intensified by transferring some downside risk as well.

Second, 15 of the 18 initiatives have been implemented in the United States. In part, this may be due to the adopted language restriction in this review. Another potential explanation can be found in the specific structure and history of the U.S. health care system. Specifically, it is likely that essential preconditions for a successful introduction of VBP are better fulfilled in the United States than in other countries, enabling a jump-start of VBP in the United States. Collaborative networks of multidisciplinary providers that are able and willing to take on the role of risk-bearing accountable group are historically embedded in the U.S. health care system (Enthoven 2009). This might be partly the result of the integrated delivery systems that gained traction in the 1980s.

A third noteworthy finding is the strong reliance on primary care in all initiatives, which is evident from the explicit and central role of PCPs. In the Dutch Shared Savings Model, for example, groups of PCPs are accountable for the full continuum of primary and specialized care services. As gatekeepers, Dutch PCPs have at least some control over both primary and specialist care, legitimating their role as main contractor. The central focus on primary care across all initiatives is consistent with the global trend toward primary care–oriented systems. This trend is understandable given the many studies showing that areas with higher ratios of PCPs to population are associated with better health outcomes and lower total cost of health services compared with other areas (Starfield et al. 2005).

Fourth, the majority of identified initiatives adopt spending targets with risk-sharing arrangements built on existing (FFS) payment systems. This finding is consistent with the recommendation derived from a major VBP initiative in California to start with 'virtual' targets and shift to 'real' prospective payments at a later stage (Williams & Yegian 2014). Virtual payments can

potentially realize the same goal as real payments, without the regulatory and administrative burdens of replacing current payment and billing systems that could disrupt momentum. In addition, initially testing the model using virtual payments offers the possibility of developing a reliable benchmark from which the fixed payment level can be reasonably negotiated (Williams & Yegian 2014). However, the incentives emanating from virtual payments may be perceived as weaker than those from real prospective payments (Struijs et al. 2018). Thus, although virtual payments can be a practical first step, moving away from FFS should remain a priority (De Bakker et al. 2012; Williams & Yegian 2014).

Fifth, most initiatives apply some form of risk adjustment and incorporate risk-mitigating measures in their payment contracts. This contributes to fairness in payment, reduced incentives for risk selection, and protection against excessive random variation in spending. Apparently, the importance of these two VBP design features is not only recognized in theory (Ash & Ellis 2012; Cattel et al. 2020a) but also in practice. Regarding risk adjustment, initiatives typically use existing diagnoses-based algorithms that were originally developed in the context of health plan payment. Although this may be an efficient and pragmatic approach that could serve its purpose in the short run, in the longer run it seems preferable to customize the risk-adjustment algorithm to the specific purpose of paying providers (Ash & Ellis 2012). This may be particularly relevant to prevent the introduction of new perverse incentives such as for manipulating the diagnoses-based morbidity information used in the risk-adjustment formula to maximize payment (Geruso & Layton 2015; Landon & Mechanic 2017; Markovitz et al. 2019).

Finally, our results indicate that VBP models as defined here have the potential to improve value and contribute to the provision of VBHC. Regarding the five initiatives that have been evaluated, studies generally demonstrate similar or reduced spending growth and equal or improved quality. In this respect, it is noteworthy that the Medicare Shared Savings Program excludes prescription drugs from the VBP contract. Since prescription drugs account for a substantial proportion of total health care spending, it is possible that this initiative did not fully reach its potential for value improvement.

Our findings are consistent with results found for ACOs in the United States that indicate no association between ACO implementation and worsened health outcomes (Kaufman et al. 2019). In addition. Our findings correspond well with the results of a recent review of outcome-based P4P initiatives, which found favorable effects only when P4P was combined with global base payments (Vlaanderen et al. 2019). Conversely, our findings are in contrast with results from prior reviews on the effects of P4P, which did not find convincing evidence for P4P being (cost-) effective in improving value when the underlying, flawed base payment system is left intact (e.g., Eijkenaar et al. 2013; Mendelson et al. 2017; Vlaanderen et al. 2019). A possible explanation for the latter is that P4P typically concerns a relatively small part of the total provider payment, whereas initiatives included in this article focus on reform of the total payment system. Finally, our finding that quality does at least not seem to have deteriorated, suggests that quality—as operationalized by the chosen indicators—did not suffer from the adopted global base payments

in VBP. This is in contrast with the widespread concern about the use of capitation payments in the context of HMOs (Dudley & Luft 2001; Miller & Luft 1997).

### 5.2 Limitations and implications

Our findings should be interpreted in the light of several limitations. First, as any systematic review, this study suffers from publication bias. Second, it is possible that we missed relevant VBP initiatives as a consequence of our search strategy, specifically the restriction to articles/documents written in English or Dutch. In addition, we excluded multiple potentially relevant initiatives due to insufficient information. For example, we expect that long-standing integrated delivery systems such as Kaiser Permanente and Cleveland Clinic also adopt relevant VBP models, but since specific information on the payment structure is lacking, we could not include them. Overall, maximally twice as much VBP initiatives could have been included in this review, had sufficient information been available. Third, we were not always able to describe all relevant design features of each included initiative. In particular, information was often unavailable on the attribution methods, methods of setting the payment/target, internal payment contracts, contract duration, risk-mitigating measures, and quality incentive structure. Fourth, the overrepresentation of U.S. initiatives limits the generalizability of our findings to other settings. Finally, our findings regarding the effects on value are based studies evaluating only 5 of the 18 initiatives, with 20 of the 24 included evaluation studies pertaining to 2 initiatives: The Alternative Quality Contract and the Medicare Shared Savings Program. Moreover, the effects found in these studies are unlikely to reflect the impact of payment reform exclusively. This is because VBP is typically part of a broader approach to value improvement including other interventions that are implemented simultaneously, like structured performance feedback and public reporting.

In addition to the implications mentioned in the section 5.1, the results of this review have two other implications for research and policy. First, from both a research and a policy perspective, the design of VBP models are ideally documented more carefully in the future. Furthermore, it is important that VBP implementation goes hand-in-hand with rigorous evaluation. This is expected to result in important insights with regard to VBP design and the link with effectiveness, enabling others to learn from prior experiences. As this review shows, few initiatives have been subject to rigorous evaluation. Hence, little is still known about the effects in general, let alone about the impact of specific design choices on value. Moreover, the long-term impact of VBP is often not assessed, even though the gains from specific interventions such as investments in prevention are expected to emerge only after a longer period of time. The only two initiatives for which effects in the longer run are available confirm this statement. For example, net savings were generated only after 4 years in the AQC (Song et al. 2014).

Second, policy makers pursuing VBHC should keep in mind that although payment reform is an invaluable element in this process, it is not the only relevant factor. Other financial and nonfinancial interventions on both the supply- and demand-side of the market are likely to be important for the success of VBHC as well. Examples are a joint IT-infrastructure, physician

leadership, performance monitoring with structured feedback, and public reporting (McClellan et al. 2010; Phipps-Taylor & Shortell 2016; Robinson 2001a; Shortell & Casalino 2010). Consistent with the recommendation by Roland and Campbell (2014) that P4P needs to be combined with other improvement strategies to produce sustained improvements, implementing VBP while disregarding other relevant factors is unlikely to materially affect value. The successful AQC, for example, embraced a multifaceted improvement strategy by offering technical support for participating provider groups parallel to the intervention of payment reform (Chernew et al. 2011). The role of other value-adding aspects and the interplay with VBP is an interesting avenue for future research.

### 5.3 Conclusion

In the coming years, VBP models stimulating value in a broad sense will likely continue to gain ground, as the quest toward VBHC proceeds. This article demonstrates that VBP models consisting of global base payments combined with explicit quality incentives are operationalized in practice in various ways. In addition, our results show that this particular VBP model has the potential to improve value and contribute to VBHC. Going forward, this article may serve as inspirational material for those interested in developing new or improving on existing VBP models.

## APPENDIX

### Appendix A: Search strings

***Embase.com: 1215***

((((Value* OR variable OR performance OR explicit OR outcome OR quality OR readmission OR mortality OR complication OR coordination OR efficien* OR effectiveness* OR efficac* OR cost-conscious* OR well-coordinat* OR innovat* OR prevent*) NEAR/6 (based OR program* OR evaluat* OR assess* OR model* OR initiative* OR connect* ) NEAR/6 (Payment* OR incentive* OR remuner* OR fee OR fees OR reward* OR reimburs* OR financing OR funding OR budget OR capitat* OR bonus OR contract OR contracts OR contracting OR contracted OR spending ))):ab,ti) AND ('physician'/exp OR 'medical specialist'/exp OR 'hospital'/de OR 'general hospital'/de OR 'community hospital'/de OR 'geriatric hospital'/de OR 'mental hospital'/exp OR 'pediatric hospital'/de OR 'private hospital'/de OR 'public hospital'/de OR (physician* OR practitioner* OR doctor* OR ((health-care OR healthcare) NEAR/3 provider*) OR hospital* OR clinic OR clinics):ab,ti) NOT ([Conference Abstract]/lim OR [Letter]/lim OR [Note]/lim OR [Editorial]/lim) AND ([english]/lim OR [dutch]/lim) AND [2000-2017]/py

***Medline Ovid: 1403***

("Value-Based Purchasing"/ OR "Value-Based Insurance"/ OR (((Value* OR variable OR performance OR explicit OR outcome OR quality OR readmission OR mortality OR complication OR coordination OR efficien* OR effectiveness* OR efficac* OR cost-conscious* OR well-coordinat* OR innovat* OR prevent*) ADJ6 (based OR program* OR evaluat* OR assess* OR model* OR initiative* OR connect* ) ADJ6 (Payment* OR incentive* OR remuner* OR fee OR fees OR reward* OR reimburs* OR financing OR funding OR budget OR capitat* OR bonus OR contract OR contracts OR contracting OR contracted OR spending ))).ab,ti,kf.) AND (exp "Physicians"/ OR "Specialization"/ OR "hospitals"/ OR "Hospitals, General"/ OR "Hospitals, Community"/ OR "Hospitals, Psychiatric"/ OR "Hospitals, Pediatric"/ OR "Hospitals, Private"/ OR "Hospitals, Public"/ OR (physician* OR practitioner* OR doctor* OR ((health-care OR healthcare) ADJ3 provider*) OR hospital* OR clinic OR clinics).ab,ti,kf.) NOT ((letter OR news OR comment OR editorial OR congresses OR abstracts).pt.) AND (english.la. OR dutch.la.) AND (2000 OR 2001 OR 2002 OR 2003 OR 2004 OR 2005 OR 2006 OR 2007 OR 2008 OR 2009 OR 2010 OR 2011 OR 2012 OR 2013 OR 2014 OR 2015 OR 2016 OR 2017).yr

***Cochrane central: 103***

((((Value* OR variable OR performance OR explicit OR outcome OR quality OR readmission OR mortality OR complication OR coordination OR efficien* OR effectiveness* OR efficac* OR cost-conscious* OR well-coordinat* OR innovat* OR prevent*) NEAR/6 (based OR program* OR evaluat* OR assess* OR model* OR initiative* OR connect* ) NEAR/6 (Payment* OR incentive* OR remuner* OR fee OR fees OR reward* OR reimburs* OR financing OR funding

OR budget OR capitat* OR bonus OR contract OR contracts OR contracting OR contracted OR spending ))):ab,ti) AND ((physician* OR practitioner* OR doctor* OR ((health-care OR healthcare) NEAR/3 provider*) OR hospital* OR clinic OR clinics):ab,ti)

***Web of science: 1160***

TS=(((((Value* OR variable OR performance OR explicit OR outcome OR quality OR readmission OR mortality OR complication OR coordination OR efficien* OR effectiveness* OR efficac* OR cost-conscious* OR well-coordinat* OR innovat* OR prevent*) NEAR/5 (based OR program* OR evaluat* OR assess* OR model* OR initiative* OR connect* ) NEAR/5 (Payment* OR incentive* OR remuner* OR fee OR fees OR reward* OR reimburs* OR financing OR funding OR budget OR capitat* OR bonus OR contract OR contracts OR contracting OR contracted OR spending ))))) AND ((physician* OR practitioner* OR doctor* OR ((health-care OR healthcare) NEAR/2 provider*) OR hospital* OR clinic OR clinics))) AND DT=(article) AND LA=(english)

## Appendix B: List of consulted experts

- Erik Schut (the Netherlands)
- Richard Heijink (the Netherlands)
- Lieven Annemans (Belgium)
- Maria Trottmann (Switzerland)
- Thomas McGuire (US)
- Noaki Ikegami (Japan)

## Appendix C: Overview of included articles and documents


Acerete, B., A. Stafford & P. Stapleton. 2011. 'Spanish healthcare public private partnerships: The Alzira model.' *Critical Perspectives on Accounting* 22:533-549.

Afendulis, C., M. Fendrick, Z. Song, B.E. Landon, D.G. Safran, R.E. Mechanic, R. E. & M.E. Chernew. 2014. 'The impact of global budgets on pharmaceutical spending and utilization: Early experience from the Alternative Quality Contract.' *Inquiry* 51:1-7.

Alderwick, H., C. Ham, & D. Buck. 2015. *Population health systems: Going beyond integrated car*e. Retrieved from https://www.kingsfund.org.uk/sites/default/files/field/field_publication_file/population-health-systems-kingsfund-feb15.pdf

Anthem. 2011. *Accountable care organization.* Presentation, Indiana.

Bailit, M. & C. Hughes. 2011. *Key design elements of shared-savings payment arrangements* (Commonwealth Issue Brief No. 1539). Retrieved from http://www.commonwealthfund.org/-/media/Files/Publications/Issue%20Brief/2011/Aug/1539_Bailit_key_design_elements_sharedsavings_ib_v2.pdf

Bailit, M., C. Hughes, M. Burns, & D.H. Freedman. 2012. *Shared-savings payment arrangement in health care: Six case studies* (Commonwealth Fund No. 1624). Retrieved from http://www.commonwealthfund.org/publications/fund-reports/2012/aug/shared-savings-payment-arrangements

Barnes, A.J., L. Unruh, A. Chukmaitov & E. van Ginneken. 2014. 'Accountable care organizations in the USA: Types, developments and challenges.' *Health Policy* 118:1-7.

Barry, C.L., E.A. Stuart, J.M. Donohue, S.F. Greenfield, E. Kouri, K. Duckworth, Z. Song, R.E. Mechanic, M.E. Chernew & H.A. Huskamp. 2015. 'The early impact of the 'Alternative Quality Contract' on mental health service use and spending in Massachusetts.' *Health Affairs* 34:2077-2085.

Bartels, S.J., L. Gill, L. & J.A. Naslund. 2015. 'The Affordable Care Act, accountable care organizations, and mental health care for older adults: Implications and opportunities.' *Harvard Review of Psychiatry* 23:304-319.

Berwick, D.M. 2011. 'Making good on ACOs' promise – the final rule for the Medicare Shared Savings Program.' *The New England Journal of Medicine* 365:1753-1756.

Blewett, L.A., D. Spencer, D. & P. Huckfeldt. 2017. 'Minnesota Integrated Health Partnership Demonstration: Implementation of a Medicaid ACO model.' *Journal of Health Politics, Policy and Law* 42:1127-1142.

Bodaken, B. 2014. 'Increasing the impact and sustainability of California accountable care organizations.' *California Journal of Politics and Policy* 6:245-247.

Bodaken, B., R. Bankowitz, T. Ferris, J. Hansen, J. Hirshleifer, S. Kronlund, D. Labby, R. MacCornack, M. McClellan & L. Sandy. 2016. *Sustainable success in accountable care.* Retrieved from https://nam.edu/wp-content/uploads/2016/04/Sustainable-Success-in-Accountable-Care.pdf

Borza, J., M.K. Oerline, T.A. Skolarus, E.C. Norton, J.B. Dimick, B.L. Jacobs, L.A. Herrel, C. Ellimoottil, J.M. Hollingsworth, A.M. Ryan, D.C. Miller, V.B. Shahinian & B.K. Hollenbeck. 2019. 'Association between hospital participation in Medicare Shared Savings Program accountable care organizations and readmission following major surgery.' *Annals of Surgery* 269:873-878.

Brennan, K.F. 2017. *Geisinger's evolving provider incentive models*. Presentation, New Orleans.

Burns, L.R. & M.V. Pauly. 2012. 'Accountable care organizations may have difficulty avoiding the failures of Integrated Delivery Networks of the 1990s.' *Health Affairs* 31:2407-2416.

Busch, A.B., H.A. Huskamp & J.M. McWilliams. 2016. 'Early efforts by Medicare accountable care organizations have limited effect on mental illness care and management.' *Health Affairs* 25:1247-1256.

Busse, R. & J. Stahl. 2014. 'Integrated care experiences and outcomes in Germany, the Netherlands, and England.' *Health Affairs* 33:1549-1558.

Caballer-Tarazona, M. & D. Vivas-Consuelo. 2016. 'A cost and performance comparison of public private partnership and public hospitals in Spain.' *Health Economics Review* 6:2-7.

Carlin, C.S. 2014. 'Patient loyalty in a mature IDS market: Is population health management worth it?' *Health Services Research* 49:1011-1033.

Casalino, L.P. & N. Chenven. 2017. 'Independent practice associations: Advantages and disadvantages of alternative form of physician practice organization.' *Healthcare* 5:46-52.

Chang, A.M., D. Cohen, D. McCarty, T. Rieckmann & K.J. McConnell. 2015. 'Oregon's Medicaid transformation: Observations on organizational structure and strategy.' *Journal of Health Politics, Policy and Law* 40:257-264.

Chernew, M.E., R.E. Mechanic, B.E. Landon & D.G. Safran. 2011. 'Private-payer innovation in Massachusetts: The Alternative Quality Contract.' *Health Affairs* 30:51-61.

Chien, A.T., Z. Song, M.E. Chernew, B.E. Landon, B.J. McNeil, D.G. Safran & M.A. Schuster. 2014. 'Two-year impact of the Alternative Quality Contract on pediatric health care quality and spending.' *Pediatrics* 133:96-104.

Chien, A.T., K.H. Schiavoni, E. Sprecher, B.E. Landon, B.J. McNeil, M.E. Chernew & M.A. Schuster. 2016. 'How accountable care organizations responded to pediatric incentives in the Alternative Quality Contract.' *Academic Pediatrics* 16:200-207.

Christensen, E.W. & N.R. Payne. 2016a. 'Pediatric inpatient readmissions in an accountable care organization.' *The Journal of Pediatrics* 170:113-119.

Christensen, E.W. & N.R. Payne. 2016b. 'Effect of attribution length on the use and cost of health care for a pediatric Medicaid accountable care organization.' *JAMA Pediatrics* 170:148-154.

CMS, Centers for Medicare and Medicaid Services. 2014. *Independence at Home Demonstration shared savings methodology: Specifications.* Retrieved from https://innovation.cms.gov/Files/reports/iah-ssmethodologyrpt.pdf

CMS, Centers for Medicare and Medicaid Services. 2016a. *Accountable care organizations: What providers need to know* (ICN 907406, March 2016). Retrieved from https://www.methodisthealthsystem.org/documents/ACO_Providers_Factsheet_ICN907406.pdf

CMS, Centers for Medicare and Medicaid Services. 2016b. *Medicare Shared Savings Program quality measure benchmarks for the 2016 and 2017 reporting years*. Retrieved from https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/sharedsavingsprogram/Downloads/MSSP-QM-Benchmarks-2016.pdf

CMS, Centers for Medicare and Medicaid Services. 2016c. *Independence at Home Demonstration: Fact sheet.* Retrieved from https://innovation.cms.gov/Files/fact-sheet/iah-fs.pdf

CMS, Centers for Medicare and Medicaid Services. 2017. *Medicare Shared Savings Program: Shared savings and losses and assignment methodology – specifications*. Retrieved from https://www.cms.gov/Medicare/Medicare-Fee-for-Service- Payment/sharedsavingsprogram/Downloads/Shared-Savings-Losses-Assignment- Spec-V5.pdf

Cohen, A., S. Klein & D. McCarthy. 2014. *Hill physicians medical group: A market-drive approach to accountable care for commercially insured patients* (Commonwealth Fund No. 1770, vol 23). Retrieved from http://www.commonwealthfund.org/publications/case-studies/2014/oct/hill-physicians-aco-case-study

Colla, C.H., V.A. Lewis, L.S. Kao, A.J. O'Malley, C.H. Chang & E.S. Fisher. 2016. 'Association between Medicare accountable care organization implementation and spending among clinically vulnerable beneficiaries.' *The Journal of the American Medical Association Internal Medicine* 176:1167–1175.

Colorado Department of Health Care Policy and Financing. (Unclear). *The Accountable Care Collaborative*. Presentation, Colorado.

Colorado Department of Health Care Policy and Financing. 2015. *Accountable Care Collaborative phase II concept paper*. Retrieved from https://www.colorado.gov/pacific/sites/default/files/ACC%20Phase%20II%20Concept%20Paper.pdf

Colorado Department of Health Care Policy and Financing. 2017. *Supporting a culture of coverage: Accountable Care Collaborative 2015 annual report*. Retrieved from http://leg.colorado.gov/sites/default/files/6__accountable_care_collaborative_2014-15_annual_report.pdf.

Colorado Health Institute. 2017. *The route to the RAEs: Analyzing the next phase of Medicaid's Accountable Care Collaborative in Colorado*. Retrieved from https://www.coloradohealthinstitute.org/sites/default/files/file_attachments/ACC%20Phase%20Two_0.pdf

Conrad, D.A., D. Grembowski, S.E. Hernandez, B. Lau & M. Marcus-Smith. 2014. 'Emerging lessons from regional and state innovation in value-based payment reform: Balancing collaboration and disruptive innovation.' *The Milbank Quarterly* 92:568-623.

Damberg, C.L., M.E. Sorbero, S.L. Lovejoy, G.R. Martsofl, L. Raaen & D. Mandel. 2014. *Measuring success in health care value-based purchasing programs: Findings from an environmental scan, literature review, and expert panel discussions*. Retrieved from https://www.rand.org/content/dam/rand/pubs/research_reports/RR300/RR306/RAND_RR306.pdf

DeJonge, K.E., G. Taler & P.A. Boling. 2009. 'Independence at Home: Community-based care for older adults with severe chronic illness.' *Clinical Geriatric Medicine* 25:155-169.

Delbanco, S.F., K.M. Anderson, C.E. Major, M.B. Kiser & B.W. Toner. 2011. *Promising payment reform: Risk-sharing with accountable care organizations* (Commonwealth Fund No. 1530). Retrieved from https://www.chcs.org/media/Creating_ACOs_in_Medicaid.pdf

DeVore, S. & R.W. Champion. 2011. 'Driving population health through accountable care organizations.' *Health Affairs* 30:41-50.

DuGoff, E.H., S. Dy, E.R. Giovannetti, B. Leff & C.M. Boyd. 2013. 'Setting standards at the forefront of delivery system reform: Aligning care coordination quality measures for multiple chronic conditions.' *Healthcare Quality* 35:58–69

Feldman, R. 2015. 'The economics of provider payment reform: Are accountable care organizations the answer?' *Journal of Health Politics, Policy and Law* 40:745-760.

Friedberg, M.W., P.G. Chen, C. White, O. Jung, L. Raaen, S. Hirshman, E. Hoch, C. Stevens, P.B. Ginsburg, L.P. Casalino, M. Tutty, C. Vargo & L. Lipinski. 2015. *Effects of health care payment models on physician practice in the United States*. Retrieved from https://www.rand.org/pubs/periodicals/health-quarterly/issues/v5/n1/08.html

Geisinger Health System. 2017. *Community Health Needs Assessment Update: 2017*. Retrieved from https://www.geisinger.org//media/OneGeisinger/pdfs/ghs/about-geisinger/chna/2017-reports/ghs-chna-2017.pdf?la=en

Gilfillan, R.J., J. Tomcavage & M.B. Rosenthal. 2010. 'Value and the medical home: Effects of transformed primary care.' *The American Journal of Managed Care* 16:607-614.

Gleeson, S., K. Kelleher & W. Gardner. 2016. 'Evaluating a pay-for-performance program for Medicaid children in an accountable care organization.' *JAMA Pediatrics* 170:259-266.

Harris, J.M., I. Elizondo & A.M. Brow. 2016. 'Orchestrating ACO success: How top performers achieve shared savings.' *Healthcare Financial Management* 70:42-50.

Hayen, A.P., M.J. van den Berg, B.R. Meijboom, J.N. Struijs & G.P. Westert. 2015. 'Incorporating shared savings programs into primary care: From theory to practice.' *BMC Health Services Research* 15:1-15.

Herrel, L.A., E.C. Norton, S.R. Hawken, Z. Ye, B.K. Hollenbeck & D.C. Miller. 2016. 'Early impact of Medicare accountable care organizations on cancer surgery outcomes.' *Cancer* 122:2739-2746.

Hildebrandt, H. 2014a. *Crossing the boundaries of medical care towards regional public health*. Presentation, Bucharest.

Hildebrandt, H. 2014b. 'Crossing the boundaries from individual medical care to regional public health outcomes: The triple aim of Gesundes Kinzigtal – better health + improved care + affordable costs.' *International Journal of Integrated Care* 14:1-2.

Hildebrandt, H., C. Hermann & R. Knittel. 2010. 'Gesundes Kinzigtal integrated care: Improving population health by a shared health gain approach and a shared savings contract.' *International Journal of Integrated Care* 10:e046-e061.

Hildebrandt, H., T. Schulte & B. Stunder. 2012. 'Triple aim in Kinzigtal, Germany: Improving population health, integrating health care and reducing costs of care – lessons for the UK?' *Journal of Integrated Care* 20:205-222.

Howard, S.W., S.L. Bernell, J. Yoon & J. Luck. 2014. 'Oregon's Coordinated Care Organizations: A promising and practical reform model.' *Journal of Health Politics, Policy and Law* 39:933-940.

Howard, S.W., S.L. Bernell, J. Yoon, J. Luck & C.M. Ranit. 2015. 'Oregon's experiment in health care delivery and payment reform: Coordinated Care Organizations replacing managed care.' *Journal of Health Politics, Policy and Law* 40:246-255.

Huskamp, H.A., S.F. Greenfield, E.A. Stuart, J.M. Donohue, K. Duckworth, E.M. Kouri, Z. Song, M.E. Chernew & C.L. Barry. 2016. 'Effects of global payment and accountable care on tobacco cessation service use: An observational study.' *Journal of General Internal Medicine* 31:1134-11140.

Kaufman, B.G., B.S. Spivack, S.C. Stearns, P.H. Song & E.C. O'Brien. 2019. 'Impact of accountable care organizations on utilization, care, and outcomes: A systematic review.' *Medical Care Research and Review* 76:255-290.

Kelleher, K.J., J. Cooper, K. Deans, P. Carr, R.J. Brilli, S. Allen & W. Gardner. 2015. 'Cost savings and quality of care in a pediatric accountable care organization.' *Pediatrics* 135(3):e582-e589.

Kim, D.H., C. Lloyd, D.K. Fernandez, A. Spielman & D. Bradshaw. 2017. 'A direct experience in a new accountable care organization: Results, challenges, and the role of the neurosurgeon.' *Neurosurgery* 80:S42-S49.

Kinosian, B., G. Taler, P. Boling & D. Gilden. 2016. 'Projected savings and workforce transformation from converting Independence at Home to a Medicare benefit.' *Journal of the American Geriatrics Society* 64:1531-1536.

Larson, B.K., A.D. van Citters, S.A. Kreindler, K.L. Carluzzo, J.N. Gbemudu, F.M. Wu, E.C. Nelson, S.M. Shortell & E.S. Fisher. 2012. 'Insight from transformations under way at four Brookings-Dartmouth accountable care organization pilot sites.' *Health Affairs* 31:2395-2406.

Lupianez-Villanueva, F. & A. Theben. 2014. *Strategic intelligence monitor on personal health systems phase 3: Gesundus Kinzigtal (Germany) case study report* (Report EUR 27057 EN). Sevilla, Spain: European Commission, Joint Research Centre, Institute for Prospective Technological Studies.

Maeng, D.D., T.R. Graf, D.E. Davis, J. Tomcavage & F.J. Bloom 2012. 'Can a patient-centered medical home lead to better patient outcomes? The quality implications of Geisinger's ProvenHealth Navigator.' *American Journal of Medical Quality* 27:210-216.

Maeng, D.D., N. Khan, J. Tomcavage, T.R. Graf, D.E. Davis & G.D. Steele. 2015. 'Reduced acute inpatient care was largest savings component of Geisinger Health System's patient-centered medical home.' *Health Affairs* 34:636-644.

Makni, N., A. Rothenburger & K. Kelleher. 2015. 'Survey of twelve children's hospital-based accountable care organizations.' *Journal of Hospital Administration* 4:64-73.

Markovich, P. 2012. 'A global budget pilot project among provider partners and Blue Shield of California led to savings in first two years.' *Health Affairs* 9:1969-1976.

McConnell, K.J., A.M. Chang, D.J. Cohen, N. Wallace, M.E. Chernew, G. Kautz, D. McCarty, B. McFarland, B. Wright & J. Smith. 2014. 'Oregon's Medicaid transformation: An innovative approach to holding a health system accountable for spending growth.' *Healthcare* 2:163-167.

McConnell, K.J. 2016. 'Oregon's Medicaid Coordinated Care Organizations.' *The Journal of the American Medical Association* 315:869-870.

McConnell, K.J., S. Renfro, B.K. Chan, T.H. Meath, A. Mendelson, D. Cohen, J. Waxmonsky, D. McCarty, N. Wallace & R.C. Lindrooth. 2017. 'Early performance in Medicaid accountable care organizations: A comparison of Oregon and Colorado.' JAMA *Internal Medicine* 177:538-545.

McGinnis, T. & D.M. Small. 2012. *Accountable care organizations in Medicaid: Emerging practices to guide program design* (Policy Brief Center for Health Care Strategies, February 2012). Retrieved from https://www.chcs.org/media/Creating_ACOs_in_Medicaid.pdf

McWilliams, J.M., B.E. Landon & M.E. Chernew. 2013. 'Changes in health care spending and quality for Medicare beneficiaries associated with a commercial ACO contract.' *The Journal of the American Medical Association* 310:829-836.

McWilliams, J.M., B.E. Landon, M.E. Chernew & A.M. Zaslavsky. 2014. 'Changes in patients' experiences in Medicare accountable care organizations.' *The New England Journal of Medicine* 371:1715-1724.

McWilliams, J.M., L.A. Hatfield, M.E. Chernew, B.E. Landon & A. Schwartz, A. 2016. 'Early performance of accountable care organizations in Medicare.' *The New England Journal of Medicine* 374:2357-2366.

McWilliams, J.M., L.G. Gilstrap, D.G. Stevenson, M.E. Chernew, H.A. Huskamp & D.C. Grabowski. 2017. 'Changes in postacute care in Medicare Shared Savings Program.' *The Journal of the American Medical Association Internal Medicine* 177:518-526.

Mechanic, R.E., P. Santon, B.E. Landon & M.E. Chernew. 2011. 'Medical group responses to global payment: Early lessons from the 'Alternative Quality Contract' in Massachusetts.' *Health Affairs* 9:1734-1742.

Meyer, H. 2012. 'Many accountable care organizations are now up and running, if not off to the races.' *Health Affairs* 31:2363-2367.

Minnesota Department of Human Services Health Care Administration. 2017. *Request for proposals for a qualified grantee to provide health care services to medical assistance and Minnesota care enrolees under alternative payment arrangements through the Integrated Health Partnerships (IHP) demonstration*. Retrieved from https://mn.gov/dhs/assets/2017-ihp-rfp_tcm1053-294430.pdf

Nationwide Children's. 2016. *Partners for Kids: Saving money by improving health for our most vulnerable children*. Retrieved from https://www.nationwidechildrens.org/impact-quality/partners-for-kids-pediatric-accountable-care

O'Halloran, K., A. Depalma, V. Joseph, N. Cobelli & A. Sharan. 2012. 'The role of accountable care organizations in delivering value.' *Current Reviews in Musculoskeletal Medicine* 5:283-289.

OECD. 2016. *Better ways to pay for health care*. Paris, France: OECD Health Policy Studies, OECD Publishing.

Ouayogodé, M., C.H. Colla & V.A. Lewis. 2017. 'Determinants of success in shared savings programs: Analysis of ACO and market characteristics.' *Healthcare* 5:53-61.

Pimperl, A., H. Hildebrandt, O. Groen, T. Schulte, I. Meyer & M. Wetzel. 2017a. *Case study: Gesundes Kinzigtal, Germany*. Durham, North Carolina: Duke University Margolis Center for Health Policy.

Pimperl, A., T. Schulte, A. Mühlbacher, M. Rosenmöller, B. Busse, O. Groene, H.P. Rodriguez & H. Hildebrandt. 2017b. 'Evaluating the impact of an accountable care organization in population health: The quasi-experimental design of the German Gesundes Kinzigtal.' *Population Health Management* 20:239-248.

Rodin, D. & S. Silow-Caroll. 2013. *Medicaid payment and delivery reform in Colorado: ACOs at the regional level* (Commonwealth Fund No. 1666, vol 11). Retrieved from http://www.commonwealthfund.org/publications/case-studies/2013/mar/colorado-medicaid-payment

Rotenberg, J., B. Kinosian, P. Boling & G. Taler. 2018. 'Home-based primary care: Beyond extension of the Independence at Home Demonstration.' *Journal of the American Geriatrics Society* 66:812-818.

Salako, A., X. Zhu, C. MacKinney, F. Ullrich & K. Mueller. 2015. *Characteristics of rural accountable care organizations (ACOs) – a survey of Medicare ACOs with rural presence* (Rural Policy Brief No. 2015, 8 May). Retrieved fromhttps://www.ruralcenter.org/resource-library/characteristics-of-rural-acos-%E2%80%93-a-survey-of-medicare-acos-with-rural-presence

Schwartz, A.L., M.E. Chernew, B.E. Landon & J.M. McWilliams. 2015. 'Changes in low-value services in year 1 of the Medicare Pioneer accountable care organization program.' *JAMA Internal Medicine* 175:1815-1825.

Scott, A., M. Liu & J. Yong. 2018. 'Financial incentives to encourage value-based health care.' *Medical Care Research and Review* 75:3-32.

Serrano, C., M. Ferrer & A. Toner. 2009. 'Alzira model: Hospital de la Ribera, Valencia, Spain.' In: B. Rechel, J. Erskine, B. Dowdeswell et al. (eds.), *Capital investment for health: Case studies from Europe*, 11-25. Copenhagen, Denmark: WHO Publications.

Sharp, A.L., Z. Song, D.G. Safran, M.E. Chernew & M.A. Fendrick. 2013. 'The effect of bundled payment on emergency department use: Alternative Quality Contract effects after year one.' *Academic Emergency Medicine* 20:961-964.

Shields, M. 2011. 'From clinical integration to accountable care.' *Annals of Health Law* 20:151-164.

Shields, M.C., P.H. Patel, M. Manning & L. Sacks. 2011. 'A model for integrating independent physicians into accountable care organizations.' *Health Affairs* 30:161-172.

Shortell, S.M., C.H. Colla, V.A. Lewis, E. Fisher, E. Kessell & P. Ramsay. 2015. 'Accountable care organizations: The national landscape.' *Health Politics, Policy and Law* 40:647-668.

Silow-Caroll, S., J.N. Edwards & D. Rodin. 2013. *How Colorado, Minnesota, and Vermont are reforming care delivery and payment to improve health and lower costs* (Commonwealth Fund No. 1665, vol 10). Retrieved from http://www.commonwealthfund.org/~/media/Files/Publications/Case%20Study/2013/Mar/1665_Silow-Carroll_Medicaid_synthesis_FINAL_v2.pdf

Song, Z. & B.E. Landon. 2012. 'Controlling health care spending: The Massachusetts experiment.' *The New England Journal of Medicine* 366:1560-1561.

Song, Z., D.B. Safran, B.E. Landon, Y. He, R.P. Ellis, R.E. Mechanic, M.P. Day & M.E. Chernew. 2011. 'Health care spending and quality in year 1 of the Alternative Quality Contract.' *The New England Journal of Medicine* 365:909-918.

Song, Z., D.B. Safran, B.E. Landon, M.B. Landrum, Y. He, R.E. Mechanic, M.P. Day & M.E. Chernew. 2012. 'The Alternative Quality Contract in Massachusetts, based on global budgets, lowered medical spending and improved quality.' *Health Affairs* 31:1885-1894.

Song, Z., A.M. Fendrick, D.G. Safran, B.E. Landon & M.E. Chernew. 2013. 'Global budgets and technology-intensive medical services.' *Healthcare* 1:15-21.

Song, Z., S. Rose, D.G. Safran. 2014. 'Changes in health care spending and quality 4 years into global payment.' *The New England Journal of Medicine* 371:1704-1714.

Song, Z., S. Rose, M.E. Chernew & D.G. Safran. 2017. 'Lower- versus higher-income populations in the Alternative Quality Contract: Improved quality and similar spending.' *Health Affairs* 36:74-82.

Struckmann, V., W. Boerma & E. Ginneken. 2015. *The Gesundes Kinzigtal programme, Germany.* Retrieved from http://www.icare4eu.org/pdf/Gesundes_Kinzigtal.pdf

Struckmann, V., W. Quentin, R. Busse & E. van Ginneken. 2016. *How to strengthen financing mechanisms to promote care for people with multimorbidity in Europe?* (Policy Brief ICARE4EU 24). Utrecht, the Netherlands: Nivel.

Stuart, E.A., C.L. Barry, J.M. Donohue, S.F. Greenfield, K. Duckworth, Z. Song, R. Mechanic, E.M. Kouri, C. Ebnesajjad, M.E. Chernew & H.A. Huskamp. 2017. 'Effects of accountable care and payment reform on substance use disorder treatment: Evidence from the initial three years of the Alternative Quality Contract.' *Addiction* 112:124-133.

The Advisory Board Company. 2016. *The business of population health management: Gesundes Kinzigtal GmbH, Black Forest, Germany*. Washington, USA: Advisory Board International.

Weier, R.C., W. Gardne, K. Conkol, K. Pajer & K.J. Kelleher. 2017. 'Partners for Kids care coordination: Lessons from the field.' *Pediatrics* 139:S109-S116.

Williams, J. 2013. 'A new model for care population management.' *Healthcare Financial Management* 67:69-76.

Winblad, U., V. Mor, J.P. McHugh & M. Rahman. 2017. 'ACO-affiliated hospitals reduced rehospitalizations from skilled nursing facilities faster than other hospitals.' *Health Affairs* 36:67-73.

Zimmerman, M. 2015. *Medicaid payment and delivery system innovation: Minnesota's experience*. Presentation, Minnesota.

# Chapter 5

## How to manage financial risk for capitated primary care providers? The impact of care package, risk adjustment, risk sharing, and patient panel size

With Frank Eijkenaar

Preparing for submission

## ABSTRACT

To strengthen incentives for cost-consciousness, capitation payment models for primary care providers (PCPs) covering comprehensive care packages are becoming increasingly popular. A key question is how to design these models in such a way to keep financial risk manageable for PCPs and prevent unintended consequences. This paper assesses the relative impact of four key determinants of PCPs' financial risk: (1) scope of the care package, (2) sophistication of risk adjustment, (3) risk sharing, and (4) patient panel size. Using rich administrative data (N=4.2 million individuals), we simulate capitation payments and assess the impact on financial risk for both real and simulated PCPs. Our simulations show that the scope of the care package has the greatest impact. Furthermore, irrespective of panel size, improving the risk adjustment and applying high-cost risk sharing sharply decreases risk, particularly for more comprehensive packages. To notably reduce financial risk, increases in panel size should be substantial.

## 1. INTRODUCTION

A strong primary care system in which primary care physicians (PCPs) play a central, coordinating role is an essential pillar of value-based health care (Kringos et al. 2013; McClellan et al. 2010; Cattel et al. 2020a; Cattel & Eijkenaar 2020b). In many countries including Canada, the Netherlands, Norway, Spain, and the United Kingdom (UK) a patient rostering model was introduced in which individuals formally register with a PCP. In these settings, PCPs are responsible for managing cost and quality of care for their patients. Typically, PCPs are the first point of contact for patients in need of care, prescribe medication, function as gatekeepers to non-emergency care provided in hospitals, and remain involved in later stages of care trajectories (Singh et al. 2019; Bodenheimer et al. 1999). As such, PCPs have considerable influence over costs and quality across the care continuum (Vermaas 2006).

While the pivotal roles envisioned for PCPs in healthcare systems remain largely undisputed, predominant models for paying PCPs generally match poorly with these roles. Specifically, PCPs' payments often still depend strongly on the number of services provided and only cover core primary care services. This rewards volume instead of value and discourages integration across the care continuum (Landon 2014; Westert et al. 2014; Berenson & Rich 2010; Hayen et al. 2015). Policymakers and payers have therefore been exploring alternative payment models (Cattel & Eijkenaar 2020b; Bazemore et al. 2018; Struijs et al. 2020; APMF FPT Work Group 2016; Chernew et al. 2020). A growing number of these models rely on comprehensive capitation payments in which providers receive a prospectively determined fixed amount for each enrolled individual in their practice, covering a specified care package for a defined period (Vlaanderen et al. 2019; Scott et al. 2018; Cattel & Eijkenaar 2020b). Increasingly, these payments do not only pertain to primary care services, but also to prescription medication and secondary care, distinguishing them from conventional primary care capitation. Under such models, (groups of) PCPs will typically function as main contracting entity for payers and employ or subcontract other providers required for delivering the covered care services. Examples from practice include General Practitioner (GP) Fundholding in the UK, the Gesundes Kinzigtal project in Germany, the Menzis Shared Savings Program in the Netherlands, some Patient-Centered Medical Homes in the United States (US), and the payment options under the new CMS Primary Cares Initiative in the US (Ross 2019; CMS 2019; Cattel & Eijkenaar 2020b).

A key characteristic of comprehensive capitation payments for PCPs is that – because of their prospective nature and the care package stretching beyond primary care services – PCPs are exposed to greater amounts of financial risk for medical spending than under conventional payment models. This provides them with incentives to act cost-consciously in providing primary care services, prescribing medication and referring to secondary care, and to actively take up their role as coordinators and efficiently organize care processes across the care continuum (Jegers et al. 2002; Miller 2009; Robinson 2001b; Berenson 2010; Anderson & Weller 1999; Hayen et al. 2015; Hayen et al. 2021; Frakt & Mayes 2012). For example, PCPs facing financial risk for

follow-up care are encouraged to critically review the possibilities for treating patients in primary care settings instead of in more expensive hospital settings.

Although exposure to more financial risk strengthens PCPs' incentives for cost-consciousness, a potential danger is that PCPs might be exposed to too much financial risk, which could threaten PCPs' financial viability and lead to provider opposition, low participation rates, and undesired behavior like quality skimping and risk selection. Prior studies have shown that these are not just theoretical concerns (Kay 2002; Frakt & Mayes 2012; Ellis 1998). Therefore, a key question in this regard is how to keep financial risk manageable for PCPs, while maintaining incentives for cost-consciousness. Answering this question requires insight in the determinants of financial risk and the interplay between them. The objective of this paper is to gain this insight by examining the relative impact of four key elements in the design of capitation payments on financial risk for PCPs: (1) the scope of the care package covered by the payment, (2) the sophistication of risk adjustment of the payment, (3) the application of risk sharing, and (4) the size of PCPs' patient panels. Although the direction of the effect of each element separately is clear (section 2), no prior study has analyzed these elements simultaneously regarding their relative impact on financial risk in the context of primary care payment reform.

Using rich administrative data on medical spending and risk characteristics of over 4.2 million individuals enrolled with a large Dutch health insurer, we simulate prospective capitation payments and analyze financial risk for both real PCPs in the Netherlands and simulated, larger PCP entities with substantially larger patient panels. Our study contributes to the body of knowledge concerning smarter choices in the design of provider payment systems. Specifically, this study could help those involved in primary care payment reform in making better-informed decisions regarding design and appropriate levels of financial risk for providers.

This paper is organized as follows. The next section provides a conceptual framework on financial risk in relation to the four payment design elements analyzed in this paper. Section 3 describes the data and methods used, and section 4 presents the results. We conclude with a discussion of our findings.

## 2. CONCEPTUAL FRAMEWORK

In the context of provider payment incentives, financial risk concerns the question 'who bears the financial consequences of healthcare spending at the margin, and to what extent'. Clearly, the relative shares of financial risk borne by the payer (e.g., an insurer) and the provider (e.g., a PCP) depend heavily on the payment model in place, which in turn influences the incentives faced by the provider. For example, providers paid by fee-for-service face limited financial risk because each provided service is reimbursed separately, with limited incentives to contain costs as a result. In contrast, capitation payments expose providers to greater amounts of risk because providers receive a fixed periodic payment per enrolled person that is independent of the amount

of care provided. At the margin, providers are therefore fully risk bearing for the care covered by the payment, resulting in strong incentives for cost-conscious behavior.

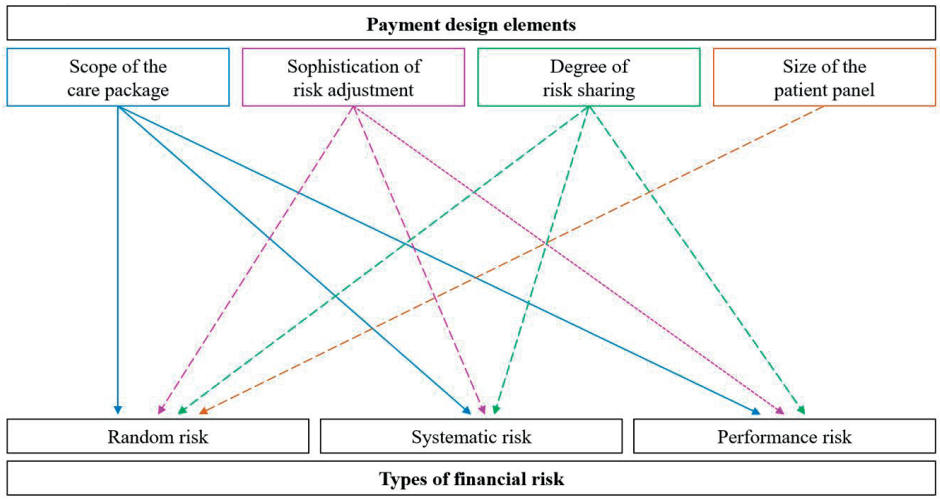
Medical spending and the variation therein are the result of the random nature of the occurrence of health problems (random risk), systematic differences in health risk and behavior between individuals (systematic risk), and provider behavior (performance risk) (Vermaas 2006). When transferring financial risk to providers to introduce incentives for cost-consciousness, providers are ideally only confronted with performance risk, with the payer retaining all random risk and systematic risk. This is because performance risk is the only type of risk that providers can directly influence and thus can reasonably be held accountable for.[12] Unfortunately, in practice it is virtually impossible to unravel the various types of risk because medical spending generally has multiple overlapping causes that are difficult to observe and attribute. For example, it is often difficult to assess the extent to which spending is the result of something that could not have been predicted, the natural course of the disease, the health and behavior of the patient, or actions of the provider. Thus, though theoretically the first-best choice, splitting the risk and transferring only performance risk to providers is unfeasible in practice.

Because placing providers fully at risk for spending under a comprehensive care package is neither likely to be viable (e.g., because of provider opposition) nor desirable (i.e., because of possible unwanted provider bankruptcies and strategic provider behavior like quality skimping and risk selection), payment reform efforts usually rely on financial risk sharing between payers and providers. Though second-best because providers also become liable for a portion of random and systematic risk, risk sharing is feasible in practice and still introduces incentives for cost-conscious behavior (Miller 2009; Vermaas 2006; De Brantes & Rastogi 2008; Frakt & Mayes 2012).

The magnitude of the financial risk providers are exposed to depends on various elements related to the design of the payment model (Spector et al. 2015, 2018; Cattel et al. 2020a; Vermaas 2006). This paper focuses on the relative impact on PCP-level financial risk of four key elements (Figure 5.1).

12 Note that to some extent providers can influence future health risk by effective prevention and treatment. In addition, providers may be able to influence patients' health behavior, for example by encouraging a healthy lifestyle and compliance with treatment plans. Insofar providers can influence such behavior, related spending can be considered performance risk. If not, this spending can be considered random risk (if the behavior is idiosyncratic) or systematic risk (if the behavior is related to certain patient characteristics).

**Figure 5.1.** Conceptual framework describing the effects of four payment design elements on three types of financial risk




Note: Solid dash = positive impact on risk. Long dash = negative impact on risk. Squared dots = unknown impact on risk.

## 2.1 Scope of the care package

All else equal, random risk will be higher for a PCP accountable for a comprehensive care package including primary care, prescription medication, and hospital care than for a PCP accountable for narrow a set of services (e.g., primary care only). This is because spending variation tends to be higher for pharmaceutical and hospital care than for primary care. Systematic risk is likely to be higher too because more complex care is included in the package, resulting in increased disease severity and more pronounced differences in individual-level health risk. Finally, performance risk is also higher as the financial consequences of (in)efficient care are larger (Vermaas 2006; Spector et al. 2018).

## 2.2 Sophistication of risk adjustment

Risk adjustment accounts for predictable variation in spending due to differences in the risk profile of providers' patient panels. In this way, risk adjustment contributes to fair and accurate allocation of payments and reduces systematic risk, thereby mitigating incentives for cherry picking profitable patients and dumping unprofitable ones (Newhouse et al. 1997; Anderson & Weller 1999; Ash & Ellis 2012; Rose et al. 2016; Spector et al. 2018). In addition, in principle performance risk is a positive function of the sophistication of the risk-adjustment model because better risk adjustment encourages providers to focus on cost-consciousness. However, when the model uses risk factors based on historical spending/utilization (which tends to be the case in practice), these incentives are mitigated because providers with low past spending/utilization are penalized relative to providers with high past spending/utilization. Depending on the design,

risk adjustment may also negatively impact random risk; all else equal, a model using risk factors based on historical spending/utilization confronts providers with less random risk than a model that does not contain such factors.

### 2.3 Degree of risk sharing

The distribution of healthcare spending is highly skewed to the right, with a small share of the population accounting for a disproportionately large share of total spending. In addition, the (unexplained) variation in spending is particularly large in the tail. For example, a recent study found that while the top 1% of patients accounts for approximately 20% of all spending, these patients are responsible for almost 75% of unexplained variance under sophisticated risk adjustment (McGuire et al. 2020b). Consequently, even with sophisticated risk adjustment, caring for relatively many of these high-cost patients is likely to lead to high levels of risk for providers. Risk sharing can help protecting providers against such excessive risk. With risk sharing, the payer instead of the provider accounts for some share of high spending. Because the risk of large random shocks in spending is (at least partly) shifted to the payer, providers' random risk reduces. In addition, systematic risk reduces if predictable (health-related) spending not accounted for by the risk-adjustment model is part of the risk-sharing arrangement. Finally, risk sharing also reduces performance risk as it is based on actual spending.

### 2.4 Size of the patient panel

Due to the law of large numbers, random risk decreases as panel size increases (Christianson & Conrad 2011; Van de Ven 2014; Vermaas 2006). Therefore, random risk is lower for a large group of PCPs caring for many patients than for an individual PCP with a small patient panel. Ceteris paribus, providers with large panels are better equipped to spread risk and absorb random (unpredictable) spending shocks compared to providers with small panels (Frakt & Mayes 2012; Spector et al. 2018).

When it comes to provider payment reform, the importance of carefully considering (the interaction between) these four elements to keep financial risk manageable and prevent the reform from defeating their aims, is broadly recognized in theory and practice (Hayen et al. 2015; Shortell et al. 2014; Ash & Ellis 2012; Cattel et al. 2020a). For example, panel size requirements tend to be stricter when providers are held accountable for more comprehensive care packages. In the GP Fundholding initiative in the UK (1991-1997), (consortia of) GP practices received payments covering not only primary care provided by physicians, but also elective surgery, diagnostic tests and examinations, outpatient referrals, community health services, and prescription medication. Participation in the initiative was initially restricted to practices with panels of at least 11,000 patients. In the UK Total Purchasing Pilot (1994-1997), budgets covering virtually the entire continuum of care were introduced for providers with at least 20,000 registered patients (Lewis 2004). In the US Medicare Shared Savings Program, providers' upside risk (for savings) is maximized at 70% and assuming downside risk (for losses) is voluntary, while in the Medicare Next

Generation ACO Model providers assume either 80% or 100% risk for both savings and losses; consequently, the minimum required panel size is lower for the former than for the latter initiative, i.e., 5,000 versus 10,000 Medicare beneficiaries (CMS 2017; Cattel & Eijkenaar, 2020b). Finally, in the new CMS Primary Cares Initiative in the US two basic pathways are created with varying levels of financial risk for participating providers. In the pathway for providers with small patient panels ('Primary Care First') providers can lose up to 10% of their revenue but gain as much as 50%, while in the pathway for providers with large panels ('Direct Contracting') providers are accountable for 50% to 100% of both savings and losses (Ross 2019; CMS 2019).

## 3. DATA AND METHODS

### 3.1 Study setting and data

This study was conducted using data collected in the context of the Dutch healthcare system, which is based on the principles of regulated competition among insurers and among providers. Competition occurs on price and quality, while the government imposes regulation to enforce the public objectives of quality, accessibility, and affordability (Van Kleef et al. 2018). An important feature of the Dutch health insurance system is a sophisticated risk-adjustment model compensating insurers for predictable variation in individual medical spending. Without risk adjustment, insurers would be confronted with strong incentives for risk selection given that they are not allowed to risk-rate their premiums (ibid.).

In the Netherlands, virtually all citizens are registered with a single PCP or primary care practice and patient panels are more or less fixed (Kringos et al. 2013). Citizens typically have a long-term relationship with their PCP who acts as gatekeeper to secondary care. In 2017, there were over 12,000 registered PCPs, working in over 5,000 practices with an average panel size of around 2,200 patients per full-time PCP (Versteeg & Batenburg 2017). While in the past most PCPs worked in solo practices, they are increasingly working in groups which also comprise other primary care providers; while in 2001 33% of all Dutch PCPs worked in a solo practice, in 2017 this was 17% (NZa 2012; Versteeg & Batenburg 2017; Van der Velden et al. 2017). This trend is consistent with the development in other countries towards large(r) provider entities comprising multiple (primary care) disciplines. In the Netherlands, spending on primary care provided by physicians accounts for approximately 4% of total medical spending covered by the comprehensive basic benefit package under the Health Insurance Act (LHV 2018).

For our study, we relied on two large administrative datasets, which could be merged at the individual level using a unique identification key that was anonymized by a trusted third party. First, we used individual-level administrative data on medical spending and health risk. This

dataset, which was originally composed and used for calculating insurers' risk-adjusted capitation payments for the year 2015, contains various categories of spending in 2012 and the set of risk-adjustor variables included in the risk-adjustment model of 2015. Specifically, the data contain information on all somatic spending covered by the basic benefit package, divided into the following categories: primary care provided by physicians (primarily registration fees and fees for office and home visits), primary care diagnostics provided in hospitals and treatment centers (e.g., laboratory tests), physiotherapy, durable medical equipment (e.g., insulin infusion pumps), prescription medication, and hospital care. In addition, the data include the following risk characteristics (see Van Kleef et al. 2018 for details): age (20 classes), gender, socio-economic status based on household income (4 classes based on deciles of the income distribution), source of income (6 classes), pharmacy-based cost groups (25 PCGs based on prior use of medication prescribed for chronic illnesses), diagnosis-based cost groups (15 DCGs based on diagnoses from certain hospital treatments in the prior year), and multiple-year high-cost groups (7 MYHCGs based on high spending in the three prior years). PCGs, DCGs, and MYHCGs are morbidity-based characteristics and can be considered as direct proxies for health.

The second dataset contains individual-level data obtained from a large Dutch health insurer with information on the PCP that individuals were registered with in 2012. The data include a unique provider identifier representing a PCP, as well as the 4-digit zip code of the address of the PCP's practice. In our data, a PCP can be an individual physician, a group practice, or a health center (i.e., an entity in which multiple PCPs and other primary care providers provide and coordinate care, usually from the same building). The data represent real PCP patient panels from 2012, and include all individuals enrolled with the insurer that provided the data. Importantly, people who did not visit their PCP could still be identified in our data because Dutch PCPs receive a fixed registration fee for every individual enrolled in their practice, regardless of actual healthcare utilization.

Merging both datasets at the individual level resulted in approximately 4.5 million individuals registered with approximately 7,000 PCPs. In the merged data, we dropped individuals with missing data on medical spending, risk characteristics, and/or PCP, as well as individuals registered with PCPs with panel sizes smaller than 100 or larger than 5,000 patients. The lower threshold is common in practice and in the literature (Eijkenaar & Van Vliet 2014) and resulted in removal of 1,346 PCPs. The upper threshold is the result of visual inspection of the panel size distribution of PCPs showing that the vast majority of PCPs had panel sizes of maximally 5,000 patients, with 13 larger entities as clear outliers.[13] The final sample contains roughly 5,600 PCPs that served approximately 4.2 million patients in 2012.

13 The 13 outlier PCPs appear to be parent companies with practices located across the country and are therefore not comparable with the typical PCP in the data.

## 3.2 Analyses

This section describes our approach to simulate prospective capitation payments for PCPs and assesses how PCP-level financial risk depends on the scope of the care package covered by the payment, the sophistication of risk adjustment, the application of risk sharing, and patient panel size. Our approach comprised the following five steps.

### *3.2.1 Step 1: Constructing care packages*

In the first step, we constructed four different care packages by summing individual-level, annualized spending on various types of care covered by the basic benefits package of the Health Insurance Act. We only analyzed types of care for which the actions of PCPs will dictate, at least in part, their patients' journey through the healthcare system. Package 1 (P1) includes all spending on primary care provided by physicians. Package 2 (P2) equals P1 supplemented with spending on primary care diagnostics, physiotherapy, and durable medical equipment. For package 3 (P3), we added spending on prescription medication to P2. Finally, package 4 (P4) is P3 supplemented with spending on hospital care. P1 reflects a narrow primary care package representing about 7% of total somatic spending in 2012 covered by the basic benefit package, while P4 is the most comprehensive package covering approximately 93%.

Next, for each package we defined 'at risk' spending for PCPs for two scenarios: no risk sharing and a form of high-cost risk sharing between payer and PCPs. No risk sharing implies that PCPs are 100% accountable for all spending under the relevant package.[14] In contrast, with high-cost risk sharing the payer accounts for some preset proportion of spending above a certain threshold. In this paper, we applied high-cost risk sharing with 100% of spending above a threshold set at the 99$^{th}$ percentile of the relevant spending distribution being reimbursed by the payer. In our data, this implies thresholds of €500 (P1), €2,898 (P2), €5,604 (P3), and €20,892 (P4).

### *3.2.2 Step 2: Simulating risk-adjusted capitation payments*

The first step resulted in eight definitions of 'at risk' spending (i.e., four care packages with and without risk sharing). Next, we calculated individual-level predicted annual spending (i.e., simulated annual capitation payments) for each of these eight definitions while applying no, simple, and sophisticated risk adjustment (so 8 x 3 = 24 configurations in total). 'No risk adjustment' implies that the payment per individual equals the grand mean spending as observed in the data for the relevant package. For simple and sophisticated risk adjustment, we ran linear regression models predicting annualized spending for each package with and without risk sharing (i.e., eight dependent variables) from two sets of risk adjustors (independent variables) using individuals' duration of the insurance contract in 2012 for analytic weighting. 'Simple risk adjustment' only uses sociodemographic information (i.e., age interacted with gender, socioeconomic status, and

14 Although we acknowledge that many forms of risk sharing are possible, for ease of presentation/interpretation we chose to include just two risk-sharing scenarios in our simulations (including the scenario 'no risk sharing').

source of income), while 'sophisticated risk adjustment' also includes the morbidity indicators PCGs, DCGs, and MYHCGs. In constructing the models, we mirrored the structure of the Dutch risk-adjustment model 2015 used for calculating insurers' payments, mainly because P4 approximates the package for which Dutch health insurers are financially accountable. To enable meaningful comparison, the (definitions of) risk-adjustor variables were kept the same across all models.[15]

At both the individual level and the level of PCPs, we calculated R-squared values for each package (with and without risk sharing) while applying simple or sophisticated risk adjustment. High R-squared values at the PCP level would suggest that a substantial proportion of PCP-level variation in spending can be explained by differences in the risk profiles of PCPs' patient panels, which would underline the importance of adjusting PCPs' capitation payments for these risk factors (Ash & Ellis 2012).

### *3.2.3 Step 3: Measuring financial risk at the PCP level*

In this paper, our primary measure of financial risk is the standard deviation of residual spending at the PCP level (Ash & Ellis 2012). For each of the 24 payment configurations, we first calculated individual-level residual spending by subtracting predicted spending generated by the relevant regression model (or the grand mean in case of no risk adjustment) from actual spending. Next, for each PCP we calculated the mean residual spending using the residual spending of the individuals registered with each PCP. Finally, we calculated the mean of the mean residual spending across all PCPs. The standard deviation of this mean of means constitutes our measure of PCP-level financial risk.

We also examined financial risk by analyzing PCPs' 'risk of ruin' under each payment configuration. Following Layton & McGuire (2016), we defined risk of ruin as the probability of a PCP suffering a 'catastrophic loss'. First, we calculated the mean financial result (i.e., profit or loss) for each PCP in our sample. Next, in case of a loss (i.e., the mean actual spending for a PCP exceeds the mean predicted spending or payment) we determined whether this loss could be considered as a 'catastrophic loss', which we defined for P4 as a loss which exceeds the payment by at least 5%. Because the spending levels of P1-P3 are (much) lower than the level of P4, for these packages we increased this percentage commensurate to the decrease in mean (predicted) spending relative to P4.[16] Finally, for each payment configuration we calculated the risk of ruin as the percentage of PCPs with a catastrophic loss.

15 For more information on the Dutch risk-adjustment model for insurer payments, see Van Kleef et al. (2018).

16 For example, with our application of risk sharing the mean spending for P1 is €119 and for P4 €1,493, implying that a loss under P1 will be deemed 'catastrophic' if the loss exceeds the payment by at least 5x1,493/119=63%.

#### *3.2.4 Step 4: Assessing the impact of panel size on financial risk*

To analyze the impact of PCPs' panel size on financial risk, we divided PCPs into five subgroups with roughly the same number of PCPs but with increasing mean panel size, and then calculated financial risk separately for each subgroup.[17] To simulate the effect on financial risk of provider groups increasing further in size because of, for example, future consolidation as a response to increasing financial accountability, we also clustered PCPs based on geographic proximity into three virtual groups of PCPs. We used the 4-digit zip code of PCPs' practice locations and clustered PCPs working in the same neighborhood (i.e., based on all four digits of the zip code; 2,078 simulated entities), area (i.e., based on the first two digits; 90 simulated entities), and region (i.e., based on the first digit; nine simulated entities). We refer to these larger PCP groups as 'ACOs'.[18]

#### *3.2.5 Step 5: Calculating financial-risk ratios*

For ease of interpretation, for each combination of care package, type of risk adjustment, and PCP/ACO subgroup, we divided the standard deviation of PCP-/ACO-level residual spending (i.e., our main measure of financial risk) by the standard deviation of PCP-level residual spending for a baseline configuration, separately for the scenario with and without risk sharing. This baseline configuration is defined as P1 without risk adjustment for all PCPs. Given the data, this configuration best approximates the current Dutch payment model for PCPs. A financial-risk ratio of 0.5 would denote 50% less financial risk compared to baseline, while values of 1.0 and 1.5 would mean equal or 50% greater risk, respectively.

### 3.3 Sensitivity analyses

To determine the sensitivity of our results to analytic choices made, we performed multiple sensitivity checks, focusing on the analyses for real PCPs. First, we redid our analyses on similar data from other years, i.e., 2010 and 2011 instead of 2012 (analysis S1). Second, we limited the analyses to the patients of PCPs with panel sizes between 500 and 2,500 patients instead of between 100 and 5,000, leaving roughly 2,700 PCPs serving 2.9 million individuals (analysis S2). Finally, because including PCPs with significant fluctuations in panel size over time may have distorted our findings and resulted in conclusions that are not generally applicable to all PCPs, we dropped PCPs (and their patients) with panel sizes that increased/decreased by more than 50% between 2011 and 2012, leaving 5,300 PCPs and 4.1 million individuals (analysis S3). To quantify the extent to which our results are affected by these choices, we estimated Pearson correlation coefficients between financial-risk ratios generated by our main analysis and those generated by sensitivity analyses S2 and S3.

---

17 Note that because the data contain information of individuals enrolled with one specific health insurer, PCPs' panel sizes depend on the market share of the insurer in the region where the PCPs' hold practice.

18 We acknowledge that these simulated groups only approximate ACOs in terms of their (large) panel size and not in terms of other characteristics of the provider group and governance structure.

# 4. RESULTS

## 4.1 Descriptive statistics

Table 5.1 shows some descriptive statistics of the sample, which is representative for the total Dutch population in 2012 (N = 16.5 million) in terms of sociodemographic and morbidity-based characteristics (Van Kleef et al. 2017, Table 1). Among the 4.2 million individuals, 19% is categorized in at least one PCG, 9% in a DCG, and 6% in a MYHCG. Mean spending per individual (SD) ranges from €122 (109) for P1 to €1,699 (5,794) for P4 without risk sharing and from €119 (81) for P1 to €1,493 (3,151) for P4 with 100% risk sharing above a threshold set at the 99$^{th}$ percentile of the spending distribution. The standard deviation ranges from 109 for P1 to 5,794 for P4 without risk sharing and from 81 for P1 to 3,151 for P4 with risk sharing. This indicates that risk sharing does not only reduce the mean of at-risk spending, but also the variation around the mean.

**Table 5.1.** Descriptive statistics of the sample at the individual enrollee-level (2012-data)

| | |
|---|---|
| N (unweighted) [a] | 4,249,929 |
| N (weighted) [a] | 4,204,662 |
| **Age/gender** | |
| Male, 0-17 years | 11% |
| Male, 18-34 years | 10% |
| Male, 35-44 years | 7% |
| Male, 45-54 years | 7% |
| Male, 55-64 years | 6% |
| Male, 65 years and older | 8% |
| Female, 0-17 years | 10% |
| Female, 18-34 years | 10% |
| Female, 35-44 years | 7% |
| Female, 45-54 years | 7% |
| Female, 55-64 years | 6% |
| Female, 65 years and older | 10% |
| **Source of income** | |
| Disability benefits | 5% |
| Social security benefits | 3% |
| Student | 3% |
| Self-employed | 4% |
| Other (including employment) | 46% |
| **Socioeconomic status** | |
| Lowest income class (deciles 1-3 of the income distribution) | 31% |
| Middle income class (deciles 4-7 of the income distribution) | 39% |
| Highest income class (deciles 8-10 of the income distribution) | 29% |

**Table 5.1.** Descriptive statistics of the sample at the individual enrollee-level (2012-data) (continued)

| N (unweighted) [a] | 4,249,929 |
|---|---|
| **Morbidity indicators** [b] | |
| In at least 1 pharmacy-based cost group (PCG) | 19% |
| In a diagnosis-based cost group (DCG) | 9% |
| In a multiple-year high-cost group (MYHCG) | 6% |
| **Mean spending in Euros (SD)** [c] | |
| P1 without risk sharing | 122 (109) |
| P1 with risk sharing | 119 (81) |
| P2 without risk sharing | 295 (717) |
| P2 with risk sharing | 272 (430) |
| P3 without risk sharing | 580 (1,679) |
| P3 with risk sharing | 529 (877) |
| P4 without risk sharing | 1,699 (5,794) |
| P4 with risk sharing | 1,493 (3,151) |

Note: SD = standard deviation.

a. Unweighted = number of individuals in the sample. Weighted = number of individuals weighted by the duration of the insurance contract in 2012 (i.e., the number of insured-years).

b. Individuals can be classified in only one DCG per year (i.e., the one with the highest follow-up costs), but in multiple PCGs (Van Kleef et al. 2018).

c. Annualized and weighted by the duration of the insurance contract in 2012. P1 is the sum of spending on primary care provided by physicians. P2 is P1 but supplemented with primary care diagnostics, physiotherapy, and durable medical equipment. P3 is P2 but supplemented with pharmaceutical care. P4 is P3 but supplemented with hospital care. Without risk sharing, PCPs are financially accountable for 100% of spending under the relevant package. With risk sharing, the payer accounts for 100% of spending above a threshold set at the 99$^{th}$ percentile of the spending distribution under the relevant package.

Table 5.2 provides information on the number of PCPs, panel size, and spending with risk sharing at the PCP and ACO level. The mean panel size for the 5,584 PCPs in the full sample is 753 patients with a standard deviation of 662. This is smaller than the panel of a typical Dutch PCP, which can be explained by the fact that we only observe individuals enrolled with one specific insurer (section 3.1). With ascending PCP subgroup number, mean panel size increases, while the number of PCPs included in the subgroups remains roughly constant (i.e., about 1,117 PCPs). Mean spending at the level of all PCPs ranges from €115 for P1 to €1,419 for P4. Clustering PCPs working in the same neighborhood, area, and region leads to a reduction of the number of provider entities from 5,584 to 2,078, 90, and 9, respectively. Clustering to the neighborhood level results in a mean panel size that is a factor 2.7 greater than the mean panel size of real PCPs observed in the data (i.e., 2,023 vs. 753). For clustering to the highest level (region), this factor is 620 (i.e., 467,185 vs. 753).

**Table 5.2.** Number of PCPs/ACOs, panel size, and medical spending with risk sharing at the PCP and ACO level (2012-data) [a]

| | # of PCPs /ACOs | Mean panel size (SD) [c] | Mean spending in Euros (SD) [d] | | | |
|---|---|---|---|---|---|---|
| | | | P1 | P2 | P3 | P4 |
| PCP total | 5,584 | 753 (662) | 115 (16) | 260 (46) | 500 (109) | 1,419 (308) |
| PCP subgroup 1 | 1,116 | 204 (50) | 113 (17) | 246 (48) | 464 (108) | 1,326 (327) |
| PCP subgroup 2 | 1,118 | 347 (40) | 111 (14) | 248 (44) | 471 (100) | 1,347 (292) |
| PCP subgroup 3 | 1,116 | 514 (65) | 113 (16) | 252 (45) | 478 (105) | 1,363 (286) |
| PCP subgroup 4 | 1,117 | 901 (163) | 116 (16) | 269 (46) | 521 (108) | 1,478 (303) |
| PCP subgroup 5 | 1,117 | 1,798 (721) | 123 (14) | 285 (37) | 565 (89) | 1,582 (249) |
| ACO neighborhood [b] | 2,078 | 2,023 (2,288) | 113 (14) | 255 (40) | 486 (93) | 1,386 (260) |
| ACO area [b] | 90 | 46,718 (53,878) | 114 (8) | 258 (26) | 493 (65) | 1,401 (171) |
| ACO region [b] | 9 | 467,185 (342,458) | 116 (6) | 266 (18) | 511 (45) | 1,441 (111) |

Note: PCP = primary care provider. ACO = accountable care organization. SD = standard deviation.

a. With risk sharing, the payer accounts for 100% of spending above a threshold set at the 99$^{th}$ percentile of the spending distribution under the relevant package.

b. To simulate ACOs based on geographic proximity with larger panel sizes than real PCPs observed in the data, PCPs were clustered to the level of neighborhood (i.e., based on all four digits of the zip code of the address of PCPs practice locations), area (i.e., based on the first two digits), and region (i.e., based on the first digit).

c. In insured-years, i.e., weighted by the duration of the insurance contract in 2012.

d. Annualized and weighted by the duration of the insurance contract in 2012. P1 is the sum of spending on primary care provided by physicians. P2 is P1 but supplemented with primary care diagnostics, physiotherapy, and durable medical equipment. P3 is P2 but supplemented with pharmaceutical care. P4 is P3 but supplemented with hospital care.

## 4.2 Predictive power

Table 5.3 shows R-squared values at the individual- and PCP-level for the four care packages with and without risk sharing, while applying simple and sophisticated risk adjustment. Holding the care package constant, values are smaller at the individual level than at the PCP level and for no risk sharing versus high-cost risk sharing. In addition, as expected, the sophisticated risk-adjustment model explains a larger proportion of the variation in PCP-level spending than simple risk adjustment, for all packages and regardless of risk sharing. At both levels but especially the PCP level, R-squared values tend to be considerably higher for the more comprehensive care packages, particularly under sophisticated risk adjustment.

**Table 5.3.** Individual-level and PCP-level R-squared values (x100%) for four packages with and without risk sharing, while applying simple and sophisticated risk adjustment (2012-data) [a]

| | Individual level (N = 4,2 million) | | PCP level (N = 5,584) | |
|---|---|---|---|---|
| Care package [b] | Simple risk adjustment [c] | Sophisticated risk adjustment [c] | Simple risk adjustment [c] | Sophisticated risk adjustment [c] |
| | No \| Yes risk sharing [d] | No \| Yes risk sharing [d] | No \| Yes risk sharing [d] | No \| Yes risk sharing [d] |
| P1 | 11.7 \| 16.9 | 16.2 \| 21.9 | 22.2 \| 22.0 | 30.1 \| 30.2 |
| P2 | 6.3 \| 12.7 | 23.5 \| 32.0 | 48.7 \| 51.5 | 64.9 \| 66.6 |
| P3 | 6.6 \| 19.4 | 35.8 \| 58.1 | 53.4 \| 64.9 | 80.8 \| 86.3 |
| P4 | 4.9 \| 11.6 | 26.2 \| 33.0 | 57.1 \| 65.8 | 75.1 \| 83.6 |

Note: PCP = primary care provider.

a. R-squared = proportion explained variation in spending = 1 minus the sum of squared residuals divided by the total sum of squares.

b. P1 is the sum of spending on primary care provided by physicians. P2 is P1 but supplemented with primary care diagnostics, physiotherapy, and durable medical equipment. P3 is P2 but supplemented with pharmaceutical care. P4 is P3 but supplemented with hospital care.

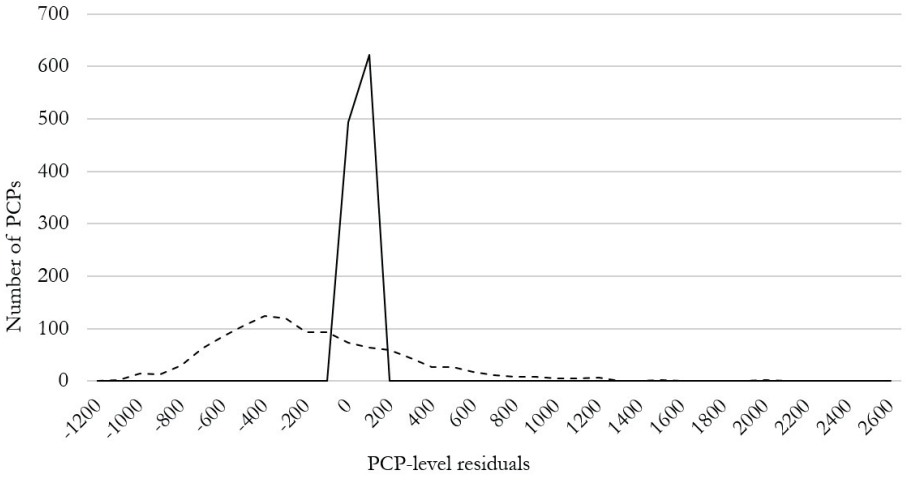
c. Simple risk adjustment includes the risk adjustors age interacted with gender, socioeconomic status, and source of income. Sophisticated risk adjustment includes the same risk adjustors as the simple model but supplemented with the morbidity-based risk adjustors (i.e., pharmacy-based cost groups, diagnosis-based cost groups, and multiple-year high-cost groups).

d. Without risk sharing, PCPs are financially accountable for 100% of spending under the relevant package. With risk sharing, the payer accounts for 100% of spending above a threshold set at the 99$^{th}$ percentile of the spending distribution under the relevant package.

## 4.3 Financial-risk ratios at the PCP level

In this paper, our primary measure of financial risk for PCPs is the standard deviation of mean residual spending at the PCP level. Figure 5.2 shows the distribution of PCP-level residual spending for the two configurations with the smallest and greatest financial risk in our sample. The figure shows a completely different picture for the two configurations. For the configuration with the smallest risk (solid line), the mean residual concentrates near €0 and the range of residuals is limited (SD = €13), while for the configuration with the greatest risk (dashed line) the mean residual is negative (€-217) with a much wider range (SD = €470).

**Figure 5.2.** Distribution of PCP-level residuals for the configuration with the smallest financial risk [a] (solid line), and the configuration with the greatest financial risk [b] (dashed line) (2012-data)



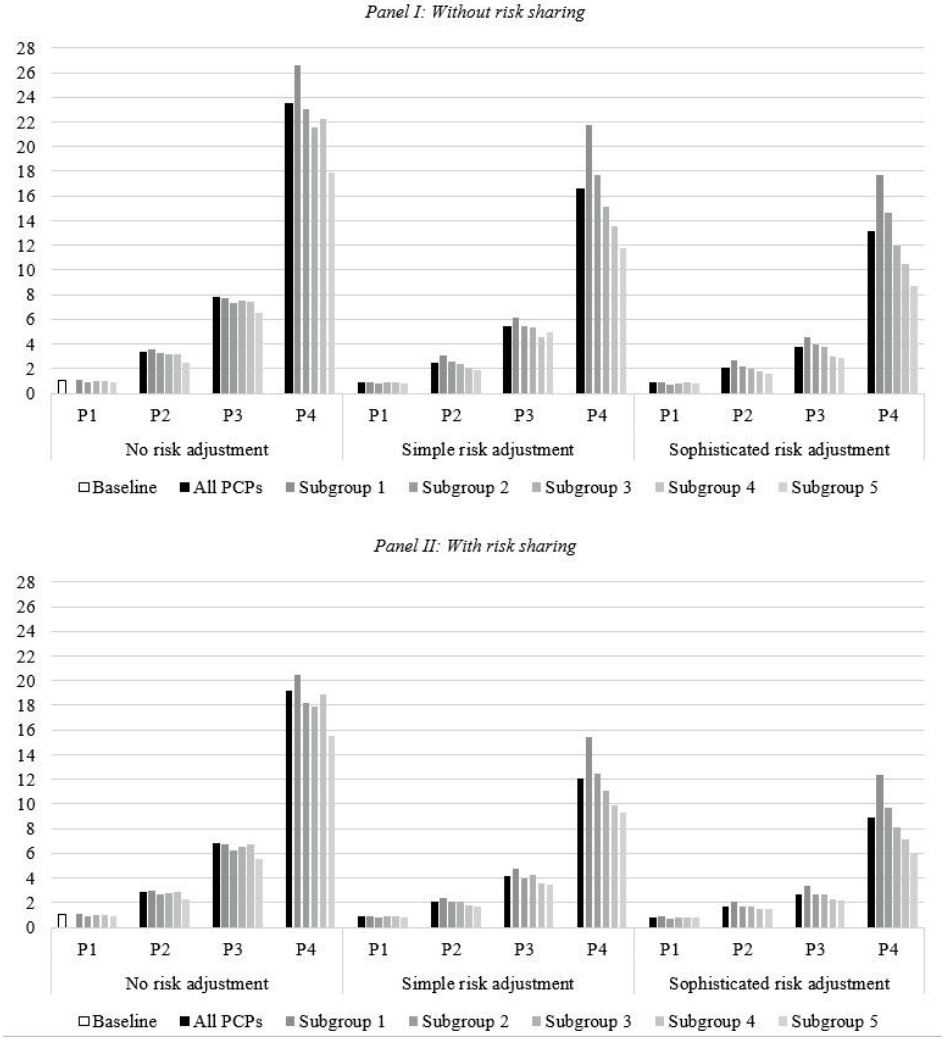

Note: PCP = primary care provider.

a. P1 with 100% risk sharing above a threshold set at the 99th percentile of the spending distribution and sophisticated risk adjustment for the subgroup of PCPs (N = 1,117) with the largest patient panels. The mean residual is €2 with a standard deviation of €13. The kurtosis is 0.7.

b. P4 without risk sharing and risk adjustment for the subgroup of PCPs (N = 1,116) with the smallest patient panels. The mean residual is €-217 with a standard deviation of €470. The kurtosis is 3.4.

Figure 5.3 shows financial-risk ratios for the twelve payment configurations (four care packages and three types of risk adjustment) without risk sharing (panel I) and with risk sharing (panel II). The black bars represent results for all PCPs and the grey bars those for the five subgroups of PCPs with – from left to right – increasing patient panel size. The white, leftmost bar represents the baseline configuration (i.e., P1 and no risk adjustment, for all PCPs) with which the other configurations are compared to calculate the financial-risk ratios.

**Figure 5.3.** Financial-risk ratios relative to baseline [a] for all PCPs (N = 5,584) and five subgroups of PCPs with increasing panel size [b], for four care packages [c] by type of risk adjustment [d] and with and without risk sharing [e] (2012-data)




Note: PCP = primary care provider.

a. Baseline = P1 and no risk adjustment, for all PCPs. A financial-risk ratio of 0.5 denotes 50% lesser financial risk compared to the baseline while values of 1.0 and 1.5 mean similar or 50% greater risk, respectively. For panel I, the absolute value of financial risk (i.e., the standard deviation of mean PCP-level residual spending) for the baseline configuration is €18. Mean spending for all PCPs under P4 is a factor 14 greater than under P1 (i.e., €1,699 vs. €122). For panel II, the absolute value of financial risk for the baseline configuration is €16. Mean spending for all PCPs under P4 is a factor 12 greater than under P1 (i.e., €1,419 vs. €115).

b. The full sample of PCPs is divided into five equally sized subgroups with increasing panel size, ranging from on average 204 patients for subgroup 1 to on average 1,798 patients for subgroup 5.

c. P1 is the sum of spending on primary care provided by physicians. For panel I, average spending is €118 for all PCPs and ranges from €115 for subgroup 1 to €126 for subgroup 5. For panel II, average spending is €115 for all PCPs and ranges from €113 for subgroup 1 to €123 for subgroup 5.

P2 is P1 but supplemented with primary care diagnostics, physiotherapy, and durable medical equipment. For panel I, average spending is €281 for all PCPs and ranges from €265 for subgroup 1 to €310 for subgroup 5. For panel II, average spending is €260 for all PCPs and ranges from €246 for subgroup 1 to €285 for subgroup 5.
P3 is P2 but supplemented with pharmaceutical care. For panel I, average spending is €545 for all PCPs and ranges from €500 for subgroup 1 to €623 for subgroup 5. For panel II, average spending is €500 for all PCPs and ranges from €464 for subgroup 1 to €565 for subgroup 5.
P4 is P3 but supplemented with hospital care. For panel I, average spending is €1,603 for all PCPs and ranges from €1,482 for subgroup 1 to €1,812 for subgroup 5. For panel II, average spending is €1,419 for all PCPs and ranges from €1,326 for subgroup 1 to €1,582 for subgroup 5.

d. Under no risk adjustment, the payment equals the grand mean spending in the data for the specific package. Simple risk adjustment uses the risk adjustors age interacted with gender, socioeconomic status, and source of income to predict individual-level spending, which in turn is used to calculate the payment per PCP. Sophisticated risk adjustment includes the same risk adjustors as the simple model but supplemented with the morbidity-based risk adjustors (i.e., pharmacy-based cost groups, diagnosis-based cost groups, and multiple-year high-cost groups).
e. Without risk sharing, PCPs are financially accountable for 100% of spending under the relevant package. With risk sharing, the payer accounts for 100% of spending above a threshold set at the 99$^{th}$ percentile of the spending distribution under the relevant package.

Figure 5.3 contains five main findings. First, as expected, financial risk is highest for PCPs with relatively small panels (subgroup 1) that are held accountable for the most comprehensive care package (P4) without risk sharing and without risk adjustment (financial-risk ratio = 26, calculated by dividing a SD of €470 by the baseline SD of €18). Correspondingly, financial risk is lowest for PCPs with relatively large panels (subgroup 5), accountability for P1 with risk sharing and sophisticated risk adjustment (financial-risk ratio = 0.8, i.e., a SD of €13 divided by the baseline SD of €16).

Second, of the payment design elements analyzed here, the scope of the care package appears to be the most important determinant of financial risk. Irrespective of whether risk sharing is applied, the type of risk adjustment used, and patient panel size, financial risk increases considerably when the care package becomes more comprehensive and is by far greatest when the package contains spending on hospital care (P4). For example, focusing on all PCPs and with risk sharing but without risk adjustment, financial risk is a factor 19 greater when comparing P4 to P1 (baseline), while mean spending under P4 is 'only' a factor 12 greater than baseline-level spending. Financial risk also increases considerably when pharmaceutical spending is added to the package (P3); with risk sharing but without risk adjustment, financial risk is a factor 7 greater compared to baseline. Without risk sharing, the impact of the scope of the care package is naturally larger. For example, focusing on all PCPs and assuming no risk adjustment, financial risk is a factor 23 greater when comparing P4 to P1, while mean spending under P4 is a factor 14 greater than baseline-level spending.

Third, the impact of risk adjustment is rather limited for narrow packages, irrespective of risk sharing. For all PCPs, financial risk for P1 under simple and sophisticated risk adjustment is very similar to baseline (i.e., no risk adjustment), as indicated by the financial-risk ratios being just below 1. For the more comprehensive care packages, however, improving the risk adjustment

has a more pronounced impact. For example, in the sample comprising all PCPs and with application of risk sharing, the financial-risk ratio for P3 reduces from 7 to 4 to 3 when no, simple, and sophisticated risk adjustment is applied, respectively; for P4, these figures are 19, 12, and 9. Without risk sharing, these figures are 8, 5, 4 for P3 and 24, 17, and 13 for P4. Even with sophisticated risk adjustment and high-cost risk sharing, however, financial risk for P4 is still quite high (i.e., a SD of €143).

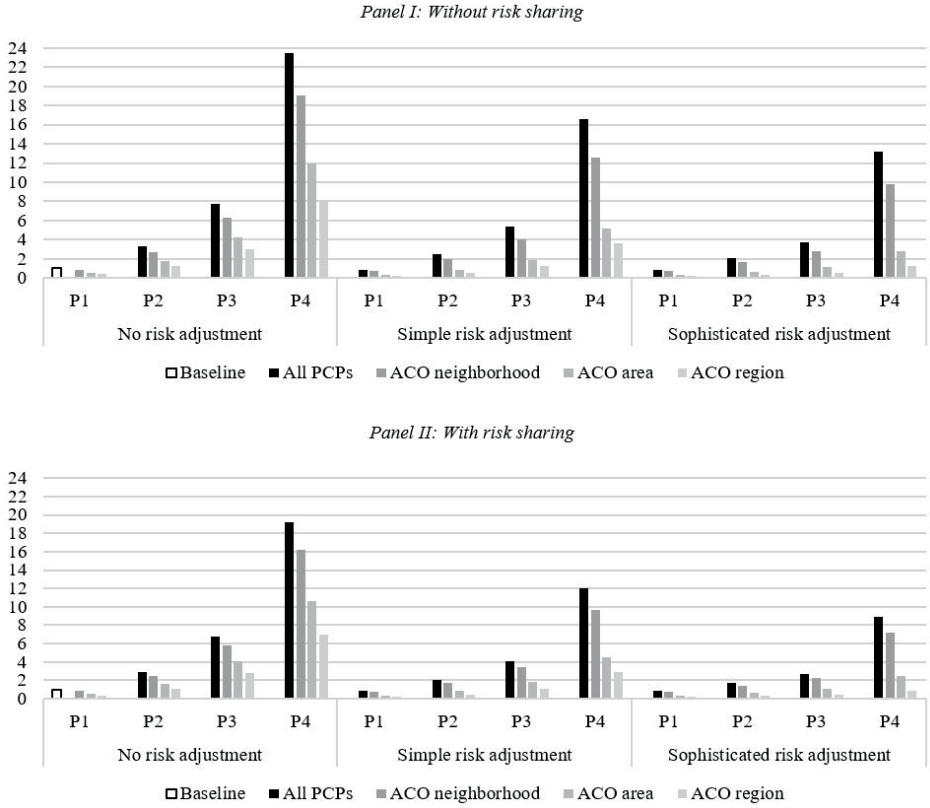
Fourth, comparison of panel I and II in Figure 5.3 shows that risk sharing is an effective method to reduce financial risk, especially for PCPs with small patient panels. The impact on financial risk is most prominent for the care packages covering spending on prescription medication and hospital care (i.e., P3 and P4), which both have a high potential of extreme spending outliers. Interestingly, the effect of risk sharing seems rather independent from the sophistication of the risk adjustment.

Finally, financial risk for PCPs with relatively small panel sizes (i.e., PCP subgroup 1 with 204 patients on average) is consistently greater than for PCPs with relatively large panel sizes (i.e., PCP subgroup 5 with 1,798 patients on average). The impact of panel size is most prominently visible for the more comprehensive packages (i.e., P3 and P4) without risk sharing and with sophisticated risk adjustment. For example, under P4 the financial-risk ratio more than halves when moving from subgroup 1 to subgroup 5, while the reduction is much smaller when looking at narrower packages and/or less sophisticated risk adjustment. Note that the general trend of lower financial risk for larger panels is not consistently discernible when comparing subgroups 2, 3, and 4. The panel sizes of these groups differ in the order of several hundreds of patients, which is apparently too little to significantly affect financial risk.

### 4.4 Financial-risk ratios at the level of simulated ACOs

Figure 5.4 shows financial-risk ratios for the four care packages (with and without risk sharing) and three types of risk adjustment for the three groups of simulated ACOs (grey bars), with mean panel sizes of 2,023 (neighborhood), 46,718 (area), and 467,185 patients (region). For easy comparison, financial-risk ratios for all PCPs as observed in the data (mean panel size = 753 patients) are displayed as well (black bars). The white, leftmost bar again represents the baseline configuration. Results are in line with expectations (section 2) and the findings for real PCPs (Figure 5.3). Financial risk for ACOs is always lower than for real PCPs, underlining the relevance of panel size for financial risk. Relatively large ACOs are confronted with relatively little risk as compared to small ACOs. In panel I (without risk sharing), the financial-risk ratio ranges from 0.2 (region-type ACO, P1, and sophisticated risk adjustment) to 19 (neighborhood-type ACO, P4, and no risk adjustment). In panel II (with risk sharing), these ratios range from 0.2 to 16. The impact of risk sharing is stronger for providers with relatively small patient panels (all PCPs and neighborhood-type ACO) compared to providers with relatively large panels (area-type ACO and region-type ACO), particularly under P4.

**Figure 5.4.** Financial-risk ratios relative to baseline [a] for all PCPs (N = 5,584) and three types of simulated ACOs with increasing size [b], for four care packages [c] by type of risk adjustment [d] and with and without risk sharing [e] (2012-data)




Note: PCP = primary care provider. ACO = accountable care organization.

a. Baseline = P1 and no risk adjustment, for all PCPs. A financial-risk ratio of 0.5 denotes 50% lesser financial risk compared to the baseline while values of 1.0 and 1.5 mean similar or 50% greater risk, respectively. For panel I, the absolute value of financial risk (i.e., the standard deviation of mean PCP-level residual spending) for the baseline configuration is €18. Mean spending for all PCPs under P4 is a factor 14 greater than under P1 (i.e., €1,699 vs. €122). For panel II, the absolute value of financial risk for the baseline configuration is €16. Mean spending for all PCPs under P4 is a factor 12 greater than under P1 (i.e., €1,419 vs. €115).

b. To simulate ACOs based on geographic proximity with larger panel sizes than real PCPs observed in the data, PCPs were clustered to the level of neighborhood (i.e., based on all four digits of the zip code of the address of PCPs practice locations, mean panel size = 2,023 patients), area (i.e., based on the first two digits, mean panel size = 46,718 patients), and region (i.e., based on the first digit, mean panel size = 467,185 patients).

c. P1 is the sum of spending on primary care provided by physicians. For panel I, average spending is €118 for all PCPs and ranges from €116 for the smallest ACOs to €119 for the largest ACOs. For panel II, average spending is €115 for all PCPs and ranges from €113 for the smallest ACOs to €116 for the largest ACOs.
P2 is P1 but supplemented with primary care diagnostics, physiotherapy, and durable medical equipment. For panel I, average spending is €281 for all PCPs and ranges from €276 for the smallest ACOs to €288 for the largest ACOs. For panel II, average spending is €260 for all PCPs and ranges from €255 for the smallest ACOs to €266 for the largest ACOs.

P3 is P2 but supplemented with pharmaceutical care. For panel I, average spending is €545 for all PCPs and ranges from €528 for the smallest ACOs to €556 for the largest ACOs. For panel II, average spending is €500 for all PCPs and ranges from €486 for the smallest ACOs to €511 for the largest ACOs.
P4 is P3 but supplemented with hospital care. For panel I, average spending is €1,603 for all PCPs and ranges from €1,562 for the smallest ACOs to €1,630 for the largest ACOs. For panel II, average spending is €1,419 for all PCPs and ranges from €1,386 for the smallest ACOs to €1,441 for the largest ACOs.

d. Under no risk adjustment, the payment equals the grand mean spending in the data for the specific package. Simple risk adjustment uses the risk adjustors age interacted with gender, socioeconomic status, and source of income to predict individual-level spending, which in turn is used to calculate the payment per PCP. Sophisticated risk adjustment includes the same risk adjustors as the simple model but supplemented with the morbidity-based risk adjustors (i.e., pharmacy-based cost groups, diagnosis-based cost groups, and multiple-year high-cost groups).
e. Without risk sharing, PCPs are financially accountable for 100% of spending under the relevant package. With risk sharing, the payer accounts for 100% of spending above a threshold set at the 99$^{th}$ percentile of the spending distribution under the relevant package.

## 4.5 Risk of ruin

In addition to measuring financial risk by the standard deviation of residual spending at the PCP level, we also measure PCPs' and ACOs' risk of ruin under each of the 24 different payment configurations. For simplicity, we only show results for all PCPs and for two types of simulated ACOs.

As shown in Table 5.4, for all payment configurations risk of ruin is smaller for ACOs than for PCPs. Risk of ruin is (close to) 0% for P1, irrespective the type of risk adjustment or whether risk sharing is applied. For P2, risk of ruin is also relatively low, but not negligible for the group of all PCPs when applying no or simple risk adjustment (i.e., ranging from 1.8% to 5.6% depending on whether risk sharing is applied). In contrast, for P3 and particularly P4 risk of ruin is substantial (i.e., above 20%) for PCPs and ACOs with relatively small patient panels. More than a quarter of all PCPs (i.e., approximately 1,400 PCPs) would suffer a catastrophic loss under P4, even with risk sharing and sophisticated risk adjustment.

Table 5.4 also shows that in contrast to risk sharing (of which the impact on risk of ruin seems rather limited), risk adjustment strongly reduces risk of ruin, especially under P3 and P4. For example, for P3 with risk sharing, risk of ruin for all PCPs reduces from 16.9% to 8.9% to 4.4% when applying no, simple, and sophisticated risk adjustment. The impact of risk adjustment tends to be larger for P3 than for P4, except when panel sizes are substantial (i.e., ACO area). In that case, the effect of panel size seems to 'boost' the effect of risk adjustment on risk of ruin.

**Table 5.4.** Risk of ruin (in %) for all PCPs (N = 5,584) and two types of simulated ACOs, for four care packages [a] by type of risk adjustment [b] and with and without risk sharing [c] (2012-data)

| | No risk adjustment | | | | Simple risk adjustment | | | | Sophisticated risk adjustment | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **P1** | **P2** | **P3** | **P4** | **P1** | **P2** | **P3** | **P4** | **P1** | **P2** | **P3** | **P4** |
| **Without risk sharing** | | | | | | | | | | | | |
| All PCPs | 0.1 | 5.6 | 17.2 | 31.3 | 0.1 | 3.2 | 11.7 | 29.0 | 0.1 | 2.1 | 7.7 | 27.9 |
| ACO neighborhood [d] | 0.1 | 2.6 | 12.3 | 25.2 | 0.1 | 1.0 | 6.3 | 22.8 | 0.1 | 0.7 | 4.6 | 23.0 |
| ACO area [e] | 0.0 | 0.0 | 3.3 | 24.4 | 0.0 | 0.0 | 0.0 | 5.6 | 0.0 | 0.0 | 0.0 | 1.1 |
| **With risk sharing** | | | | | | | | | | | | |
| All PCPs | 0.1 | 4.3 | 16.9 | 29.9 | 0.1 | 1.8 | 8.9 | 27.6 | 0.0 | 1.1 | 4.4 | 25.9 |
| ACO neighborhood [d] | 0.1 | 1.8 | 12.4 | 25.1 | 0.1 | 0.6 | 4.5 | 20.5 | 0.1 | 0.6 | 2.7 | 20.9 |
| ACO area [e] | 0.0 | 0.0 | 2.2 | 21.1 | 0.0 | 0.0 | 0.0 | 4.4 | 0.0 | 0.0 | 0.0 | 1.1 |

Note: PCP = primary care provider. ACO = accountable care organization.

a. P1 is the sum of spending on primary care provided by physicians. P2 is P1 but supplemented with primary care diagnostics, physiotherapy, and durable medical equipment. P3 is P2 but supplemented with pharmaceutical care. P4 is P3 but supplemented with hospital care.

b. Under no risk adjustment, the payment equals the grand mean spending in the data for the specific package. Simple risk adjustment uses the risk adjustors age interacted with gender, socioeconomic status, and source of income to predict individual-level spending, which in turn is used to calculate the payment per PCP. Sophisticated risk adjustment includes the same risk adjustors as the simple model but supplemented with the morbidity-based risk adjustors (i.e., pharmacy-based cost groups, diagnosis-based cost groups, and multiple-year high-cost groups).

c. Without risk sharing, PCPs are financially accountable for 100% of spending under the relevant package. With risk sharing, the payer accounts for 100% of spending above a threshold set at the 99$^{th}$ percentile of the spending distribution under the relevant package.

d. The mean panel size for the 2,078 ACOs is 2,023 patients.

e. The mean panel size for the 90 ACOs is 46,718 patients.

## 4.6 Sensitivity analyses

To check the sensitivity of our results based on 2012-data, we redid our analyses on data from 2011 and 2010 (S1). Differences in terms of sociodemographic and morbidity-based risk characteristics between the three data years are limited and the same analyses on less recent years yield very similar results in terms of financial risk as presented in the sections above (see Appendix A).

In addition, we reproduced our main results having made different analytic choices (see Appendix B). Restricting the estimation sample to patients of PCPs with panel sizes between 500 and 2,500 patients (S2) did not lead to noticeable differences in results, except that – unsurprisingly – the impact of panel size becomes less prominent. Restricting the sample to PCPs with panel sizes that did not increase or decrease by more than 50% between 2011 and 2012 (S3) also resulted in highly similar findings. Pearson correlation coefficients between the standard deviations of PCP-level residual spending based on the main analysis and those based on the sensitivity analyses S2 and S3 are (very close to) 1 for the various payment configurations, confirming the robustness of our overall findings to analytic choices made.

# 5. DISCUSSION

## 5.1 Summary

Alternative primary care payment models increasingly rely on comprehensive capitation payments stretching beyond primary care services. Under these models, PCPs are exposed to more financial risk than under conventional payment models with the goal to strengthen incentives for cost-consciousness and improve alignment with the roles envisioned for PCPs in the healthcare system. By means of the design of the payment model and taking PCPs' patient panel sizes into account, financial risk can be 'calibrated' to levels that can be considered reasonable in a certain context. This study aimed to provide insight in the impact on PCP-level financial risk of four key determinants: (1) scope of the care package, (2) sophistication of risk adjustment, (3) risk sharing, and (4) patient panel size.

In our simulations, the scope of the care package had the greatest impact on financial risk. Financial risk for the narrower packages covering primary care, physiotherapy, and durable medical equipment is relatively limited, as illustrated by a risk of ruin ranging from 0% to maximally 6% depending on the exact configuration. Irrespective of whether risk sharing is applied, the type of risk adjustment used, and patient panel size, adding prescription medication and even more so hospital care to the package increases financial risk sharply.

In addition, our simulations show sophisticated risk adjustment to be an effective risk-reducing tool, especially for the more comprehensive care packages. This is also shown by high PCP-level R-squared values. Without morbidity-based risk adjustment financial accountability for broad care packages would undesirably expose PCPs to substantial amounts of systematic risk. This finding is consistent with Ash & Ellis (2012), who find that 42% of the variation in PCP-level spending on primary care services plus a portion of spending on prescription medication and hospital care can be explained by a simple demographic risk-adjustment model and 72% by a sophisticated morbidity-based model. We find R-squared values ranging from 22% to 86%, depending on the care package and whether simple or sophisticated risk adjustment is applied. Our finding of a particularly strong impact of morbidity-based risk adjustment for a package covering prescription medication is probably related to the high predictive power of the PCG risk adjustor for pharmaceutical spending. Although a sophisticated model does good work for the package that also covers hospital care, the decrease in financial risk compared to adjustment for socio-demographic factors only is less prominent than for a package not covering hospital care.

To a lesser extent than risk adjustment, high-cost risk sharing can also be effective in mitigating PCPs' financial risk, especially in case of relatively small patient panels and accountability for hospital care (although the effect of on risk of ruin was limited). Importantly, however, combining risk adjustment and risk sharing did not guarantee low levels of financial risk: even with sophisticated risk adjustment and full risk sharing for the top-1% spenders, financial risk associated with the package covering hospital care remains high with more than a quarter of all PCPs in our sample expected to suffer a catastrophic loss.

When it comes to panel size, we found the negative effect on financial risk to be most prominent for comprehensive care packages in combination with sophisticated risk adjustment and without risk sharing. Our results do suggest, however, that increases in panel size should be substantial to have a meaningful impact. In our simulations several hundreds of additional patients did not notably reduce risk, so at this range the law of large numbers does not seem to work effectively. Instead, increases in the order of several thousands of patients are required, which according to our findings would also 'boost' the negative effect of risk adjustment on risk of ruin.

### 5.2 Limitations

This study has several limitations. First, we used administrative data from just one insurer, which means that our results are not based on PCPs' complete patient panels. In a context with multiple competing payers, however, this does not seem problematic. In fact, we believe that simulating comprehensive capitation payments for a share of PCPs' patients, as we have done in this paper, reflects current practice well as it is unlikely that all payers would simultaneously adopt the same alternative payment model.

Second, in our analysis we assumed that PCPs' patient panels are more or less fixed, with limited room for patients to switch PCP and thus limited room for risk selection. This implies that our results are to some extent determined by the specific distribution of patients over the PCPs in our data. In a system in which switching PCP would be common and risk selection thus possible, PCPs would be able to influence their ex-ante financial risk by, for example, attracting patients with low (expected) spending. This scenario, however, was beyond the scope of this paper.

Third, our data did not allow for distinguishing between different types of PCP practices and more specific types of care services. Consequently, we were unable to construct more refined care packages (for example by 'carving-out' high-cost services such as cancer treatments and expensive medication or by restricting hospital care to ambulatory-care sensitive conditions) or to restrict the payment to practices with certain characteristics, such as group practices.

Fourth, besides the financial aspects that were the focus of this paper, other more qualitative considerations (e.g., the need for PCPs to develop entrepreneurial skills) are important as well when considering capitation models for PCPs. In addition, we have examined just one specific form of risk sharing (i.e., high-cost risk sharing) in comparison to no risk sharing. In practice, however, many other applications of high-cost risk sharing as well as other forms of risk sharing are possible (Cattel et al. 2020a), potentially with different effects on financial risk. These other forms and applications, however, were also beyond this paper's scope.

Finally, we simulated capitation payments under the assumption of a contract duration of one year. Expectedly, however, using multiyear contracting also has an impact on PCPs' financial risk, for example through the opportunity of financing potential losses with accumulated financial reserves and spreading these losses over time.

## 5.3 Implications for policy and future research

An important implication of this study is that stakeholders involved in primary care payment reform should carefully decide on the scope of the care package covered by the payment, while accounting for the available possibilities regarding risk adjustment and risk sharing and at given patient panel sizes. After all, this decision has serious consequences for PCPs' financial risk and thereby their incentives for (un)desired behavior. Specifically, accountability for spending on pharmaceutical care but especially hospital care is only warranted if morbidity-based risk adjustment is technically and practically possible and patient panel sizes are substantial. But even then, extensive risk sharing will likely be required to protect providers from excessive levels of residual random and systematic risk that is beyond their control. Examples of risk-sharing measures are a low spending threshold above which risk sharing kicks in, applying risk corridors, and/or limiting 'at risk' spending for PCPs to a portion and/or specific types of hospital care (Ash & Ellis 2012; Cattel et al. 2020a). If the payment can only be adjusted using sociodemographic information, PCPs' financial accountability should be limited to primary care (possibly supplemented with a relatively small portion of pharmaceutical care).

Another important implication of this study is that panel sizes should be substantial (i.e., in the order of tens of thousands of patients per PCP) in order to get the law of large numbers to work effectively, especially under more comprehensive care packages. Regardless of the package, PCPs with smaller panels are confronted with significant levels of risk, and for these PCPs, several hundreds of additional patients will not lead to notable risk reductions. These findings underline the importance for payers to use panel size requirements when implementing comprehensive capitation payments and may imply that implementing these payments in competitive healthcare markets is most realistic in large, urbanized areas. In turn, panel size requirements could stimulate the development of larger provider entities that are better equipped to bear risk and reduce wasteful spending. In this regard, insofar applicable, issues related to competition and antitrust enforcement (e.g., sufficient choice possibilities for patients) should be taken into account, particularly in relatively small competitive healthcare markets, such as that in the Netherlands.

In this paper we have studied the impact of four key determinants on financial risk, under specific assumptions. For various reasons, analysis of PCP-level financial risk for other determinants (e.g., contract duration), more refined forms and applications of the determinants studied here (e.g., care package and risk sharing), and alternative assumptions (e.g., room for patients to switch PCPs) was beyond the scope of this paper but remains an interesting topic for future research.

Another important avenue for further exploration is how risk adjustment can be tailored to the specific purpose of provider payment. Importantly, the risk adjustors used in this study were developed over a period of more than two decades for risk adjusting the capitation payments for insurers, not providers. Arguably, incentives emanating from the same risk-adjustment model are not identical for insurers and providers and, consequently, not all risk-adjustor variables may be equally suitable. For example, both incentives and opportunities for gaming might be stronger

for providers than for insurers because the risk profile of the relatively small patient panel is likely to have more impact on both payment and financial result. When stakes are high, providers might be tempted to change (coding of) medication dosages and/or diagnoses such that more patients receive a morbidity flag, with the aim of receiving a higher payment (Eijkenaar & Van Vliet 2014). Although ethical constraints on such behavior may be stronger for providers than for insurers, potential issues such as these should be taken into account when designing risk adjustment for the specific purpose of provider payment

# APPENDIX

## Appendix A: Results using 2010- and 2011-data (S1)

**Table A.5.1.** Descriptive statistics of the samples at the individual enrollee-level (2010- and 2011-data)

| | 2010-data | 2011-data |
|---|---|---|
| N (unweighted) [a] | 4,207,551 | 4,171,108 |
| N (weighted) [a] | 4,170,356 | 4,136,500 |
| **Age/gender** | | |
| Male, 0-17 years | 11% | 11% |
| Male, 18-34 years | 10% | 10% |
| Male, 35-44 years | 7% | 7% |
| Male, 45-54 years | 7% | 7% |
| Male, 55-64 years | 6% | 6% |
| Male, 65 years and older | 7% | 7% |
| Female, 0-17 years | 10% | 10% |
| Female, 18-34 years | 10% | 10% |
| Female, 35-44 years | 7% | 7% |
| Female, 45-54 years | 7% | 7% |
| Female, 55-64 years | 6% | 6% |
| Female, 65 years and older | 10% | 10% |
| **Source of income** | | |
| Disability benefits | 5% | 5% |
| Social security benefits | 3% | 3% |
| Student | n.a. [b] | 3% |
| Self-employed | 4% | 4% |
| Other (including employment) | 50% | 47% |
| **Socioeconomic status** | | |
| Lowest income class (deciles 1-3 of the income distribution) | 31% | 32% |
| Middle income class (deciles 4-7 of the income distribution) | 39% | 38% |
| Highest income class (deciles 8-10 of the income distribution) | 29% | 29% |
| **Morbidity indicators** [c] | | |
| In at least 1 pharmacy-based cost group | 18% | 18% |
| In a diagnosis-based cost group | 9% | 9% |
| In a multiple-year high-cost group | 6% | 6% |
| **Mean observed spending in Euros (SD)** [d] | | |
| P1 without risk sharing | 121 (112) | 136 (112) |
| P1 with risk sharing | 118 (89) | 134 (91) |
| P2 without risk sharing | 291 (688) | 316 (709) |
| P2 with risk sharing | 268 (420) | 293 (433) |
| P3 without risk sharing | 593 (1,559) | 616 (1,651) |

**Table A.5.1.** Descriptive statistics of the samples at the individual enrollee-level (2010- and 2011-data) (continued)

| | 2010-data | 2011-data |
|---|---|---|
| P3 with risk sharing | 544 (903) | 567 (906) |
| P4 without risk sharing | 1,680 (5,146) | 1,763 (5,628) |
| P4 with risk sharing | 1,506 (3093) | 1,566 (3152) |

Note: SD = standard deviation.

a. Unweighted = number of individuals in the sample. Weighted = number of individuals weighted by the duration of the insurance contract in 2010 and 2011 (i.e. the number of insured-years).
b. N.a. = Not available because the risk equalization model did not include this risk class.
c. Individuals can be classified in only one DCG per year (i.e. the one with the highest follow-up costs), but in multiple PCGs (Van Kleef et al. 2018).
d. Annualized and weighted by the duration of the insurance contract in 2010 and 2011. P1 is the sum of spending on primary care provided by physicians. P2 is P1 but supplemented with primary care diagnostics, physiotherapy, and durable medical equipment. P3 is P2 but supplemented with pharmaceutical care. P4 is P3 but supplemented with hospital care. Without risk sharing, PCPs are financially accountable for 100% of spending under the relevant package. With risk sharing, the payer accounts for 100% of spending above a threshold set at the 99$^{th}$ percentile of the spending distribution under the relevant package.

**Table A.5.2.** Number of PCPs/ACOs, panel size, and medical spending with risk sharing at the PCP and ACO level (2010-data) [a]

| | # of PCPs | Mean panel size (SD) [c] | Mean spending in Euros (SD) [d] | | | |
|---|---|---|---|---|---|---|
| | | | P1 | P2 | P3 | P4 |
| PCP total | 5,581 | 747 (670) | 116 (15) | 257 (48) | 515 (114) | 1,430 (306) |
| PCP subgroup 1 | 1,115 | 195 (46) | 115 (19) | 246 (19) | 485 (129) | 1,354 (330) |
| PCP subgroup 2 | 1,119 | 330 (40) | 112 (14) | 243 (42) | 481 (101) | 1,341 (287) |
| PCP subgroup 3 | 1,115 | 498 (65) | 114 (15) | 250 (43) | 493 (101) | 1,372 (272) |
| PCP subgroup 4 | 1,116 | 900 (179) | 117 (14) | 265 (45) | 537 (111) | 1,489 (299) |
| PCP subgroup 5 | 1,116 | 1,814 (709) | 121 (13) | 281 (38) | 579 (93) | 1,597 (256) |
| ACOs neighborhood [b] | 2,083 | 2,002 (2,340) | 114 (12) | 252 (39) | 501 (94) | 1,395 (257) |
| ACOs area [b] | 90 | 46,337 (53,817) | 115 (6) | 255 (25) | 509 (64) | 1,412 (169) |
| ACOs region [b] | 9 | 463,373 (344,009) | 117 (4) | 262 (16) | 526 (43) | 1,455 (106) |

Note: PCP = primary care provider. ACO = accountable care organization. SD = standard deviation.

a. With risk sharing, the payer accounts for 100% of spending above a threshold set at the 99$^{th}$ percentile of the spending distribution under the relevant package.
b. To simulate ACOs based on geographic proximity with larger panel sizes than real PCPs observed in the data, PCPs were clustered to the level of neighborhood (i.e. based on all four digits of the zip code of the address of PCPs practice locations), area (i.e. based on the first two digits), and region (i.e. based on the first digit).
c. In insured-years, i.e. weighted by the duration of the insurance contract in 2010.
d. Annualized and weighted by the duration of the insurance contract in 2010. P1 is the sum of spending on primary care provided by physicians. P2 is P1 but supplemented with primary care diagnostics, physiotherapy, and durable medical equipment. P3 is P2 but supplemented with pharmaceutical care. P4 is P3 but supplemented with hospital care.

**Table A.5.3.** Number of PCPs/ACOs, panel size, and medical spending with risk sharing at the PCP and ACO level (2011-data) [a]

| | # of PCPs | Mean panel size (SD) [c] | Mean spending in Euros (SD) [d] | | | |
|---|---|---|---|---|---|---|
| | | | P1 | P2 | P3 | P4 |
| PCP total | 5,556 | 745 (661) | 130 (17) | 281 (48) | 536 (113) | 1,490 (311) |
| PCP subgroup 1 | 1,111 | 198 (47) | 127 (18) | 268 (51) | 503 (114) | 1,393 (321) |
| PCP subgroup 2 | 1,112 | 334 (40) | 125 (15) | 266 (43) | 502 (103) | 1,411 (295) |
| PCP subgroup 3 | 1,111 | 501 (64) | 127 (16) | 273 (47) | 515 (107) | 1,440 (291) |
| PCP subgroup 4 | 1,111 | 895 (171) | 131 (16) | 288 (48) | 556 (114) | 1,546 (306) |
| PCP subgroup 5 | 1,111 | 1,795 (702) | 138 (15) | 307 (39) | 604 (92) | 1,662 (254) |
| ACOs neighborhood [b] | 2,071 | 1,992 (2,312) | 128 (14) | 275 (42) | 522 (98) | 1,457 (269) |
| ACOs area [b] | 90 | 45,961 (53,516) | 129 (8) | 279 (27) | 530 (67) | 1,470 (172) |
| ACOs region [b] | 9 | 459,611 (342,751) | 132 (6) | 287 (18) | 548 (45) | 1,512 (111) |

Note: PCP = primary care provider. ACO = accountable care organization. SD = standard deviation.

a. With risk sharing, the payer accounts for 100% of spending above a threshold set at the 99$^{th}$ percentile of the spending distribution under the relevant package.

b. To simulate ACOs based on geographic proximity with larger panel sizes than real PCPs observed in the data, PCPs were clustered to the level of neighborhood (i.e. based on all four digits of the zip code of the address of PCPs practice locations), area (i.e. based on the first two digits), and region (i.e. based on the first digit).

c. In insured-years, i.e. weighted by the duration of the insurance contract in 2011.

d. Annualized and weighted by the duration of the insurance contract in 2011. P1 is the sum of spending on primary care provided by physicians. P2 is P1 but supplemented with primary care diagnostics, physiotherapy, and durable medical equipment. P3 is P2 but supplemented with pharmaceutical care. P4 is P3 but supplemented with hospital care.

**Table A.5.4.** Individual-level and PCP-level R-squared values (x100%) [a] for four packages [b] with and without risk sharing [c], while applying simple and sophisticated risk adjustment [d] (2010- and 2011-data)

| | 2010-data | | | | 2011-data | | | |
|---|---|---|---|---|---|---|---|---|
| | Individual level | | PCP level | | Individual level | | PCP level | |
| | Simple RA | Sophisticated RA | Simple RA | Sophisticated RA | Simple RA | Sophisticated RA | Simple RA | Sophisticated RA |
| | *Without risk sharing* | | | | | | | |
| **P1** | 13.6 | 19.5 | 17.7 | 25.5 | 15.9 | 20.8 | 21.7 | 29.3 |
| **P2** | 7.4 | 25.7 | 49.9 | 65.5 | 7.1 | 24.5 | 47.0 | 63.8 |
| **P3** | 8.4 | 40.6 | 56.4 | 81.8 | 7.5 | 37.4 | 53.2 | 80.3 |
| **P4** | 6.3 | 29.0 | 59.8 | 77.6 | 5.5 | 28.4 | 56.7 | 75.8 |
| | *With risk sharing* | | | | | | | |
| **P1** | 16.7 | 22.6 | 17.4 | 24.7 | 19.6 | 24.1 | 21.0 | 28.8 |
| **P2** | 14.5 | 34.2 | 52.8 | 67.3 | 14.1 | 33.0 | 49.2 | 65.4 |
| **P3** | 20.4 | 59.2 | 66.3 | 86.1 | 20.3 | 58.7 | 63.4 | 85.3 |
| **P4** | 13.0 | 34.2 | 67.1 | 84.2 | 12.7 | 34.0 | 65.8 | 83.9 |

Note: PCP = primary care provider. RA = risk adjustment.

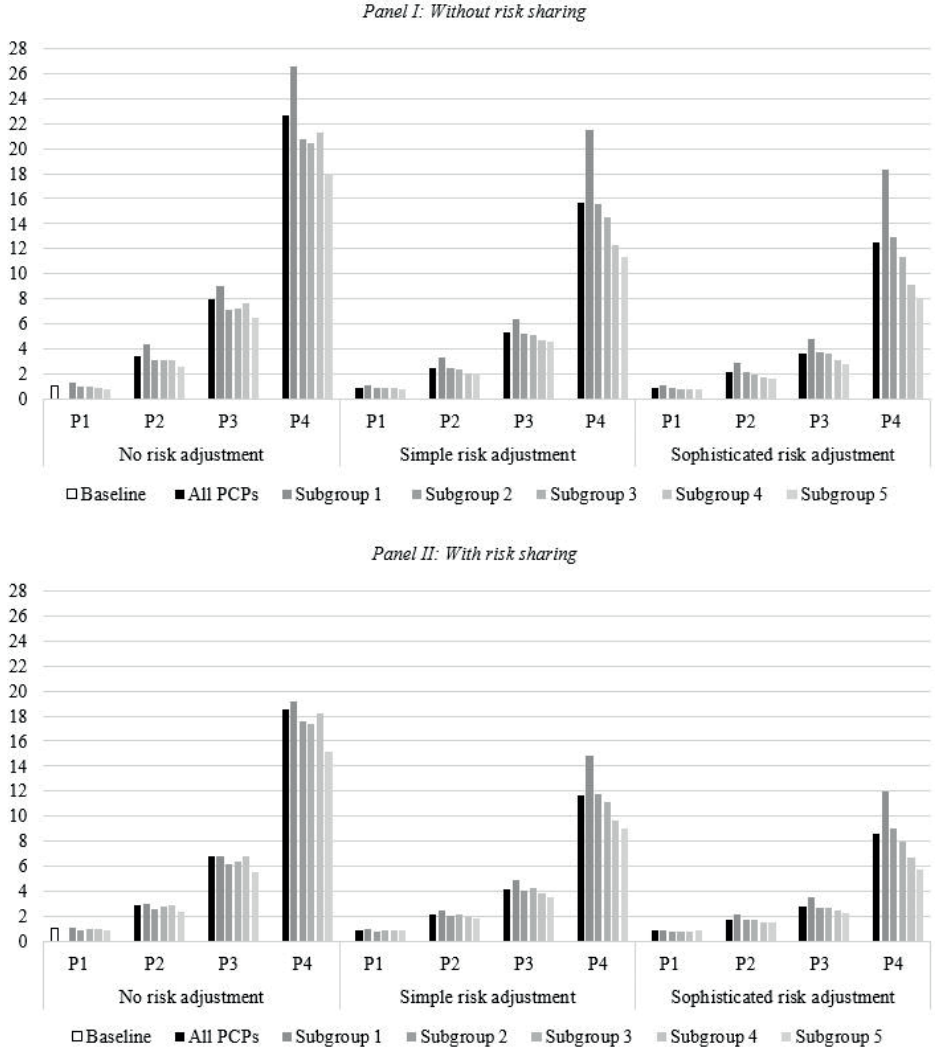
a. R-squared = proportion explained variation in spending = 1 minus the sum of squared residuals divided by the total sum of squares.

b. P1 is the sum of spending on primary care provided by physicians. P2 is P1 but supplemented with primary care diagnostics, physiotherapy, and durable medical equipment. P3 is P2 but supplemented with pharmaceutical care. P4 is P3 but supplemented with hospital care.

c. Without risk sharing, PCPs are financially accountable for 100% of spending under the relevant package. With risk sharing, the payer accounts for 100% of spending above a threshold set at the 99$^{th}$ percentile of the spending distribution under the relevant package.

d. Simple risk adjustment uses the risk adjusters age interacted with gender, socioeconomic status, and source of income to predict individual-level spending, which in turn is used to calculate the payment per PCP. Sophisticated risk adjustment includes the same risk adjusters as the simple model but supplemented with the morbidity-based risk adjusters (i.e. pharmacy-based cost groups, diagnosis-based cost groups, and multiple-year high-cost groups).

**Figure A.5.1.** Financial-risk ratios relative to baseline [a] for all PCPs (N = 5,581) and five subgroups of PCPs with increasing panel size [b], for four care packages [c] by type of risk adjustment [d] and with and without risk sharing [e] (2010-data)



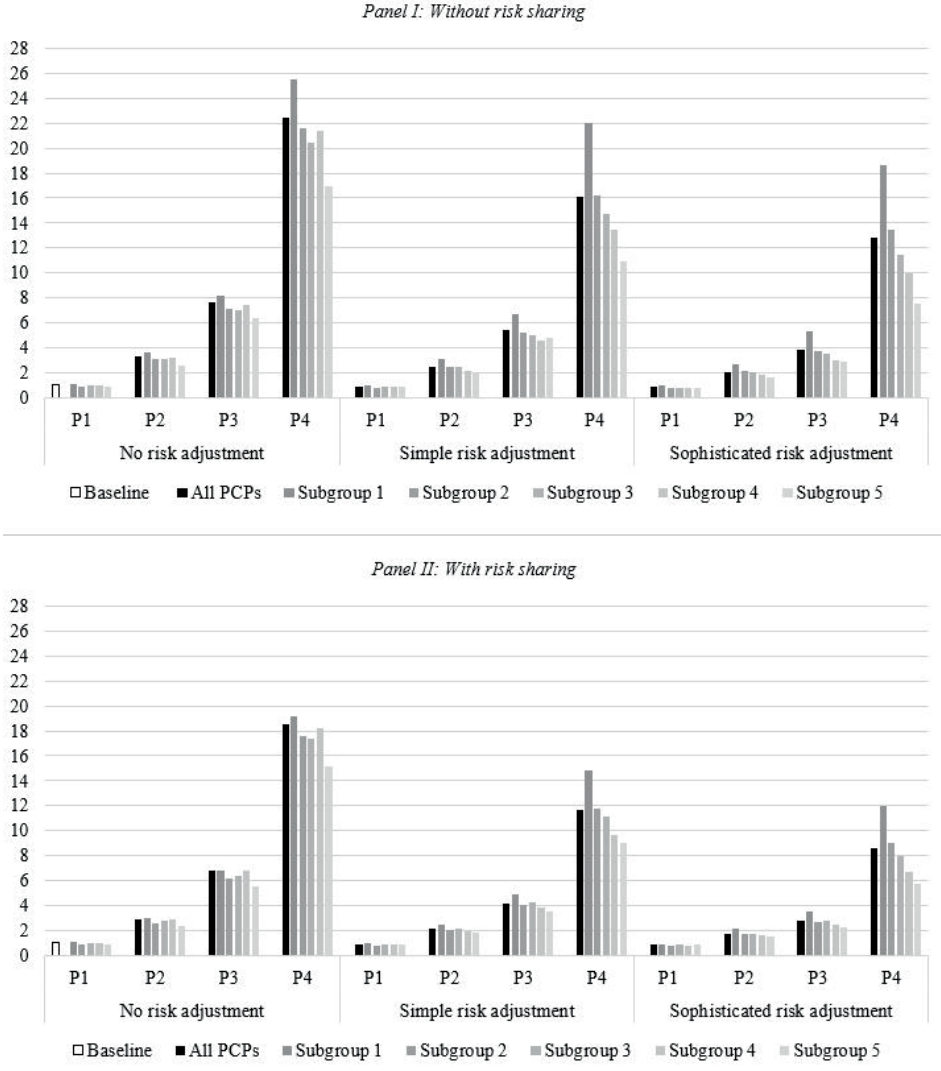

Note: PCP = primary care provider.

a. Baseline = P1 and no risk adjustment, for all PCPs. A financial-risk ratio of 0.5 denotes 50% lesser financial risk compared to the baseline while values of 1.0 and 1.5 mean similar or 50% greater risk, respectively. For panel I, the absolute value of financial risk (i.e. the standard deviation of mean PCP-level residual spending) for the baseline configuration is €17. Mean spending for all PCPs under P4 is a factor 14 greater than under P1 (i.e. €1,680 vs. €121). For panel II, the absolute value of financial risk for the baseline configuration is €15. Mean spending for all PCPs under P4 is a factor 12 greater than under P1 (i.e. €1,430 vs. €116).

b. The full sample of PCPs is divided into five equally sized subgroups with increasing panel size, ranging from on average 195 patients for subgroup 1 to on average 1,814 patients for subgroup 5.

c. P1 is the sum of spending on primary care provided by physicians. For panel I, average spending is €118 for all PCPs and ranges from €117 for subgroup 1 to €124 for subgroup 5. For panel II, average spending is €116 for all PCPs and ranges from €115 for subgroup 1 to €121 for subgroup 5.
   P2 is P1 but supplemented with primary care diagnostics, physiotherapy, and durable medical equipment. For panel I, average spending is €278 for all PCPs and ranges from €266 for subgroup 1 to €306 for subgroup 5. For panel II, average spending is €257 for all PCPs and ranges from €246 for subgroup 1 to €281 for subgroup 5.
   P3 is P2 but supplemented with pharmaceutical care. For panel I, average spending is €558 for all PCPs and ranges from €523 for subgroup 1 to €634 for subgroup 5. For panel II, average spending is €515 for all PCPs and ranges from €485 for subgroup 1 to €579 for subgroup 5.
   P4 is P3 but supplemented with hospital care. For panel I, average spending is €1,587 for all PCPs and ranges from €1,499 for subgroup 1 to €1,791 for subgroup 5. For panel II, average spending is €1,430 for all PCPs and ranges from €1,354 for subgroup 1 to €1,597 for subgroup 5.
d. Under no risk adjustment, the payment equals the grand mean spending in the data for the specific package. Simple risk adjustment uses the risk adjusters age interacted with gender, socioeconomic status, and source of income to predict individual-level spending, which in turn is used to calculate the payment per PCP. Sophisticated risk adjustment includes the same risk adjusters as the simple model but supplemented with the morbidity-based risk adjusters (i.e. pharmacy-based cost groups, diagnosis-based cost groups, and multiple-year high-cost groups).
e. Without risk sharing, PCPs are financially accountable for 100% of spending under the relevant package. With risk sharing, the payer accounts for 100% of spending above a threshold set at the 99$^{th}$ percentile of the spending distribution under the relevant package.

**Figure A.5.2.** Financial-risk ratios relative to baseline [a] for all PCPs (N = 5,556) and five subgroups of PCPs with increasing panel size [b], for four care packages [c] by type of risk adjustment [d] and with and without risk sharing [e] (2011-data)



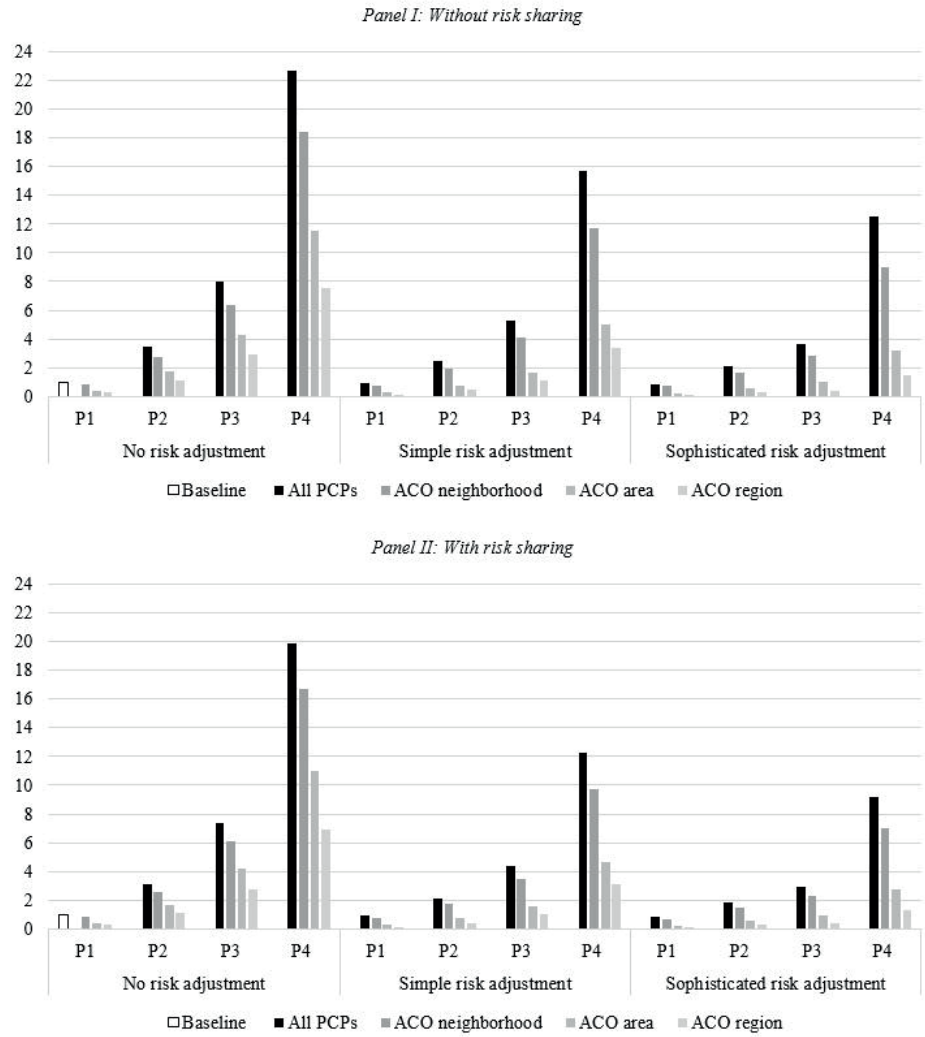

Note: PCP = primary care provider.

a. Baseline = P1 and no risk adjustment, for all PCPs. A financial-risk ratio of 0.5 denotes 50% lesser financial risk compared to the baseline while values of 1.0 and 1.5 mean similar or 50% greater risk, respectively. For panel I, the absolute value of financial risk (i.e. the standard deviation of mean PCP-level residual spending) for the baseline configuration is €18. Mean spending for all PCPs under P4 is a factor 13 greater than under P1 (i.e. €1,669 vs. €132). For panel II, the absolute value of financial risk for the baseline configuration is €17. Mean spending for all PCPs under P4 is a factor 11 greater than under P1 (i.e. €1,490 vs. €130).

b. The full sample of PCPs is divided into five equally sized subgroups with increasing panel size, ranging from on average 198 patients for subgroup 1 to on average 1,795 patients for subgroup 5.

c. P1 is the sum of spending on primary care provided by physicians. For panel I, average spending is €132 for all PCPs and ranges from €129 for subgroup 1 to €141 for subgroup 5. For panel II, average spending is €130 for all PCPs and ranges from €127 for subgroup 1 to €138 for subgroup 5.
   P2 is P1 but supplemented with primary care diagnostics, physiotherapy, and durable medical equipment. For panel I, average spending is €302 for all PCPs and ranges from €288 for subgroup 1 to €332 for subgroup 5. For panel II, average spending is €281 for all PCPs and ranges from €268 for subgroup 1 to €307 for subgroup 5.
   P3 is P2 but supplemented with pharmaceutical care. For panel I, average spending is €580 for all PCPs and ranges from €540 for subgroup 1 to €660 for subgroup 5. For panel II, average spending is €536 for all PCPs and ranges from €503 for subgroup 1 to €604 for subgroup 5.
   P4 is P3 but supplemented with hospital care. For panel I, average spending is €1,669 for all PCPs and ranges from €1,554 for subgroup 1 to €1,881 for subgroup 5. For panel II, average spending is €1,490 for all PCPs and ranges from €1,393 for subgroup 1 to €1,662 for subgroup 5.
d. Under no risk adjustment, the payment equals the grand mean spending in the data for the specific package. Simple risk adjustment uses the risk adjusters age interacted with gender, socioeconomic status, and source of income to predict individual-level spending, which in turn is used to calculate the payment per PCP. Sophisticated risk adjustment includes the same risk adjusters as the simple model but supplemented with the morbidity-based risk adjusters (i.e. pharmacy-based cost groups, diagnosis-based cost groups, and multiple-year high-cost groups).
e. Without risk sharing, PCPs are financially accountable for 100% of spending under the relevant package. With risk sharing, the payer accounts for 100% of spending above a threshold set at the 99$^{th}$ percentile of the spending distribution under the relevant package.

**Figure A.5.3.** Financial-risk ratios relative to baseline [a] for all PCPs (N = 5,581) and three types of simulated ACOs with increasing size [b], for four care packages [c] by type of risk adjustment [d] and with and without risk sharing [e] (2010-data)



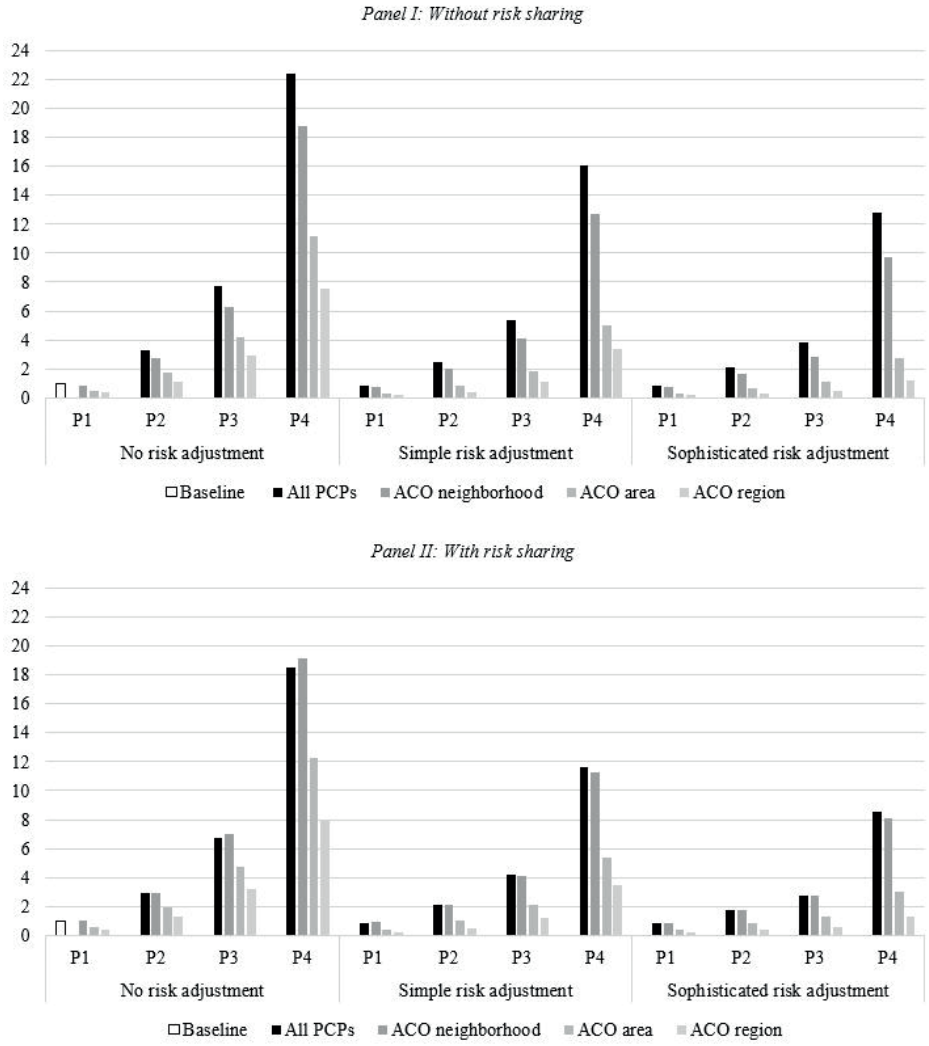

Note: PCP = primary care provider. ACO = accountable care organization.

a. Baseline = P1 and no risk adjustment, for all PCPs. A financial-risk ratio of 0.5 denotes 50% lesser financial risk compared to the baseline while values of 1.0 and 1.5 mean similar or 50% greater risk, respectively. For panel I, the absolute value of financial risk (i.e. the standard deviation of mean PCP-level residual spending) for the baseline configuration is €17. Mean spending for all PCPs under P4 is a factor 14 greater than under P1 (i.e. €1,680 vs. €121). For panel II, the absolute value of financial risk for the baseline configuration is €15. Mean spending for all PCPs under P4 is a factor 12 greater than under P1 (i.e. €1,430 vs. €116).

b. To simulate ACOs based on geographic proximity with larger panel sizes than real PCPs observed in the data, PCPs were clustered to the level of neighborhood (i.e. based on all four digits of the zip code of the address of PCPs practice locations, mean panel size = 2,002 patients), area (i.e. based on the first two digits, mean panel size = 46,337 patients), and region (i.e. based on the first digit, mean panel size = 463,373 patients).

c. P1 is the sum of spending on primary care provided by physicians. For panel I, average spending is €118 for all PCPs and ranges from €117 for the smallest ACOs to €119 for the largest ACOs. For panel II, average spending is €116 for all PCPs and ranges from €114 for the smallest ACOs to €117 for the largest ACOs.
   P2 is P1 but supplemented with primary care diagnostics, physiotherapy, and durable medical equipment. For panel I, average spending is €278 for all PCPs and ranges from €273 for the smallest ACOs to €284 for the largest ACOs. For panel II, average spending is €257 for all PCPs and ranges from €252 for the smallest ACOs to €262 for the largest ACOs.
   P3 is P2 but supplemented with pharmaceutical care. For panel I, average spending is €558 for all PCPs and ranges from €542 for the smallest ACOs to €570 for the largest ACOs. For panel II, average spending is €515 for all PCPs and ranges from €501 for the smallest ACOs to €526 for the largest ACOs.
   P4 is P3 but supplemented with hospital care. For panel I, average spending is €1,587 for all PCPs and ranges from €1,545 for the smallest ACOs to €1,619 for the largest ACOs. For panel II, average spending is €1,430 for all PCPs and ranges from €1,395 for the smallest ACOs to €1,455 for the largest ACOs.
d. Under no risk adjustment, the payment equals the grand mean spending in the data for the specific package. Simple risk adjustment uses the risk adjusters age interacted with gender, socioeconomic status, and source of income to predict individual-level spending, which in turn is used to calculate the payment per PCP. Sophisticated risk adjustment includes the same risk adjusters as the simple model but supplemented with the morbidity-based risk adjusters (i.e. pharmacy-based cost groups, diagnosis-based cost groups, and multiple-year high-cost groups).
e. Without risk sharing, PCPs are financially accountable for 100% of spending under the relevant package. With risk sharing, the payer accounts for 100% of spending above a threshold set at the 99$^{th}$ percentile of the spending distribution under the relevant package.

**Figure A.5.4.** Financial-risk ratios relative to baseline [a] for all PCPs (N = 5,556) and three types of simulated ACOs with increasing size [b], for four care packages [c] by type of risk adjustment [d] and with and without risk sharing [e] (2011-data)




Note: PCP = primary care provider. ACO = accountable care organization.

a. Baseline = P1 and no risk adjustment, for all PCPs. A financial-risk ratio of 0.5 denotes 50% lesser financial risk compared to the baseline while values of 1.0 and 1.5 mean similar or 50% greater risk, respectively. For panel I, the absolute value of financial risk (i.e. the standard deviation of mean PCP-level residual spending) for the baseline configuration is €18. Mean spending for all PCPs under P4 is a factor 13 greater than under P1 (i.e. €1,669 vs. €132). For panel II, the absolute value of financial risk for the baseline configuration is €17. Mean spending for all PCPs under P4 is a factor 11 greater than under P1 (i.e. €1,490 vs. €130).

b. To simulate ACOs based on geographic proximity with larger panel sizes than real PCPs observed in the data, PCPs were clustered to the level of neighborhood (i.e. based on all four digits of the zip code of the address of PCPs practice locations, mean panel size = 1,992 patients), area (i.e. based on the first two digits, mean panel size = 45,961 patients), and region (i.e. based on the first digit, mean panel size = 459,611 patients).

c. P1 is the sum of spending on primary care provided by physicians. For panel I, average spending is €132 for all PCPs and ranges from €130 for the smallest ACOs to €134 for the largest ACOs. For panel II, average spending is €130 for all PCPs and ranges from €128 for the smallest ACOs to €132 for the largest ACOs.
   P2 is P1 but supplemented with primary care diagnostics, physiotherapy, and durable medical equipment. For panel I, average spending is €302 for all PCPs and ranges from €296 for the smallest ACOs to €309 for the largest ACOs. For panel II, average spending is €281 for all PCPs and ranges from €275 for the smallest ACOs to €287 for the largest ACOs.
   P3 is P2 but supplemented with pharmaceutical care. For panel I, average spending is €580 for all PCPs and ranges from €563 for the smallest ACOs to €592 for the largest ACOs. For panel II, average spending is €536 for all PCPs and ranges from €522 for the smallest ACOs to €548 for the largest ACOs.
   P4 is P3 but supplemented with hospital care. For panel I, average spending is €1,669 for all PCPs and ranges from €1,629 for the smallest ACOs to €1,696 for the largest ACOs. For panel II, average spending is €1,490 for all PCPs and ranges from €1,457 for the smallest ACOs to €1,512 for the largest ACOs.
d. Under no risk adjustment, the payment equals the grand mean spending in the data for the specific package. Simple risk adjustment uses the risk adjusters age interacted with gender, socioeconomic status, and source of income to predict individual-level spending, which in turn is used to calculate the payment per PCP. Sophisticated risk adjustment includes the same risk adjusters as the simple model but supplemented with the morbidity-based risk adjusters (i.e. pharmacy-based cost groups, diagnosis-based cost groups, and multiple-year high-cost groups).
e. Without risk sharing, PCPs are financially accountable for 100% of spending under the relevant package. With risk sharing, the payer accounts for 100% of spending above a threshold set at the 99$^{th}$ percentile of the spending distribution under the relevant package.

**Table A.5.5.** Risk of ruin (in %) for all PCPs (N = 5,581) and two types of simulated ACOs, for four care packages [a] by type of risk adjustment [b] and with and without risk sharing [c] (2010-data)

| | No risk adjustment | | | | Simple risk adjustment | | | | Sophisticated risk adjustment | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **P1** | **P2** | **P3** | **P4** | **P1** | **P2** | **P3** | **P4** | **P1** | **P2** | **P3** | **P4** |
| **Without risk sharing** | | | | | | | | | | | | |
| All PCPs | 0.3 | 5.4 | 17.5 | 30.4 | 0.2 | 3.3 | 12.0 | 29.4 | 0.2 | 2.1 | 8.2 | 28.8 |
| ACO neighborhood [c] | 0.2 | 2.2 | 12.8 | 24.0 | 0.1 | 1.3 | 7.3 | 22.1 | 0.1 | 1.5 | 5.5 | 22.8 |
| ACO area [d] | 0.0 | 0.0 | 2.2 | 20.0 | 0.0 | 0.0 | 0.0 | 5.6 | 0.0 | 0.0 | 0.0 | 2.2 |
| **With risk sharing** | | | | | | | | | | | | |
| All PCPs | 0.2 | 4.0 | 16.3 | 29.4 | 0.2 | 1.6 | 10.2 | 27.7 | 0.2 | 0.9 | 5.3 | 25.9 |
| ACO neighborhood [c] | 0.1 | 1.2 | 11.3 | 23.5 | 0.0 | 0.4 | 4.7 | 20.8 | 0.1 | 0.5 | 2.9 | 19.5 |
| ACO area [d] | 0.0 | 0.0 | 3.3 | 20.0 | 0.0 | 0.0 | 0.0 | 5.6 | 0.0 | 0.0 | 0.0 | 1.1 |

Note: PCP = primary care provider. ACO = accountable care organization.

a. P1 is the sum of spending on primary care provided by physicians. P2 is P1 but supplemented with primary care diagnostics, physiotherapy, and durable medical equipment. P3 is P2 but supplemented with pharmaceutical care. P4 is P3 but supplemented with hospital care.

b. Under no risk adjustment, the payment equals the grand mean spending in the data for the specific package. Simple risk adjustment uses the risk adjusters age interacted with gender, socioeconomic status, and source of income to predict individual-level spending, which in turn is used to calculate the payment per PCP. Sophisticated risk adjustment includes the same risk adjusters as the simple model but supplemented with the morbidity-based risk adjusters (i.e. pharmacy-based cost groups, diagnosis-based cost groups, and multiple-year high-cost groups).

c. Without risk sharing, PCPs are financially accountable for 100% of spending under the relevant package. With risk sharing, the payer accounts for 100% of spending above a threshold set at the 99$^{th}$ percentile of the spending distribution under the relevant package.

d. The mean panel size for the 2,083 ACOs is 2,002 patients.

e. The mean panel size for the 90 ACOs is 46,337 patients.

**Table A.5.6.** Risk of ruin (in %) for all PCPs (N = 5,556) and two types of simulated ACOs, for four care packages [a] by type of risk adjustment [b] and with and without risk sharing [c] (2011-data)

| | No risk adjustment | | | | Simple risk adjustment | | | | Sophisticated risk adjustment | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **P1** | **P2** | **P3** | **P4** | **P1** | **P2** | **P3** | **P4** | **P1** | **P2** | **P3** | **P4** |
| **Without risk sharing** | | | | | | | | | | | | |
| All PCPs | 0.1 | 5.1 | 17.5 | 31.1 | 0.1 | 3.2 | 11.8 | 29.2 | 0.1 | 2.5 | 7.7 | 27.6 |
| ACO neighborhood [c] | 0.1 | 2.5 | 11.9 | 25.4 | 0.2 | 1.2 | 6.1 | 23.3 | 0.2 | 1.3 | 4.7 | 22.1 |
| ACO area [d] | 0.0 | 0.0 | 1.1 | 22.2 | 0.0 | 0.0 | 0.0 | 4.4 | 0.0 | 0.0 | 0.0 | 1.1 |
| **With risk sharing** | | | | | | | | | | | | |
| All PCPs | 0.1 | 4.3 | 16.3 | 29.4 | 0.1 | 2.1 | 10.0 | 27.1 | 0.1 | 1.2 | 5.2 | 25.8 |
| ACO neighborhood [c] | 0.1 | 1.7 | 12.0 | 23.9 | 0.1 | 0.5 | 5.2 | 20.3 | 0.2 | 0.4 | 2.9 | 19.7 |
| ACO area [d] | 0.0 | 0.0 | 1.1 | 20.0 | 0.0 | 0.0 | 0.0 | 4.4 | 0.0 | 0.0 | 0.0 | 1.1 |

Note: PCP = primary care provider. ACO = accountable care organization.

a. P1 is the sum of spending on primary care provided by physicians. P2 is P1 but supplemented with primary care diagnostics, physiotherapy, and durable medical equipment. P3 is P2 but supplemented with pharmaceutical care. P4 is P3 but supplemented with hospital care.

b. Under no risk adjustment, the payment equals the grand mean spending in the data for the specific package. Simple risk adjustment uses the risk adjusters age interacted with gender, socioeconomic status, and source of income to predict individual-level spending, which in turn is used to calculate the payment per PCP. Sophisticated risk adjustment includes the same risk adjusters as the simple model but supplemented with the morbidity-based risk adjusters (i.e. pharmacy-based cost groups, diagnosis-based cost groups, and multiple-year high-cost groups).

c. Without risk sharing, PCPs are financially accountable for 100% of spending under the relevant package. With risk sharing, the payer accounts for 100% of spending above a threshold set at the 99$^{th}$ percentile of the spending distribution under the relevant package.

d. The mean panel size for the 2,077 ACOs is 1,992 patients. The mean panel size for the 90 ACOs is 45,961 patients.

## Appendix B. General information about models S2 and S3

**Table B.5.1.** Number of individuals, number of PCP entities, and medical spending in the estimation samples S2 and S3 constructed for sensitivity analyses (2012-data)

| | **S2** [a] | **S3** [a] |
|---|---|---|
| N(unweighted) [b] | 2,904,620 | 4,136,064 |
| N(weighted) [b] | 2,873,065 | 4,092,293 |
| N(PCP) | 2,687 | 5,327 |
| **Without risk sharing** [c] | | |
| Mean spending P1 (SD) [d] | 122 (109) | 121 (109) |
| Mean spending P2 (SD) [d] | 300 (735) | 295 (718) |
| Mean spending P3 (SD) [d] | 596 (1,740) | 580 (1,679) |
| Mean spending P4 (SD) [d] | 1,740 (5,921) | 1,699 (5,802) |
| **With risk sharing** [c] | | |
| Mean spending P1 (SD) [d] | 120 (82) | 119 (81) |
| Mean spending P2 (SD) [d] | 276 (440) | 272 (431) |
| Mean spending P3 (SD) [d] | 543 904) | 529 (876) |
| Mean spending P4 (SD) [d] | 1,529 (3,228) | 1,494 (315) |

Note: PCP = primary care provider. SD = standard deviation.

a. S2 = Patients of PCPs with panel sizes between 500 and 2,500 patients. S3 = Patients of PCPs with panel size fluctuations of maximally (-)50% between 2011 and 2012.

b. Unweighted = number of individuals in the sample. Weighted = number of individuals weighted by the duration of the insurance contract in 2012 (i.e. the number of insured-years).

c. Without risk sharing, PCPs are financially accountable for 100% of spending under the relevant package. With risk sharing, the payer accounts for 100% of spending above a threshold set at the 99$^{th}$ percentile of the spending distribution under the relevant package.

d. Annualized and weighted by the duration of the insurance contract in 2012. P1 is the sum of spending on primary care provided by physicians. P2 is P1 but supplemented with primary care diagnostics, physiotherapy, and durable medical equipment. P3 is P2 but supplemented with pharmaceutical care. P4 is P3 but supplemented with hospital care.

# Chapter 6

## Getting the incentives right: Simulating the effects of residual-based risk-sharing for primary care providers under global payment

With Frank Eijkenaar and Richard van Kleef

Preparing for submission

## ABSTRACT

In provider payment system design, an important challenge is to maximize incentives for cost control while minimizing the risk of unwanted effects. This paper examines an innovative form of risk sharing that might prove helpful in this regard: residual-based risk sharing. Despite its potential, this form of risk sharing has not been studied in the context of provider payment. The goal of this paper is to provide insight into the incentive effects and the tradeoffs associated with the design of residual-based risk sharing in the presence of morbidity-based risk adjustment. Using rich administrative data, we simulate the effects of various modalities of residual-based risk sharing for primary care providers under global payment on cost-control incentives, risk selection incentives, upcoding incentives, and excessive provider-level losses/profits. We show that reducing undesirable incentives through residual-based risk sharing inherently involves a sacrifice of incentives for cost control. Though small levels of risk sharing can achieve much in terms of less risk on unwanted effects, an acceptable reduction of that risk still requires a sizeable sacrifice of cost-control incentives. We conclude that residual-based risk sharing is a promising design feature of global provider payment models and that it is up to decision makers to weigh the pros and cons associated with various design choices, given context-specific preferences.

## 1. INTRODUCTION

Over the past decade, healthcare regulators, purchasers and policymakers have started to explore alternative provider payment models to help steering healthcare systems towards value (Friedberg et al. 2020; Scott et al. 2018; Struijs et al. 2019; Chernew et al. 2020). A growing number of these models rely on global payment arrangements in which collaborating healthcare providers jointly accept clinical and financial responsibility for the provision of a comprehensive care package to a predefined patient population (Vlaanderen et al. 2019; Cattel & Eijkenaar 2020b). Examples from practice include the Medicare Shared Savings Program, the Medicare Next Generation ACO Model, and the Alternative Quality Contract in the United States (US), the Gesundes Kinzigtal project in Germany, and the Menzis Shared Savings Program in the Netherlands (Song et al. 2019; Ginsburg & Patel 2017; Cattel & Eijkenaar 2020b).

Because of their prospective nature and the care package stretching beyond single services, diseases or treatments, global payment confronts providers with greater amounts of financial risk for medical spending than under conventional fee-for-service and episode-based payment models. This means that providers become to some extent accountable for discrepancies between spending and payments, which strengthens their incentives to control costs (Jegers et al. 2002; Miller 2009). Given the huge variability in individual-level medical spending, however, a potential danger of global payment is that providers might be tempted to engage in risk selection, i.e., actions to attract (deter) individuals that are predictably (un)profitable. Risk selection may threaten efficiency and equity (Newhouse 1989; Barros 2003). Another potential drawback is that providers – particularly those with small patient panels – might be confronted with excessive losses if their panels happen to contain some extremely underpaid patients. This might result in low participation rates in the payment program, unwanted bankruptcies, and/or strategic provider behavior (Werbeck et al. 2020; Shen 2003; Hofer et al. 1999; Dranove et al. 2003; McDonald and Roland 2009; Chen et al. 2011; Chang et al. 2012; Hsieh et al. 2016).

Both in theory and practice, risk adjustment and risk sharing are important measures to reap the benefits of financial risk for providers under global payment while mitigating adverse effects related to risk selection and excessive losses (Ash & Ellis 2012; Cattel et al. 2020a; Cattel & Eijkenaar 2020b). With risk adjustment, provider payments are based on the predicted spending of the patient population given a predefined set of characteristics (e.g., age, gender and morbidity).

With risk sharing, provider payments are (partly) based on observed spending. While risk adjustment and risk sharing mitigate incentives for selection and the risk of excessive losses, they also come with a price. Specifically, risk adjustment might confront providers with incentives for upcoding. Insofar risk adjustment is based on morbidity indicators derived from prior healthcare utilization and hospital diagnoses (which in practice typically is the case), providers receive higher payments for individuals flagged by these indicators. To qualify for higher payments, providers might manipulate the codes that are used to determine the payment or strategically provide

services that indirectly generate codes for more serious diagnoses or higher-than-necessary medication dosages. Risk sharing comes with a different problem: because it creates a direct link between observed spending and payment, providers' incentives for cost control are reduced.

Designing risk adjustment and risk sharing for the purpose of global provider payment thus involves a tradeoff between incentives for cost control, incentives for risk selection, incentives for upcoding, and the risk of excessive losses. In this respect, we believe the field of provider payment might benefit from insights from the field of health plan payment (i.e., payments to insurers for health plan enrollees). Recently, Schillo et al. (2016) and McGuire et al. (2020a, 2020b) have proposed an innovative combination of morbidity-based ex-ante risk adjustment and ex-post risk sharing for health insurance plans. The essence of Schillo et al.'s proposal is to complement risk adjustment with *individual-level payments based on residual spending*, with residual spending being defined as observed spending less risk-adjusted payment. Under this approach, health plans would receive extra payments for those individuals most heavily *underpaid* by the risk-adjustment model. However, focusing on residual spending also calls attention to individuals who are heavily *overpaid* by the risk-adjustment model. Therefore, McGuire et al. (2020a, 2020b) recommend using residual-based *repayments* in addition to residual-based payments. In addition, they recommend accounting for the presence of risk sharing in optimizing the payment weights in the risk-adjustment model (section 3).

For four reasons, we believe that these innovations from the health plan payment literature can be very helpful in the design of global provider payment in light of the abovementioned tradeoff. First, basing risk sharing on *residual* spending net of risk adjustment and optimizing payment weights for the presence of risk sharing avoids 'double payment' for high-cost patients via the morbidity-based ex-ante risk-adjustment model and the ex-post risk-sharing model. This minimizes the reduction in incentives for cost control, given a fixed risk-sharing budget. Second, residual-based risk sharing results in a more targeted reduction of risk selection incentives by reducing predictable over- and underpayments net of risk adjustment. Third, repayments limit the overpayment of individuals flagged by morbidity-based risk adjustors for whom spending is lower than expected. Therefore, requiring repayments reduces incentives for upcoding relative to using payments only. Finally, targeting risk sharing to the highest under- and overpayments net of risk adjustment mitigates excessive losses and profits for providers.

Despite their potential advantages, the effects of residual-based (re)payments have not been studied in the context of provider payment. Therefore, the goal of this paper is to provide empirical insight in the effect of this innovative form of risk sharing in terms of (1) incentives for cost control, (2) incentives for risk selection, (3) incentives for upcoding, and (4) the risk of excessive provider-level financial results. Specifically, using rich administrative data on medical spending and risk characteristics of over 4.4 million individuals enrolled with a large Dutch health insurer, we simulate risk-adjusted global payments for primary care providers (PCPs) for a comprehensive care package, and apply various risk-sharing modalities that differ in the funds devoted to risk sharing and in whether only residual-based payments or both payments and repayments are used.

We expect the resulting insights to be of substantial value for healthcare providers, purchasers, and policymakers in designing better provider payment models.

This paper is structured as follows. Using theoretical and empirical literature on provider and health plan payment, the next section discusses in more detail the potential effects of residual-based risk sharing on incentives for cost control, risk selection, and upcoding, as well as on providers' risk of excessive financial results. In section 3 the data and the empirical approach are described. The results are presented in section 4 and in section 5 the main findings and their implications are discussed.

## 2. THEORETICAL BACKGROUND

### 2.1 Risk adjustment and risk sharing

Providers under global payment bear financial risk for the medical spending of individuals enrolled with or attributed to them with respect to a predefined package of care. For this to be successful, providers should only be accountable for the share of spending that they can influence (i.e., for the 'performance risk' related to their clinical skills and the efficiency of 'care production'). However, individual-level spending variation also has other sources related to the random nature of the occurrence of health problems (insurance risk) and differences in health risk (systematic risk). Given their role and sphere of influence, providers should not be held responsible for these risks. In practice, however, it is virtually impossible to split performance risk from insurance and systematic risk, implying that without additional measures, global payment confronts providers with at least some risk that is ideally borne by the payer, which may lead to undesirable incentives and effects (Vermaas 2006; Hussey et al. 2011; Cattel & Eijkenaar 2020b). In this paper we investigate how the specific design of risk sharing can help in striking a balance between desirable and undesirable incentives and effects (Ash & Ellis 2012; Rose et al. 2016; Cattel et al. 2020a). Specifically, we focus on the design of risk sharing in the presence of morbidity-based risk adjustment.

With risk adjustment, providers receive higher (lower) payments for individuals with higher (lower) predicted spending. To predict spending, demographic (age, gender), socioeconomic (e.g., income and educational level) and morbidity-based information (e.g., prior utilization and diagnoses) can be used (McGuire & van Kleef 2018). Designing a risk-adjustment model that adequately corrects for systematic spending variation due to differences in population risk characteristics has proven very challenging in practice, with even highly sophisticated models still leaving substantial gaps between observed and predicted spending. For example, in a recent study on the design of health plan payment McGuire et al. (2020b) found that in the Netherlands, Germany and the US Marketplaces, for one in a thousand people observed spending exceeds predicted spending by €70,636, €87,494, and €189,918, respectively. On the other side of the distribution, for one in a thousand people observed spending falls below predicted spending by

more than €24,539, €27,842, and €95,335. Importantly, residual spending from risk adjustment is to some extent predictable, resulting in incentives for risk selection (McGuire et al. 2020b; Van Veen 2015b).

Risk sharing can protect providers against predictable spending variation (i.e., systematic risk) not accounted for by risk adjustment (Van de Ven & Ellis, 2000). In addition, risk sharing can help in protecting providers against the random risk of drawing a few high-cost cases (i.e., insurance risk), which can be financially hazardous or even ruinous for providers (especially for those with small patient panels) because a very small fraction of the total population accounts for a disproportionate share of (the variance of) total spending (Geruso & McGuire 2016). The specific design of risk sharing has consequences for the incentives for providers and can vary along several dimensions, including the group for whom the risk is shared, the types of care for which the risk is shared, and how much risk is shared (Van Barneveld 2000).

In practice, many forms of risk sharing are possible. In this paper the focus is on what McGuire et al. (2020a; 2020b) refer to as 'residual-based reinsurance and repayments' or 'residual-based risk sharing'. This form of risk sharing is similar to the commonly applied 'excess-of-loss risk sharing' in the sense that it involves ex-post payments for individual-level spending above a threshold. There are, however, three important differences between excess-of-loss risk sharing and residual-based risk-sharing as studied in this paper. First, with residual-based risk sharing, risk sharing is targeted at individual-level *residual* spending (i.e., spending net of risk adjustment) instead of just spending. The rationale of taking residual spending as a basis for risk sharing is that incentives for risk selection do not stem from high spending per se but from over- and underpayments by the risk-adjustment model (Schillo et al. 2016; McGuire et al. 2020a; McGuire et al. 2020b). Second, in contrast to excess-of-loss risk sharing, residual-based risk sharing not only involves payments to providers for highly underpaid individuals (i.e., those with large positive residuals) but can also require providers to contribute *repayments* for highly overpaid individuals (i.e., those with large negative residuals). With such a two-sided risk-sharing model, payments are more fairly distributed among providers since payments for individuals overpaid by risk adjustment are redistributed to individuals underpaid by risk adjustment. Consequently, the risk of excessive financial results for providers reduces. Third, residual-based risk sharing as proposed by McGuire et al. (2020a; 2020b) involves optimization of the payment weights in the risk-adjustment model for the presence of risk sharing. The rationale is that if a provider is not responsible for certain costs as a result of risk sharing, these costs should not be part of the dependent variable in the risk-adjustment model (McGuire et al. 2020a; 2020b). In practice, however, such optimization is typically not implemented in risk-sharing arrangements.

## 2.2 (Measuring) the effects of residual-based risk sharing in the presence of risk adjustment

### *2.2.1 Incentives for cost control*

Under a global payment, providers have strong incentives to work efficiently. Because the marginal financial benefits of diagnosis and treatment are zero, providers are stimulated to minimize cost by reducing waste (Jegers et al. 2002; Miller 2009; McGuire & van Kleef 2018). Evaluations of two prominent examples of global payment programs for providers in the US – the Medicare Shared Savings Program and the Alternative Quality Contract – have demonstrated significantly lower spending growth rates and roughly equal levels of quality as compared to control groups (Song et al. 2019; McWilliams et al. 2018).

Providers have incentives to control costs to the extent they are held financially responsible for these costs. Therefore, incentives for cost control can roughly be measured by the extent to which the provider is accountable for healthcare spending or, in other words, the share of cost at the margin borne by the provider (Van Kleef & van Vliet 2021; McGuire & van Kleef 2018).[19] Consequently, any form of risk sharing is expected to reduce incentives for cost control by introducing a direct link between (residual) spending and payment (McGuire et al. 2020b). In the context of health plan payment, McGuire et al. (2020a; 2020b) show that residual-based risk sharing with optimized payment weights can strongly reduce undesirable incentives while touching only a small share of overall spending and share of the total population, thus keeping the reduction in cost-control incentives to a minimum.

### *2.2.2 Incentives for risk selection*

Holding providers to some extent accountable for discrepancies between spending and payments might introduce incentives for risk selection. Risk selection may have several adverse effects in terms of efficiency and equity, and in the context of provider payment has been shown to not just be a theoretical concern (Werbeck et al. 2020; Shen 2003; Hofer et al. 1999; Dranove et al. 2003; McDonald and Roland 2009; Chen et al. 2011; Chang et al. 2012; Hsieh et al. 2016; McWilliams et al. 2020). Risk adjustment narrows the gaps between observed spending and predicted spending (i.e. the payment) and thereby reduces incentives for risk selection. Residual-based risk sharing reduces these incentives further by moving payments from heavily overpaid individuals to heavily underpaid individuals (Schillo et al. 2016).

Existing literature offers a wide range of measures of risk selection incentives, all based on gaps between observed and predicted spending (see McGuire & van Kleef 2018 and Van Veen et al. 2015b for an overview). In selecting the relevant measure, an important question is at which level(s) providers can engage in risk selection. Providers are likely to have detailed individual-level information about the health status and spending history of their patients. Therefore, they

19 In the literature, more sophisticated measures to quantify incentives for cost control are lacking. Therefore, in this paper we use this rough measure.

may well be able to effectively identify relatively low-risk (high-risk) individuals whose spending is expected to be below (above) average and behave accordingly, for example by skimping on the quality of care required by underpaid patients and/or by referring these patients to other providers. Because providers are likely to be able to select on the level of individual patients, individual-level measures are relevant.

Commonly used individual-level measures of selection incentives are R-squared, Payment System Fit (PSF), and Cumming's Prediction Measure (CPM). The R-squared from a regression of individual-level spending on a set of risk adjustors can be interpreted as the proportion of the total variance in spending explained by the risk-adjustment model. PSF is a generalization of the R-squared and takes account of other features influencing the gaps between payment and spending, such as risk sharing. PSF can be understood as the percentage of total variance in spending explained by the full set of payment system features. CPM is a linear version of R-squared and PSF and uses the absolute instead of squared values of payment gaps. Compared to CPM, R-squared and PSF give more weight to large gaps and therefore are more likely to be affected by individuals with high spending.

An important disadvantage of individual-level measures like R-squared, PSF, and CPM is that they are based on gaps between payment and observed spending and observed spending serves as a benchmark in calculating incentives for risk selection. Risk selection incentives, however, are in fact driven by *predictable* over- and underpayments by the risk-adjustment model. Therefore, additional measures are often used which provide insight into the *systematic gaps* between observed and predicted spending for specific patient groups. Examples of such measures are mean over- and undercompensations or predictive ratios at the level of specific groups, like those with a certain chronic condition or with persistently high prior spending.

### *2.2.3 Incentives for upcoding*

Risk adjustment might introduce incentives for providers to overstate the measured risk of their patient panel and trigger classification of individuals in risk classes with high payment weights, for example by directly manipulating the (diagnosis) codes used for classification or providing more or other services than necessary. For simplicity, we refer to all these types of behaviors as 'upcoding' (Simborg 1981; Geruso & Layton 2020). Upcoding is undesirable because it reduces efficiency, for example due to the provision of unnecessary services (Steinbusch et al. 2007; Georgescu & Hartmann 2013) and has been shown to occur in practice (Geruso & Layton 2020; Geruso & Layton 2017; McGuire et al. 2021).

Incentives for upcoding depend on the possible financial gains of the action: the less there is to gain, the weaker the incentive to 'game the system' (Politzer 2020). The effect of risk adjustment on these incentives depends on the type of risk adjustors included in the model. In contrast to adjustors based on demographic and socioeconomic information, morbidity-based adjustors derived from prior utilization and diagnoses are prone to upcoding because providers generate and report the information that will determine which individuals will receive a morbidity flag

which in turn affects the payment (McGuire & van Kleef 2018). Risk sharing in general limits incentives for upcoding because it reduces providers' perceived (financial) need to game the system. Residual-based repayments are expected to reduce these incentives even more by imposing limits on the potential gains from forcing individuals in risk classes with high payment weights.

In quantifying incentives for upcoding, an important factor is the population for which upcoding is studied. Focusing on specific groups that are likely to be subject to upcoding practices seems reasonable, because incentives for upcoding for these groups are actionable (Politzer 2020; Behrend et al. 2007). For example, it is more realistic for a provider to extend an existing medication prescription for chronically ill patients by a few days to trigger classification in a risk class than prescribing an extensive amount of unnecessary medication to perfectly healthy individuals to realize the same outcome.

Little is known about how to operationalize incentives for upcoding. Lamers et al. (1999) quantify these incentives as the ratio between payments for individuals assigned to a pharmacy-based cost group (i.e., an indicator for chronic conditions based on prior use of prescription medication) and their pharmacy costs. Behrend et al. (2007) and Politzer (2020) calculate incentives for upcoding as the incremental payment resulting from the upcoding activity (i.e., the estimated payment weight for a specific risk class) minus the minimum costs required for classification in this class.

#### *2.2.4 Risk of excessive losses/profits*

A potential danger of global payment is that providers might be exposed to too much (insurance) risk, which might even turn out to be ruinous. Therefore, additional risk-mitigating measures are warranted. The risk of excessive losses can be measured by the percentage of providers with a total loss that exceeds their total payment by at least some percentage, for example 5% (Layton & McGuire 2016).

Due to the law of the large numbers, providers with larger patient panels are better equipped to spread financial risk and absorb random spending shocks than providers with smaller patient panels, ceteris paribus (Frakt & Mayes 2012; Spector et al. 2018). For example, in a recent simulation study Cattel & Eijkenaar (2020c) found that under global payments with morbidity-based risk adjustment but without risk sharing, more than 20% of PCPs would suffer an excessive loss in case of panel sizes of maximally 2,000 patients, while this would be approximately 1% is case of 50,000 patients.

Risk adjustment and risk sharing both reduce excessive losses, though to what extent depends on the specific design and multiple other factors, including the specific care package to which the payment pertains. For example, Cattel and Eijkenaar (2020c) find that the risk of suffering an excessive loss would be close to 0% for a package covering only spending on primary care provided by physicians, while in case of a package that also includes spending on hospital care more than a quarter of all PCPs would suffer an excessive loss, regardless of whether morbidity-based risk adjustment and high-cost risk sharing is applied. This underlines the importance of

careful consideration of the design of risk-mitigating measures when applying global payments covering broad care packages. The specific features of the type of risk sharing studied in this paper, i.e., focusing on outliers in *both tails* of the *residual* spending distribution, have important advantages in this respect.

## 3. DATA AND METHODS

### 3.1 Data

For our empirical simulations we used two large administrative datasets. The first dataset includes individual-level data on medical spending and risk characteristics. This dataset was originally composed and used for calculating Dutch insurers' risk-adjusted capitation payments for the year 2015. It contains 2012-spending covered by the basic benefit package of the Health Insurance Act, for the following categories: primary care, physiotherapy, durable medical equipment, prescription medication, and hospital care. In addition, this dataset includes the risk adjustors of the 2015 risk-adjustment model (see van Kleef et al. 2018 for details): age interacted with gender (40 classes), socio-economic status using household income (4 classes), source of income (6 classes), pharmacy-based cost groups (25 PCGs based on prior use of specific types of prescription medication) and diagnosis-based cost groups (15 DCGs based on specific diagnoses from prior hospital treatments).

The second dataset, obtained from a large Dutch health insurer, contains a unique PCP identifier and the 4-digit zip code of the address of the PCP's practice. In our data, a PCP can be an individual physician, a group practice, or a health center (i.e., an entity in which multiple PCPs and other primary care providers provide and coordinate care). The data include all individuals enrolled with the insurer that provided the data and represent real PCP patient panels from 2012. Importantly, people who did not visit their PCP are still identifiable in the data because Dutch PCPs receive a fixed registration fee for each individual enrolled in their practice, regardless of actual healthcare utilization. The total sample contains approximately 4.4 million individuals served by roughly 7,000 PCPs.

We merged the two datasets at the individual patient level using a unique identification key that was anonymized by a trusted third party.

### 3.2 Methods

To prevent overfitting, we split the merged dataset into an equal-sized training and test sample. The training sample is used to calibrate our payment models. The test sample is used to simulate the effects of payment models on a set of outcome measures. Using these samples, our empirical approach consisted of three steps that are discussed below.

### *3.2.1 Step 1: Constructing the care package and calculating the risk-adjusted payment*

In line with global payment models implemented in practice, our payment system simulations assumed a comprehensive care package and morbidity-based risk adjustment (Cattel & Eijkenaar 2020b). We constructed the care package by summing individual-level, annualized spending on primary care, physiotherapy, durable medical equipment, prescription medication, and hospital care. To calculate the risk-adjusted payment (i.e., predicted annual total spending) for each individual, we ran a weighted ordinary least squares regression on the training sample (using individuals' duration of the insurance contract in 2012 as weights) with age interacted with gender, socioeconomic status, source of income, PCGs, and DCGs as risk adjustors. This provided us with the payment weights required to calculate individual-level risk-adjusted payments (i.e., combined with the individual-level scores on the risk-adjustor variables). We assumed a one-year payment contract duration, full insurance coverage for patients, and no other provider payment arrangements being in place.

### *3.2.2 Step 2: Simulating residual-based risk sharing*

For all individuals in our data we calculated residual spending by subtracting predicted spending (i.e., the risk-adjusted payment) from observed spending. Next, we simulated a series of risk-sharing scenarios in which the upper and lower thresholds where residual-based (re)payment 'kicks in' was determined such that the total (re)payment pools take the size of X% of total payments. Specifically, we simulated the following scenarios of X% payments/repayments with full risk sharing above the thresholds: 0/0, 1/0, 2/0, 5/0, 10/0, 20/0, 1/1, 2/2, 5/5, 10/10, and 20/20. Scenario 0/0 represents our baseline scenario of no risk sharing. Scenario 1/0, for example, represents a scenario with 1% payments and no repayments (i.e., 1% of total payments devoted to risk sharing), while scenario 1/1 represents a situation with 1% payments and 1% repayments (i.e., 2% of total payments devoted to risk sharing). Following McGuire et al. (2020a) we calibrated the risk-adjustor weights in the risk-adjustment model *post* residual-based risk sharing. Specifically, in an iterative procedure the regression weights were optimized accounting for the presence of risk sharing. For more details about this procedure, we refer to McGuire et al. (2020a).

### *3.2.3 Step 3: Calculating outcome measures*

Using the test sample, we next assessed the impact of our eleven risk-sharing scenarios on incentives for cost control, risk selection, and upcoding, as well as on the risk of excessive losses/profits. Below we describe how we measured these four outcomes.

*Incentives for cost control*

To quantify cost-control incentives we calculated the degree to which the provider is, on average, accountable for healthcare spending, as follows:

$$\text{Incentives for cost control} = 1 - \frac{\sum_i | R_i |}{\sum_i (P_i)} \quad (1)$$

with $|R_i|$ representing the absolute value of total residual-based (re)payments for individual *i* and *Pi* being the total payment for individual *i*.

*Incentives for risk selection*

We report three indicators for incentives for risk selection. The first is individual-level Payment System Fit (PSF), calculated as:

$$PSF = 1 - \frac{\sum_i (Y_i - P_i)^2}{\sum_i (Y_i - \overline{Y})^2} \quad (2)$$

with $Y_i$ being observed spending for individual *i*, $P_i$ representing the total payment for that individual, and $\overline{Y}$ corresponding to the mean observed spending in the sample. PSF measures the extent to which the total payment a provider receives for an individual from the risk-adjustment model plus (minus) the (re)payment from the risk-sharing model, tracks observed spending for that individual. Without risk sharing, payment $P_i$ corresponds to the risk-adjusted payment and PSF equals the R-squared. Our second measure is Cumming's Prediction Measure (CPM) at the individual level, which is a linear version of PSF:

$$CPM = 1 - \frac{\sum_i |Yi - Pi|}{\sum_i |Y_i - \overline{Y}|} \quad (3)$$

Third, we calculated the mean financial result for a specific risk group, as follows:

$$\text{Financial result} = \frac{\sum_{i \in g} (P_i - Y_i)}{n_g} \quad (4)$$

where $P_i$ equals the payment for individual *i*, $Y_i$ is observed spending for that individual, $i \in g$ relates to the individuals in the group of concern *g*, and $n_g$ is the number of individuals in that group. Note that a negative financial result would imply an undercompensation and a positive financial result an overcompensation. As the group of concern, we selected those individuals belonging to the top-10% spenders in each of the three prior years (i.e., 2009-2011). We chose a look-back period of three years based on the qualitative argument that with 1 or 2 years, high spending can be a coincidence due to incidental health problems, while high persistent spending is likely to be a more valid indicator of costly health problems in the current year. This group is a potential target for risk selection since consistently high prior spenders are likely to be (heavily) underpaid in the current year (McGuire et al. 2020b; Van Veen 2015b). Because of the long-term relationship that PCPs often have with their patients, PCPs are arguably in the position to identify (individuals belonging to) this group (Versteeg & Batenburg 2017).

*Incentives for upcoding*

We quantified upcoding incentives as the average incremental payment per person to the provider as a result of classification of a patient in a certain risk group included in the risk-adjustment model, net of risk sharing:

$$\text{Incentives for upcoding} = \frac{\sum_{i \in g} (\Delta P_i)}{n_g} \tag{5}$$

Where $\Delta P_i$ equals the incremental payment to the provider resulting from classification of patient $i$ in a specific risk group, $i \in g$ relates to the individuals in this risk group $g$, and $n_g$ to the number of patients in this group. We selected PCGs and DCGs as relevant risk groups for our analysis because we expect these adjustors to be most prone to upcoding by PCPs since they are based on prior utilization and diagnoses. We calculated the average incremental payment for PCGs and DCGs separately because PCPs arguably have more influence on prescription medication (on which PCGs are based) than on diagnoses from hospital treatments (on which DCGs are based). For each PCG and DCG, the incremental payment is limited to the repayment threshold. If the incremental payment for a specific PCG is, for example, €6,000 and the repayment threshold is €5,000, the incremental payment for the provider is capped at €5,000. Incentives for upcoding are presented relative to our baseline scenario of no risk sharing using index numbers.[20]

*Risk of excessive losses/profits*

The risk of incurring an excessive financial results (i.e., loss or profit) is measured as the percentage of PCPs with a loss or profit that exceeds their payment by at least 5% (Layton & McGuire 2016), as follows[21]:

$$\text{Risk of excessive loss of profit} = \frac{\sum_p (\text{loss\_profit}_p \geq 0.05 * P_p)}{m} \tag{6}$$

with $loss\_profit_p$ calculated as the difference between the mean per person payment for PCP $p$ ($P_p$) and the mean per person spending for that PCP, $P_p$ calculated as the risk-adjusted payment plus the residual-based payment minus the residual-based repayment, and $m$ denoting the total number of PCPs in the sample. A PCP incurs a loss when spending exceeds the payment, while a profit is made when the payment exceeds spending.

20 We acknowledge this measure provides just a rough indication of PCPs' incentives for upcoding. A more comprehensive measure would also incorporate the PCPs' additional costs of upcoding practices. However, our data do not allow for such an extension.

21 We acknowledge that 5% of the payment is an arbitrary threshold; in the context of primary care providers, it is unclear when a profit or loss can be considered 'excessive'. Therefore, in the results section we also present the distribution of PCP-level residual spending for various levels of residual-based payments and repayments.

To prevent this measure from being affected by differences in PCPs' patient panel size, we restricted our analyses on this measure to a specific group of simulated PCPs. In doing so, we assumed that a global payment contract as simulated in this paper would be available and appealing to PCPs with large patient panels. To simulate these provider entities, we clustered PCPs in our data based on geographic proximity using the first three digits of the zip code of PCPs' practice locations and excluded PCPs with panels smaller than 1,000 patients. This resulted in 472 simulated PCPs with a mean panel size of 4,276 (see Appendix). The reason for clustering PCPs based on the first three digits of the zip code is that this resulted in a reasonable number of larger simulated PCP entities while having to drop only 7% of the patients in our total (test) sample.

## 4. RESULTS

### 4.1 Descriptive statistics

Table 6.1 shows some descriptive statistics of the full, training, and test sample. Results are similar across the three samples. Among the 4,4 million individuals in the full sample and the 2,2 million individuals in the training and test samples, 19% is categorized in a PCG and 9% in a DCG. In all three samples, mean spending is just above €1,740 per person per year with a standard deviation of approximately €6,300. Given that the median (about €353) is much below the mean, spending is concentrated at the right tail of the distribution.

**Table 6.1.** Descriptive statistics of the full, training, and test sample at the individual enrollee-level (2012-data)

| | Full sample | Training sample | Test sample |
|---|---|---|---|
| N | 4,385,936 | 2,192,968 | 2,192,968 |
| N (weighted) [a] | 4,338,441 | 2,169,219 | 2,169,221 |
| Male, 0 – 17 years | 11% | 11% | 11% |
| Male, 18 – 65 years | 31% | 31% | 31% |
| Male, 65+ years | 8% | 8% | 8% |
| Female, 0 – 17 years | 10% | 10% | 10% |
| Female, 18 – 65 years | 31% | 31% | 31% |
| Female, 65+ years | 10% | 10% | 10% |
| In a PCG | 19% | 19% | 19% |
| In a DCG | 9% | 9% | 9% |
| **Spending (€)** [b] | | | |
| Mean | 1,744 | 1,746 | 1,742 |
| Standard deviation | 6,285 | 6,254 | 6,316 |
| 1st percentile | 25 | 25 | 25 |
| 10th percentile | 76 | 76 | 76 |
| Median | 353 | 354 | 353 |
| 90th percentile | 3,759 | 3,766 | 3,752 |
| 99th percentile | 21,685 | 21,767 | 21,606 |
| Maximum | 2,103,391 | 2,103,391 | 1,700,246 |

Note: PCG = pharmacy-based cost group; DCG = diagnosis-based cost group.

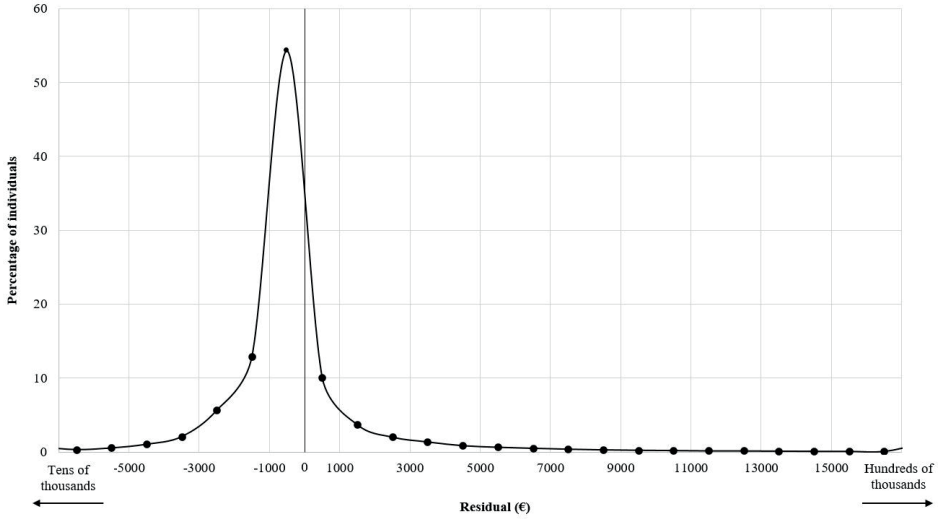
a. Number of individuals weighted by duration of insurance contract in 2012 (i.e., the number of insured years).

b. Annualized and weighted by the duration of the insurance contract in 2012.

## 4.2 Payment system simulations

Figure 6.1 shows the distribution of individual-level residual spending after risk adjustment for the baseline scenario of no risk sharing for the test sample (results for the training sample are virtually identical). Residual spending is positive when spending is higher than predicted (i.e., underpayment) and negative when spending is lower than predicted (i.e., overpayment). Residuals range from about -€91,000 to approximately €1.7 million and have a standard deviation of approximately €5,600. Residuals are negative until about the 75th percentile of the distribution implying that in the absence of risk sharing, PCPs would be overcompensated for most individuals under this global payment model. However, a considerable share of the population would still come with a (considerable) loss; for 10% of the individuals residual spending exceeds €1,400.

**Figure 6.1.** Distribution of individual-level residual spending after risk adjustment for the baseline scenario (2012-data, test-sample, annualized and weighted)




Note: Individuals on the horizontal axis are clustered based on residual spending using bins of €1,000.

Table 6.2 reports basic information on our scenarios of residual-based risk sharing. The first combination is our baseline scenario of no risk sharing. Combinations 2-6 represent scenarios with residual-based payments only and combinations 7-11 add repayments to these payments. The third column displays incentives for cost control according to equation (1) and shows to what extent PCPs are accountable for healthcare spending. Logically, higher risk-sharing percentages imply weaker cost-control incentives. The fourth and fifth column show the upper and lower thresholds where risk sharing kicks in. For scenario 1/1 (i.e., 1% payments and 1% repayments) the upper threshold is €123,827 and the lower threshold is -€14,215. Under this scenario, providers would receive additional payment for individuals who are underpaid by more than €123,827, while they would be required to make repayments for individuals overpaid by more than €14,215. Naturally, (re)payment thresholds are closer to €0 with more extensive risk sharing. In addition, for a given percentage dedicated to residual-based payments, the upper thresholds go down when repayments are added to the payment model. In line with Figure 6.1, the absolute value of the thresholds for payments is much higher than those for repayments.

The last two columns of Table 6.2 show that under most of our scenarios, only a small fraction of the total population is affected by risk sharing. For example, setting aside up to 20% of funds for residual-based payments (and 0% for repayments) would affect less than 3% of the population. In general, the fraction of the population affected by payments becomes slightly higher when repayments are added to the model. In line with the distribution of residual spending in Figure 6.1, the share of the population touched by residual-based payments is much lower than that touched by repayments (maximally 3% versus 16%). In our most extensive risk-sharing scenario (i.e., 20/20), in total approximately 19% of the population is affected (i.e., 3.02% + 16.07%).

**Table 6.2.** Incentives for cost control, (re)payment thresholds, and affected population for various scenarios of residual-based risk sharing (2012-data, test sample)

| Residual-based payment as % of total payment | Residual-based repayment as % of total payment | Incentives for cost control (%) [a] | Upper threshold for payment (€) [b] | Lower threshold for repayment (€) [b] | % of population affected by payment | % of population affected by repayment |
|---|---|---|---|---|---|---|
| 0 | 0 | 100 | - | - | 0.00 | 0.00 |
| 1 | 0 | 99 | 128,650 | - | 0.02 | 0.00 |
| 2 | 0 | 98 | 75,700 | - | 0.06 | 0.00 |
| 5 | 0 | 95 | 36,311 | - | 0.27 | 0.00 |
| 10 | 0 | 90 | 19,188 | - | 0.83 | 0.00 |
| 20 | 0 | 80 | 8,257 | - | 2.61 | 0.00 |
| 1 | 1 | 98 | 123,827 | -14,215 | 0.02 | 0.10 |
| 2 | 2 | 96 | 72,763 | -8,788 | 0.06 | 0.41 |
| 5 | 5 | 90 | 34,204 | -4,740 | 0.29 | 1.86 |
| 10 | 10 | 80 | 17,230 | -2,737 | 0.88 | 5.63 |
| 20 | 20 | 60 | 6,415 | -1,230 | 3.02 | 16.07 |

a. Calculated as the degree to which the provider is, on average, accountable for spending (equation (1)).
b. Thresholds are based on the training sample.

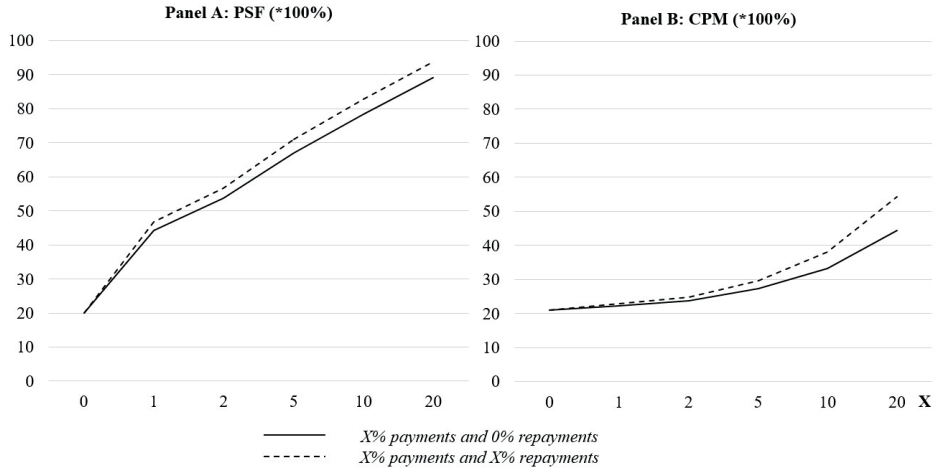
### *4.2.1 Incentives for risk selection*

Figure 6.2 shows the impact of our risk-sharing scenarios on two measures of incentives for risk selection: PSF (panel A) and CPM (panel B). The solid lines represent results for the scenarios with residual-based payments only, while the dashed lines show results for the scenarios with both payments and repayments.

Panel A of Figure 6.2 shows substantial improvements in PSF as a result of residual-based payments (solid line). Devoting only 1% of total payment to residual-based payments already more than doubles PSF compared to the baseline scenario (from 20% to 44%). Increasing residual-based payments to 20% leads to a PSF of 89%. Adding repayments to the payment model improves PSF further (to maximally 94%), especially under the more extensive scenarios.

Panel B of Figure 6.2 shows that risk sharing also improves CPM. However, compared to PSF the improvement is less pronounced and occurs especially under the more extensive scenarios. In addition, under the more extensive scenarios the positive effect of repayments is stronger (maximally 10 percentage points increase in CPM versus maximally 5 percentage points increase in PSF).

**Figure 6.2.** Payment System Fit (PSF; panel A) and Cumming's Prediction Measure (CPM; panel B) for various scenarios of residual-based risk sharing

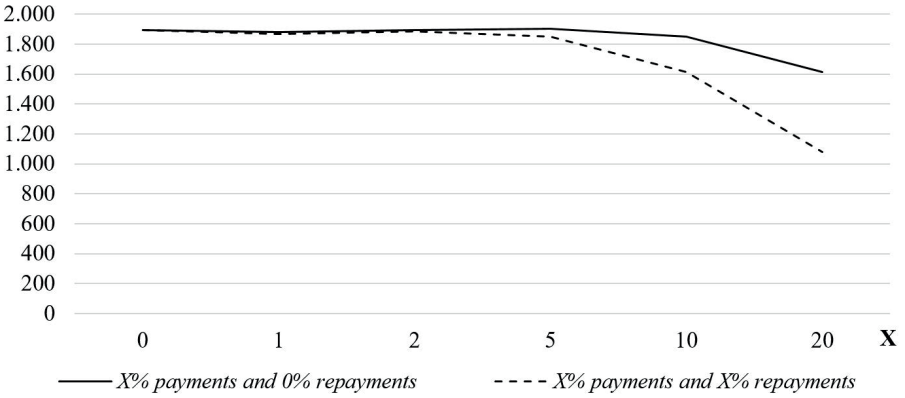



Note. PSF measures the extent to which the total payment a provider receives from the risk-adjustment model plus (minus) the (re)payment from the risk-sharing model, tracks observed spending (equation (2)). CPM is a linear version of PSF (equation (3)).

Figure 6.3 shows the results for our third measure of selection incentives: the mean undercompensation for the group individuals belonging to the top-10-% of spenders in each of the three prior years. This group encompasses 1.2% of individuals in the sample and the mean undercompensation under the baseline scenario is €1,895. Initially, the impact of residual-based risk sharing on this measure is limited, but becomes substantial under our more extensive scenarios, especially if repayments are added. For example, while the undercompensation decreases with approximately €280 when going from 0% to 20% residual-based payments (but no repayments), adding 20% repayments to this scenario leads to an extra reduction in the mean undercompensation of over €530.[22]

22 At first sight, it might be surprising that repayments result in a larger reduction of the mean undercompensation than payments. The explanation can be found in a specific feature of our residual-based risk sharing method, i.e., the optimization of risk-adjustment payment weights for the presence of (re)payments. With residual-based payments, optimization of risk adjustment results in a reduction of payment weights for morbidity indicators (due to the correlation between morbidity flags and being a high residual spender), which tempers the effect of payments on the undercompensation in Figure 6.3. With residual-based repayments, optimization of risk adjustment results in an increase of payment weights for morbidity indicators (due to the correlation between morbidity flags and being a low residual spender). Since – mechanically – the correlation between morbidity flags and being a low residual spender is higher than that between morbidity flags and being a high residual spender the increase in payment weights due to repayments is larger than the decrease in payment weights due to payments. Apparently, the group presented in Figure 6.3 benefits from these dynamics (on top of the direct effects from payment and repayments on the undercompensation for this group).

**Figure 6.3.** Mean undercompensation (€) for the top-10% of spenders in each of the three prior years for various scenarios of residual-based risk sharing



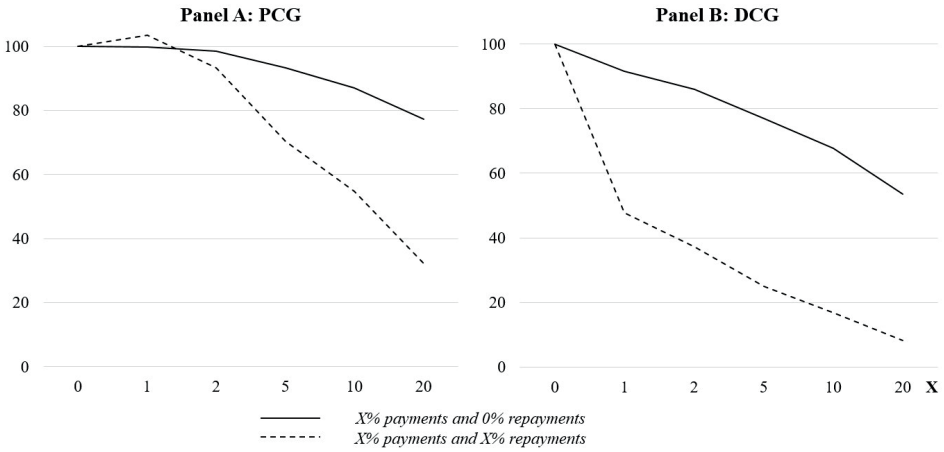

Note. The mean undercompensation is calculated using equation (4).

### *4.2.2 Incentives for upcoding*

Figure 6.4 shows the results on our measure of upcoding incentives relative to the baseline scenario of no risk sharing (index=100). The measure is based on the mean incremental payment for the two morbidity indicators in the risk-adjustment model: PCGs (panel A) and DCGs (panel B). The gradually downward sloping solid lines indicate that the introduction of payments alone already leads to (substantial) reductions in incentives for upcoding, although impact for PCGs is only visible for higher shares of payments.

As expected and as indicated by the sharp decline of the dashed lines in Figure 6.4, the impact of repayments on incentives for upcoding is much stronger than that of residual-based payments. The impact is particularly substantial for DCGs; setting aside 1% for residual-based payments leads to a decrease in the mean incremental payment for DCGs of approximately 8 percentage points in comparison to the baseline scenario, while adding 1% repayments results in a further reduction of 44 percentage points. Regarding the PCGs, setting aside 1% for payments does not lead to a notable reduction in the incentive for upcoding, while adding 1% repayments results in a small increase in the mean incremental payment (+€119; +4 percentage points). Under the most extreme risk-sharing scenario analyzed here (i.e., 20% residual-based payments in combination with 20% repayments), incentives for upcoding are reduced by 68 percentage points for PCGs and even by 92 percentage points for DCGs relative to the baseline scenario.

**Figure 6.4.** Mean incentives for upcoding (relative to the baseline scenario of no risk sharing; index=100) for having an individual flagged by a PCG (panel A) or DCG (panel B) for various scenarios of residual-based risk sharing

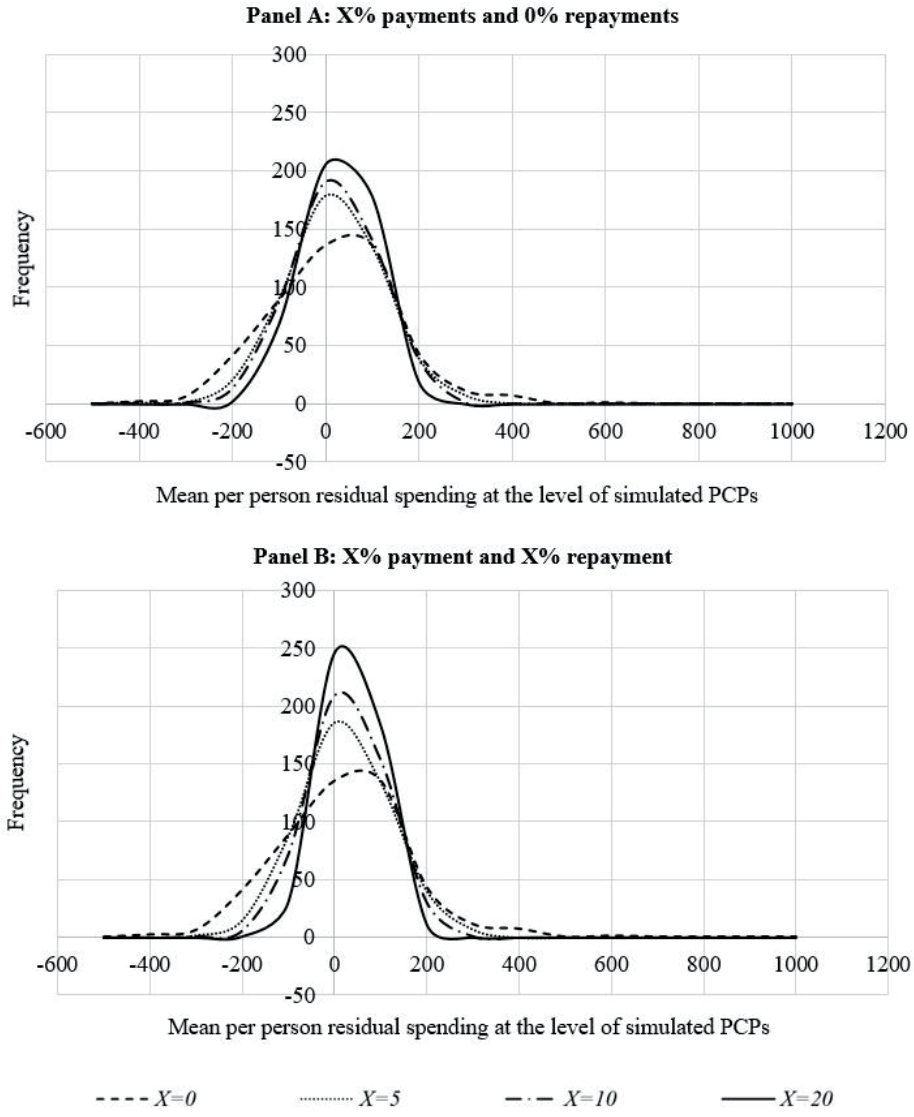



Note. The incentives for upcoding are calculated as the mean per person incremental payment to the provider as a result of classification in a pharmacy-based cost group (PCG) or diagnosis-based cost group (DCG) using equation (5). Incentives for upcoding under the various risk-sharing scenarios are presented relative to those for the baseline scenario of no risk sharing (index = 100), with a mean incremental payment of €3,344 for PCGs and of €14,747 for DCGs.

### *4.2.3 Excessive loss/profit*

Figure 6.5 shows the distribution of mean per person residual spending at the level of simulated PCPs (section 3.2.3) with residual spending defined as the difference between the mean per person spending and the payment, and payment calculated as the risk-adjusted payment plus the residual-based payment minus the residual-based repayment. Panel A of Figure 6.5 shows the impact of a selection of four scenarios with residual-based payments only, while the panel B shows the impact of these scenarios but then with repayments added. The figures show that regardless of the scenario, residual spending is positive for the majority of providers, implying that most providers are underpaid. In addition, both devoting a larger portion of total payments to risk sharing and adding repayments reduces the probability of (excessive) losses and profits for providers. This is underscored by a larger standard deviation of provider-level residual spending for the baseline scenario (i.e., €132) as compared to the scenario of 20% payments (i.e., €72) and the scenario with 20% repayments added (i.e., €58).

**Figure 6.5.** Distribution of mean per person residual spending at the level of simulated PCPs for various levels of residual-based payments (panel A) and repayments (panel B)



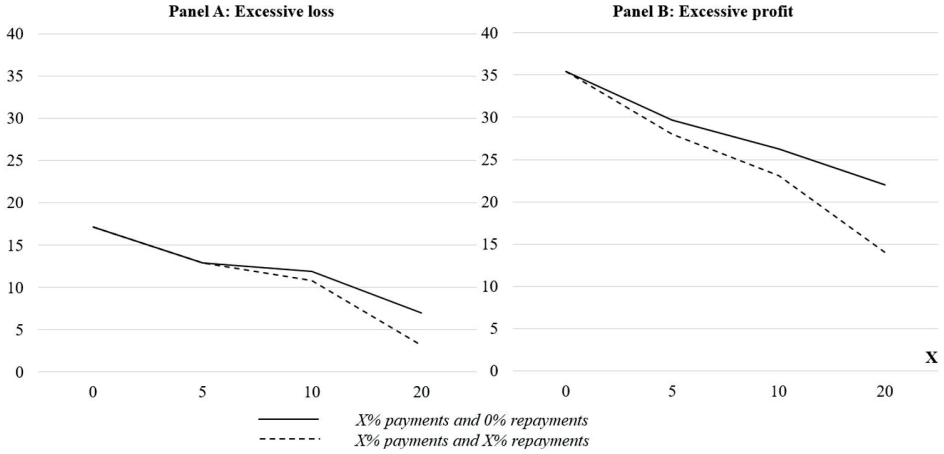

Note: In both panels, overall mean residual spending is (slightly) positive instead of €0 because provider-level results are not weighted by panel size.

Figure 6.6 shows the percentages of simulated PCPs that would suffer a loss (panel A) or make a profit (panel B) that exceeds their payment by at least 5% under our risk-sharing scenarios. In line with Figure 6.5, risk sharing reduces the risk of excessive losses/profits, although relatively high (re)payment percentages are required for a substantial impact. For the baseline scenario,

approximately 17% of all PCPs would suffer an excessive loss. This reduces to 7% when 20% of total payments is devoted to residual-based payments. Adding repayments lowers this percentage further, to 3% under our most extensive scenario.

The share of PCPs with an excessive profit is consistently higher than the share with an excessive loss, which is related to the shape of the provider-level residual spending distribution under the baseline scenario of no risk sharing (Figure 6.5). Without risk sharing, 35% of all simulated PCPs would make such a profit. This percentage reduces to 22% after introducing 20% residual-based payments and then to 14% when 20% repayments are added.

**Figure 6.6.** Percentage of simulated PCPs with an excessive loss (panel A) or profit (panel B)




Note: The risk of an excessive profit/loss is calculated using equation (6), with 'excessive' being defined as a loss or profit that exceeds the payment by at least 5%.

### *4.2.4 Tradeoffs*

Table 6.3 summarizes relevant results from the previous sections to gain insight in the inherent tradeoff between incentives for selection, incentives for upcoding, and risk of excessive loss/profit on the one hand, and incentives for cost control on the other hand across the various scenarios of residual-based risk-sharing.

Our analysis shows that devoting just a very small share of total payment to residual-based payments already substantially improves PSF, while the impact of repayments on PSF is relatively limited. The marginal returns of increasing the degree of risk sharing on PSF are diminishing: sacrificing additional incentives for cost control results in smaller and smaller increases in PSF. To considerably improve CPM, a relatively large sacrifice of incentives for cost control is required; up to 5% (re)payments, the effect of risk sharing is limited. In contrast to PSF, the marginal returns of increasing the degree of risk sharing on CPM are increasing: sacrificing additional incentives for cost control results in larger and larger increases in CPM. In addition, compared to PSF, the impact of repayments on CPM is stronger. Regarding our third measure of risk

selection – the mean undercompensation for the top-10% spenders in the three prior years –, the impact of residual-based payments is relatively limited. Sacrificing 20% of cost-control incentives by introducing residual-based payments reduces the baseline undercompensation (€1,895) by 15% to €1,612. Devoting minimally 5% of the total payment to residual-based repayments, on the contrary, already materially reduces the mean undercompensation.

The introduction of residual-based payments leads to steady reductions in incentives for upcoding, although the impact of repayments is much stronger. For DCGs, only small repayment percentages are needed to materially affect upcoding incentives, in contrast to PCGs. In general, the relative effects of risk sharing compared to baseline are more prominent for DCGs than for PCGs, implying that for DCGs a smaller sacrifice in terms of incentives for cost control is required to lower the incentives for upcoding considerably. Compared to baseline, devoting 20% of the total payment to residual-based payments lowers the mean incremental payment with 23 percentage points for PCGs and 46 percentage points for DCGs. Adding 20% repayments to this scenario (i.e., a total sacrifice of incentives for cost control 40%), further lowers the upcoding incentives considerably, leaving 32% and 8% of the initial incentives for PCGs and DCGs, respectively.

The last two columns of Table 6.3 show that the effect of relatively low percentages of risk sharing on the risk of excessive losses/profits is limited. Considerable incentives for cost control must be sacrificed to materially reduce this risk and even when 10% of the total payment is devoted to residual-based payments and 10% to residual-based repayments (i.e., 20% of incentives for cost control are sacrificed), the share of PCPs expected to suffer an excessive loss and the share of PCPs expected to make an excessive profit remains substantial (11% and 23%).

**Table 6.3.** Incentives for cost control, incentives for risk selection, incentives for upcoding, and excessive loss/profit for various scenarios of residual-based risk sharing

| Residual-based payment (%) | Residual-based repayment (%) | Incentives for cost control (%)[a] | Incentives for risk selection: PSF (%)[b] | Incentives for risk selection: CPM (%)[c] | Incentives for risk selection: undercompensation (€)[d] | Incentives for upcoding PCG (%)[e] | Incentives for upcoding DCG (%)[e] | Risk of excessive loss (%)[f] | Risk of excessive profit (%)[f] |
|---|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 100 | 20 | 21 | 1.895 | 100 | 100 | 17 | 35 |
| 1 | 0 | 99 | 44 | 22 | 1.882 | 100 | 92 | 17 | 33 |
| 2 | 0 | 98 | 54 | 24 | 1.895 | 98 | 86 | 15 | 32 |
| 5 | 0 | 95 | 67 | 27 | 1.904 | 93 | 77 | 13 | 30 |
| 10 | 0 | 90 | 78 | 33 | 1.851 | 87 | 68 | 12 | 26 |
| 20 | 0 | 80 | 89 | 44 | 1.612 | 77 | 54 | 7 | 22 |
| 1 | 1 | 98 | 47 | 23 | 1.866 | 104 | 48 | 16 | 32 |
| 2 | 2 | 96 | 57 | 25 | 1.883 | 93 | 37 | 15 | 32 |
| 5 | 5 | 90 | 71 | 30 | 1.848 | 70 | 25 | 13 | 28 |
| 10 | 10 | 80 | 83 | 38 | 1.613 | 55 | 17 | 11 | 23 |
| 20 | 20 | 60 | 94 | 54 | 1.080 | 32 | 8 | 3 | 14 |

a. Calculated as the degree to which the provider is, on average, accountable for spending (equation (1)).
b. PSF measures the extent to which the total payment a provider receives from the risk-adjustment model plus (minus) the (re)payment from the risk-sharing model, tracks observed spending (equation (2)).
c. CPM is a linear version of PSF (equation (3)).
d. The mean undercompensation is calculated using equation (4).
e. The incentives for upcoding are calculated as the mean per person incremental payment to the provider as a result of classification in a pharmacy-based cost group (PCG) or diagnosis-based cost group (DCG) using equation (5). Incentives for upcoding under the various risk-sharing scenarios are presented relative to those for the baseline scenario of no risk sharing (index = 100), with a mean incremental payment of €3,344 for PCGs and of €14,747 for DCGs.
f. The risk of an excessive profit/loss is calculated using equation (6), with 'excessive' being defined as a loss or profit that exceeds the payment by at least 5%.

# 5. DISCUSSION

## 5.1 Summary and discussion of main finding

To increase incentives for cost control, provider payment reforms increasingly rely on global payment models. These models need to be accompanied with the additional measures of risk adjustment and risk sharing, as otherwise they may well have unintended effects related to providers being confronted with too much financial risk and/or the wrong type of financial risk. However, these two risk-mitigating measures come with potential limitations themselves: risk adjustment might stimulate providers to engage in upcoding and risk sharing lowers cost-control incentives. The design of risk adjustment and risk sharing in the context of global payments for healthcare providers thus involves important tradeoffs.

The goal of this paper was to provide empirical insight in these tradeoffs and facilitate an informed design of global payment models. Specifically, the focus was on comparison of incentive effects of various scenarios of residual-based risk sharing under morbidity-based risk adjustment. This innovative form of risk sharing focuses on individual-level residual (rather than observed) spending net of risk adjustment, optimizes the payment weights of the risk-adjustment model to the presence of risk sharing, and requires providers to make repayments for overpaid patients in addition to receiving payments for underpaid patients. This approach is expected to (1) minimize the reduction in incentives for cost control by avoiding double payments from both risk adjustment and risk sharing, (2) result in a more targeted reduction of selection incentives by lowering over- and underpayments net of risk adjustment, (3) mitigate upcoding incentives by limiting individual-level overpayments, and (4) reduce the risk of excessive losses/profits for providers by targeting risk sharing to the highest under- and overpayments net of risk adjustment.

Our results show that, depending on the specific design, application of residual-based risk sharing can accomplish much in terms of reducing the risk on unwanted effects. However, this positive impact requires an inherent sacrifice of incentives for cost control, which depending on the outcome measure can be substantial. Only a relatively small sacrifice of cost-control incentives is needed to considerably reduce incentives for risk selection measured by PSF, while larger sacrifices are required to materially improve upon our other two measures for risk selection (i.e., CPM and the mean undercompensation for our subgroup of high spenders). The differences between PSF and CPM might be related to the fact that PSF is based on squared gaps between spending and payment while CPM uses absolute gaps. Our analysis further shows that the effect of reducing cost-control incentives displays diminishing returns in relation to PSF but increasing returns in relation to CPM. Regarding incentives for upcoding, we find that a small sacrifice of cost-control incentives already results in substantial improvements (particularly for DCGs), while for a substantial impact on the risk of PCP-level excessive losses/profits relatively high percentages of incentives for cost control must be sacrificed.

Our analysis further illustrates that the effect of adding residual-based repayments to the payment model depends on the outcome measure. Regarding incentives for risk selection, the impact

on PSF is relatively limited, but (much) stronger on CPM, the mean undercompensation for our subgroup of high spenders, incentives for upcoding (particularly for DCGs), and excessive losses/profits. Therefore, an important aspect of the specific design is what share of the total budget available for risk sharing is devoted to payments and what share to repayments. For example, sacrificing 20% of cost-control incentives by introducing residual-based payments has different consequences on the outcomes than sacrificing 20% of cost-control incentives by devoting 10% of the budget to payments and 10% to repayments. In terms of PSF, the first scenario is preferred (an increase from 20% to 89% instead of 83%), while in terms of risk for upcoding the latter scenario is favored (a reduction from 100% to 55% instead of 77% for PCGs and from 100% to 17% instead of 54% for DCGs).

Finally, we conclude that the relative effect of residual-based risk sharing on incentives for upcoding is more pronounced for DCGs than for PCGs. This finding might be related to the higher mean incremental payment for DCGs under the baseline scenario. Importantly, whether a PCP responds to incentives for upcoding, also depends on the extent to which a PCP can engage in upcoding behavior. Because PCPs probably have more influence on prescription medication (on which PCGs are based) than on diagnoses from hospital treatments (on which DCGs are based), scope for upcoding is expected to be wider for PCGs than for DCGs.

## 5.2 Implications for policy and practice

An important implication of this study is that (re)payments based on residual spending are a promising feature of global payment models and should be considered in current and future provider payment reform efforts. As we have shown, this innovative form of risk-sharing reduces the risk of unwanted effects, while minimizing the reduction in cost-control incentives. This tradeoff, however, remains complex and how it works out in practice (given a certain care package, risk-adjustment model and patient panel sizes) depends on the specific design of the risk-sharing model, the type of unwanted effect that is focused on, and the specific measure used to assess that effect. It is up to relevant stakeholders to weigh the pros and cons in terms of the incentive effects, given the context-specific economic and societal preferences. The most important question in this regard is: how much incentives for cost control is one willing to sacrifice to meaningfully mitigate adverse incentives and effects? To facilitate decision makers in making this tradeoff, insight in the incentive effects of various risk-sharing modalities is required. We believe this paper provided a useful start in gaining this insight.

## 5.3 Limitations and implications for future research

Our study has several limitations. Below, we discuss five and suggest several topics for future research. A first limitation is related to our outcome measures. An important goal of this paper was to measure incentives regarding several behaviors. Given our data, our measures could only serve as an indication of these incentives. With regard to incentives for cost control, for example, incorporating the possible (dilutional) effect of morbidity-based risk adjustment in which the

payment is based on prior diagnoses or utilization on cost-control incentives was beyond the scope of this paper (Pope et al. 2011; Douven et al. 2015; McGuire & van Kleef 2018; Geruso & McGuire 2016). Another example is the choice of risk group for our third measure of selection incentives; we focused on the financial result for persistently high prior spenders, while other, more specific groups (such as users of specific expensive, prescription medication) might have been more appropriate. A third example is that incentives for upcoding were quantified as the per person average incremental payment as a result of classification of an individual in a PCG or DCG. However, the cost providers have to incur to realize such classification is ideally incorporated in this measure. Finally, we arbitrarily defined 'excessive losses/profits' as a loss or profit exceeding the total payment by 5%. Given these shortcomings, the development and simulation of more refined measures that adequately capture incentives is an important avenue for future research.

A second limitation is that our focus was on incentives and tradeoffs therein and not on actual behavior. While incentives certainly drive behavior, how providers will respond to these incentives in practice largely remains an empirical question, which can only be answered through rigorous evaluation of real-world global payment initiatives. In practice, providers might not act on incentives for risk selection and upcoding because of, for example, their professional ethics, intrinsic motivation to provide optimal care to their patients, or fear of the consequences of strategic behavior (Eggleston 2000; Wynia et al. 2000; Ajzen 1991).

Third, we have studied one specific form of risk sharing. In practice, many other forms are possible and being applied, including risk corridors in which provider-level losses (and profits) are capped. As the incentive effects might be different, comparison of individual-level residual-based risk sharing with alternative forms of risk sharing is an important topic for further research.

Fourth, we simulated global payments for PCPs that are accountable for a comprehensive care package including hospital care under the assumption of morbidity-based risk adjustment and a contract duration of one year. Future research should focus on the interaction between risk sharing and other design aspects – including the care package, the risk-adjustment model, and multiyear contracts – in relation to the impact on (un)desired consequences.

Finally, our results are conditional on the characteristics of administrative data from one Dutch insurer. Nevertheless, more than a quarter of all Dutch citizens was included in our study and the cost patterns are very similar to patterns presented in McGuire et al. (2020a) for Germany, the entire Dutch population, and the US Marketplaces, supporting the generality of our results.

Though the results of our simulation study should thus be interpreted with caution, we believe this research has provided a useful start in gaining insight into the incentive effects and associated tradeoffs of residual-based risk sharing. In doing so, this study has contributed to the body of knowledge concerning smarter choices in provider payment system design.

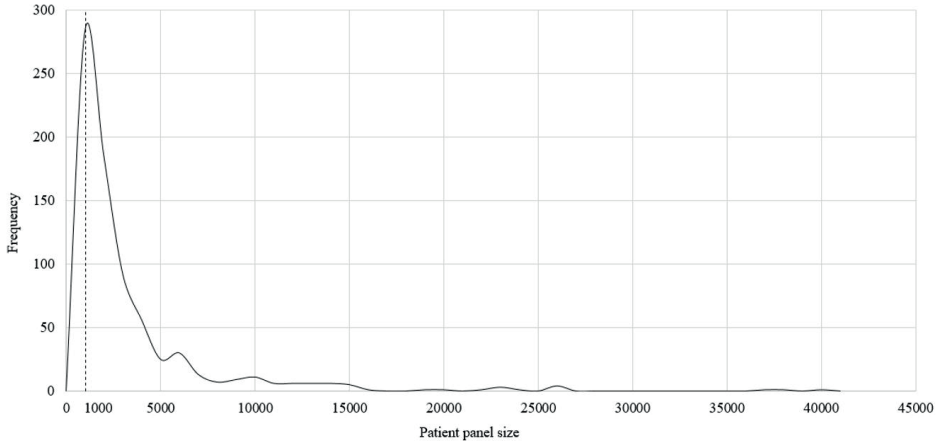
## APPENDIX

**Table A.6.1.** The number of entities, number of individuals, panel size, and medical spending for simulated PCPs

| PCPs clustered based on χ digits of PCPs' zip code | Number of entities | Number of individuals (weighted) | Mean panel size (SD) | Mean spending (SD) |
|---|---|---|---|---|
| χ = 4 | 646 | 1,542,395 | 2,388 (1,345) | 1,796 (315) |
| χ = 3 | 472 | 2,018,443 | 4,276 (5,008) | 1,650 (295) |
| χ = 2 | 90 | 2,166,308 | 24,070 (27,934) | 1,618 (224) |

Note. Number of individuals weighted by duration of insurance contract in 2012 (i.e., the number of insured years).

**Figure A.6.1.** Panel size distribution for simulated PCPs clustered based on three digits of PCPs' zip code



# Chapter 7

## Conclusions and discussion

## 1. INTRODUCTION

In this dissertation, focus has been on key conceptual and practical issues in the design of financial incentives for consumers and providers to facilitate value in health care. We synthesized evidence on the financial incentives embedded in payment systems, and conceptually and empirically analyzed important choices and tradeoffs in the design of payment systems. Findings may help stakeholders who are responsible for (re)designing existing and future cost-sharing methods for consumers and VBP models for providers in making smarter choices in payment system design. This chapter first summarizes the main conclusions of chapters 2 to 6 by answering the five research questions formulated in the introduction. Next, several implications and recommendations for policy and practice are discussed. The last section of this chapter provides some suggestions for future research.

## 2. MAIN FINDINGS

In part I of this dissertation (chapter 2) we have sought to contribute to a better understanding of VBP incentives for consumers and provided an answer to research question 1.

### Q1: How can incentives for cost-conscious behavior under various deductible designs be compared?

In chapter 2 a simulation model to approximate the relative effects of different deductible modalities on the cost-containment incentives (CCIs) is developed. Our model started from the idea that for a perfectly rational consumer the CCIs in a deductible plan depend on the marginal out-of-pocket spending given the expected spending in the contract period. We argued that the CCIs depend on (1) the probability that individual healthcare spending ends up in the relevant deductible range and (2) total expected spending given that spending ends up in the relevant deductible range. The relevant deductible range is the interval where the consumer, not the insurer, bears the costs. Ceteris paribus, CCIs are expected to reduce with the probability that spending ends up in the deductible range and higher savings potential (i.e., higher expected spending) is likely to lead to stronger CCIs. An important finding is that a deductible with an adjustable starting point based on individual's predicted healthcare spending not only results in stronger CCIs than a first-euro deductible and a doughnut hole with a uniform starting point, but also to a more equal distribution of out-of-pocket payments across consumers with low and high expected healthcare costs.

In the second part of chapter 2, our simulation model is empirically illustrated for a first-euro deductible as well as a doughnut hole with various but uniform starting points. CCIs are presented for the total population and for groups of low-risk individuals and high-risk individuals

assuming a deductible amount of €1,000. We showed that different designs result in different CCIs and incentives may differ across risk-groups. Given the data and under the assumptions made, for the total population, a doughnut hole with a uniform starting point above €0 but below €4,000 on average provides stronger CCIs than a first-euro deductible. For the low-risk individuals, this conclusion holds as long if a starting point below €3,000 is chosen. For the high-risk individuals, CCIs are (considerably) stronger under a doughnut hole with a uniform starting point compared to a first-euro deductible, even if the starting point is shifted to the right only modestly (i.e., to €500).

In part II of this dissertation we focused on VBP incentives for providers. In chapter 3 we turned to the question what the 'optimal' provider payment system in theory looks like given a five-dimensional definition of value in health care: high quality of care, cost-consciousness, well-coordinated care, cost-effective innovation, and prevention. We provided an answer to research question 2.

**Q2: What are the key design elements of a theoretically preferred value-based payment model?**

Based on a synthesis of findings of key theoretical and empirical studies on provider behavior and payment incentives, we concluded that a provider payment model that stimulates each of the five value dimensions preferably consists of two core components that must be carefully designed. The first component is a relatively large global base payment with implicit incentives for value. The second component is a relatively small variable payment with explicit incentives for value (typically: quality). Being the largest component, the base payment design is crucial but has long been largely neglected when it comes to VBP reform. The focus of chapter 3 was therefore on this component.

Our analysis revealed that the global base payment ideally consists of five key design features. In order to stimulate well-coordinated care, this payment component should be a single payment to a multidisciplinary group of healthcare providers (key design feature 1). Paying a group instead of individuals removes financial barriers between disciplines and sites, encouraging communication and cooperation across the care continuum. In addition, the global payment ideally pertains to a comprehensive set of care activities for a predefined population (key design feature 2). A global payment ideally covers all the primary and secondary care services individuals might need. Such a person-centered, holistic approach reduces fragmentation and stimulates health promotion and cost-effective prevention. Furthermore, in order to strengthen incentives for cost control and cost-effective innovation, the base payment should be fixed for a defined period (key design feature 3). Because there is no link between payment and delivered care services, providers are stimulated to contain costs.

An important consequence of the design of the base payment as described above is that providers are confronted with more financial risk than under conventional payment models. This

provides incentives for cost control but might also stimulate strategic provider behavior that may thwart value. Specifically, in order to reduce healthcare spending, providers might engage in actions to select favorable risks or skimp on quality. Therefore, the global base payment should be adjusted according to the risk profile of the population (key design feature 4) and the payment contract should include arrangements to protect providers against excessive financial risk (key design feature 5).

Under a global base payment, providers might be inclined to act too aggressively in attempts to control costs by skimping on quality or underproving necessary but expensive services. Therefore, the global base payment should be complemented with a small variable payment with explicit rewards for 'doing a good job'. This payment should trigger providers to give sufficient attention to value aspects that are unlikely to be incentivized by the global base payment but may be prone to quality skimping or underprovision. Explicit incentives should be relatively low powered to prevent a disproportionate focus on rewarded tasks. The variable payment is particularly suitable for stimulating aspects of value that can be relatively easily and objectively measured and that are difficult to incentivize implicitly. Typically, these aspects are related to high-quality care and patient-reported outcomes.

In chapter 4 we turned our attention from theory to practice. Results of a systematic review of the literature that aimed to identify and describe payment reform initiatives from practice that match the definition of a theoretically 'optimal' VBP model (i.e., a global base payment combined with explicit quality incentives) are presented. We described how these payment reform initiatives are operationalized, and their effects on spending and quality. The research question of this chapter was:

**Q3: Which initiatives exist in practice that come close to a theoretically 'optimal' value-based payment model, how are they designed, and what is their impact on value?**

We identified 18 initiatives implemented in four different countries: 15 in the US, 1 in Germany, 1 in Spain, and 1 in the Netherlands. Our analysis provides a comprehensive overview of the possibilities in terms of operationalization of the two payment components and associated design features. In most initiatives the payment is given to a large, multidisciplinary provider group consisting of various types of physicians, other healthcare providers, and facilities. Within each group, providers are jointly accountable for the provision of a comprehensive set of care activities to a delineated population. Often, these provider groups are referred to as Accountable Care Organizations (ACOs). Generally, a main contractor such as an integrated delivery system or multispecialty group practice receives the payment on behalf of the provider group and is responsible for distributing the payment and hiring individual providers. Typically, the payment covers virtually all primary and specialized medical services and prescription drugs, covered by the relevant benefit package. Sometimes, the package includes types of care beyond medical care

services only, e.g., long-term care and behavioral health care. All initiatives have a strong focus on substitution to primary care. The population can be attributed to the provider group prospectively based on prior utilization, affiliation with a provider, or region, or retrospectively based on the plurality of utilization in the completed year. A third of the 18 initiatives impose a minimum population size per provider group to reduce the effect of stochastic spending variation.

Most initiatives adopt virtual spending targets with risk-sharing arrangements built on existing (often fee-for-service like) payment and billing systems instead of actually replacing these systems with real global base payments. Most initiatives adopt multiyear contracts and apply some form of risk adjustment. Typically, initiatives adopt existing morbidity-based algorithms, originally developed in the context of risk adjustment for health plan payment. In addition to risk adjustment, a variety of risk-mitigating measures is implemented to bring financial risk for providers to acceptable levels. In about half of the initiatives, providers assume upside risk (i.e., profits) only while in the other half, providers accept downside risk (i.e., losses) as well. The risk-sharing rate varies between 50 and 100%. Most identified contracts include reinsurance provisions and carve-out some specific high-cost services from the payment contract.

We observed three main modalities of explicitly rewarding quality: add-on payments for quality (pay-for-performance), shared savings or losses dependent on quality, or a combination of these two modalities. The latter is most common in practice. A broad range of indicators is used, with clinical quality indicators being adopted most frequently. Some initiatives incorporate clinical outcome measures, patient-reported outcome measures (PROMS), or patient-reported experience measures (PREMS). The variable payment is typically low-powered (usually much lower than 10% of total payment).

Only five of the 18 initiatives have been evaluated on their impact on quality and spending. Available evaluation studies indicate that global payments in combination with variable payments for quality have the potential to improve value. Studies generally show promising results in terms of spending and quality. Importantly, these results are not the effect of payment reform only, but of a broader, multifaceted approach to value improvement that includes financial and non-financial improvement strategies.

A key question in de the design of global payments is how financial risk can be kept manageable for providers and unintended consequences can be prevented as much as possible. Answering this question requires insight in the determinants of financial risk and the interplay between these determinants. Therefore, in chapter 5 the relative impact of four key determinants related to the design of global payments on providers' financial risk was examined. The research question of this chapter was:

**Q4: Which determinants of financial risk related to global payment design can be distinguished and what is their relative impact on the financial risk of primary care providers subjected to global payments?**

We simulated prospective global payments for primary care providers (PCPs) and assessed how PCPs' financial risk depends on the scope of the care package covered by the payment, the sophistication of risk adjustment, the presence or absence of high-cost risk sharing, and patient panel size. Our primary measure of financial risk was the standard deviation of residual spending at the PCP level, with residual spending being defined as observed spending less risk-adjusted payment. In addition, we calculated PCPs' risk of ruin, defined as the probability of a PCP suffering a loss which exceeds the payment by at least 5%. To provide an answer to the research question, we relied on two large administrative datasets. The first dataset includes individual-level data on medical spending and health characteristics. The second dataset includes individual-level data obtained from a large Dutch health insurer and contains information on the PCPs that individuals were registered with.

Our simulations showed that the scope of the care package had the greatest impact on financial risk. For the narrower packages covering primary care, physiotherapy, and durable medical equipment, financial risk is relatively limited. However, irrespective of the sophistication of the risk adjustment, the use of risk sharing, and the size of the patient panel, adding prescription medication and particularly hospital care to the care package increases financial risk drastically. Our analyses further showed that morbidity-based risk adjustment is an effective measure to reduce financial risk, especially for broad care packages. Without sophisticated risk adjustment, financial accountability for comprehensive care packages would expose PCPs to excessive amounts of systematic risk. To a lesser extent than risk adjustment, full risk sharing for the 1% most costly cases can also be effective in mitigating risk, particularly when patient panels are small and the care package includes hospital care. Importantly, however, combining morbidity-based risk adjustment and high-cost risk sharing did not guarantee low levels of financial risk in absolute terms. For the care package including hospital care, more than a quarter of all PCPs in our data could be expected to suffer a loss which exceeds the payment by at least 5%. Finally, the negative impact of patient panel size on financial risk was most prominent for broad care packages in combination with morbidity-based risk adjustment but no risk sharing. We concluded that to bring financial risk for providers to appropriate levels, sufficiently large patient populations should be required.

Chapters 3, 4 and 5 have shown that both in theory and in practice, risk adjustment and risk sharing are important measures to reap the benefits of financial risk for providers under global payments while mitigating adverse effects. In chapter 6 we examined an innovative form of risk sharing, namely residual-based risk sharing. Despite its potential, this form of risk sharing has not been studied in the context of provider payment yet. Therefore, we provided insight into the incentive effects and tradeoffs associated with the design of residual-based risk sharing. We provided an answer to research question 5.

**Q5: What is the effect of residual-based risk sharing for providers on (1) incentives for cost control, (2) incentives for risk selection, (3) incentives for upcoding, and (4) excessive losses/profits for providers.**

Using the same datasets as in chapter 5, we simulated risk-adjusted global payments for PCPs for a comprehensive care package including primary care, physiotherapy, durable medical equipment, prescription medication, and hospital care. We complemented morbidity-based risk adjustment with various residual-based risk-sharing modalities that differ in the funds devoted to risk sharing and in whether only residual-based payments or both payments and repayments are used. Under this type of risk sharing, providers receive extra payments for those individuals most heavily underpaid by the risk-adjustment model and must make repayments for heavily overpaid individuals. Furthermore, in an iterative procedure we optimized the risk-adjustment payment weights for the presence of (re)payments.

Our simulation showed a substantial impact of residual-based risk sharing on cost-control incentives, risk selection incentives, upcoding incentives, and excessive provider-level losses/profits. Devoting just a very small share of total payments to residual-based payments substantially reduces incentives for risk selection as measured by Payment System Fit (PSF). The effect of adding repayments to the payment model on PSF is less prominent. Residual-based payments lead to improvements of our second measure of incentives for risk selection (i.e., Cumming's Prediction Measure; CPM) as well, but only for a relatively high degree of risk sharing. With regard to incentives for risk selection operationalized as the mean undercompensation for the group individuals belonging to the top-10-% of spenders in each of the three prior years, we observed that residual-based risk sharing only has a noticeable impact in our data if risk-sharing percentages are high. Importantly, we detected no clear pattern in the three measures for risk selection, emphasizing the importance of considering multiple measures for this outcome, given a specific context.

Incentives for upcoding are measured by the mean incremental payment for the morbidity indicators in the risk-adjustment model, i.e., pharmacy-based cost groups (PCGs) and diagnosis-based cost groups (DCGs). Our results showed that residual-based payments alone already lead to reductions in upcoding incentives, although the impact of repayments is much stronger, particularly for DCGs. Finally, we found that residual-based risk sharing can be an effective measure to reduce the share of PCPs with an excessive loss and – to a lesser extent – excessive profit, although for a substantial impact relatively high risk-sharing percentages are needed.

We concluded that less incentives for risk selection, less incentives for upcoding and less excessive losses/profits for providers, however, do come with a price. To substantially reduce the risk of unwanted effects through residual-based risk sharing, a sacrifice in incentives for cost control is required. Though small levels of risk sharing can achieve much in terms of less risk on unwanted effects, an acceptable reduction of that risk still requires a sizeable sacrifice of cost control incentives. It is up to the relevant decision makers to weigh the pros and cons of various shares of funds

devoted to residual-based payments (and potentially to repayments) in terms of incentive effects, given context-specific preferences.

## 3. IMPLICATIONS FOR POLICY AND PRACTICE

Part I of this dissertation (chapter 2) focused on the design of consumer out-of-pocket payments. A simulation model was developed and empirically illustrated to approximate the relative effects of different deductible designs on cost-containment incentives (CCIs). At least two implications can be derived. First, a deductible with an adjustable starting point based on individual's predicted healthcare spending not only results in stronger CCIs than a first-euro deductible and a doughnut hole with a uniform starting point, but also to a more equal distribution of out-of-pocket payments across consumers with low and high expected healthcare costs, confirming that such a design is an interesting design option to be (re)considered by stakeholders. Second, a doughnut hole design with a (uniform) starting point above €0 but below €4,000 on average provides stronger CCIs than a first-euro deductible implying that the starting point of the deductible should be higher than zero.

In part II of this dissertation (chapters 3 to 6) a specific manifestation of value-based provider payment reform was studied: a global base payment combined with explicit quality incentives. Such a payment model has been implemented in various settings and provides incentives for high-quality of care, cost-conscious behavior, well-coordinated care, cost-effective innovation, and prevention. Although effect studies generally show promising results we are, however, only at the beginning of the alternative payment model journey. Going forward, the insights obtained in chapters 3 to 6 may prove helpful in providing a foundation for future improvements of provider payment models. Based on our findings, at least three key implications for policy and practice can be formulated.

A first implication is that it is important to pay sufficient attention to the design of the global base payment because for several reasons the base payment constitutes the largest payment component with the strongest financial (dis)incentives for value. The scope of the care package covered by that payment component is a particularly important design aspect that designers of global payments should carefully decide on. Especially when the package covers spending on hospital care, designers should consider risk-mitigating measures to bring financial risk for providers to appropriate levels. In this regard, morbidity-based risk adjustment, (residual-based) risk sharing, and requiring sufficiently large patient populations can be highly effective measures. The latter could also imply that implementing global payments in competitive healthcare markets is most realistic in populated areas.

A second implication is that in designing VBP, decision makers should carefully consider the extent to which necessary preconditions are met, while accounting for local economic and societal preferences. This dissertation has shown that no 'one size fits all VBP design' exist that

can successfully be implemented in each setting. Adjusting the design to the specific context is a process fraught with complex tradeoffs, and it is up to decision makers to weigh the pros and cons associated with various design choices. This dissertation has provided a useful start in gaining insight in these tradeoffs. We provide three examples. First, in a setting in which providers still predominantly work in monodisciplinary 'silos', a shift from fee-for-service to global payments for multidisciplinary provider groups might be desirable but simply unfeasible, at least in the short term. Important preconditions, such as the presence of provider organizations that are able and willing to accept financial and clinical accountability for the provision of a comprehensive care package to a predefined patient population, may not be fulfilled. In that case, it seems preferable to start with adopting a less far-reaching alternative payment model than global payments, like bundled payments for specific conditions or a hybrid payment model in which part of the payment model remains fee-for-service. Second, to strengthen incentives for cost control, the global payment ideally applies to a comprehensive care package including spending on hospital care. To mitigate incentives for risk selection, morbidity-based risk adjustment is required. If, however, individual-level data on relevant population risk characteristics are not routinely available, adequate risk adjustment is unfeasible. Decision makers should then strongly consider a sacrifice in incentives for cost control by carving-out hospital care spending. Third, a tradeoff related to the local economic and societal preferences is the decision on the relative share of the global base payment and the variable quality payment. In a setting where quality of care is considered to be at an acceptable level while healthcare costs keep rising, decision makers may attach greater importance to strengthening incentives for cost control and expand the size of the global base payment. To prevent providers from acting too aggressively in attempts to control costs by skimping on quality or underproving necessary but expensive services, it is of crucial importance to carefully monitor quality of care in this context.

A third implication is that the implementation of VBP is complex and will likely require a step-by-step approach and long-term vision. It requires far-reaching changes in structures, processes, relationships, and mindsets. Building trust-based relationships and reaching consensus on, for example, the definition the care package, an appropriate form of cooperation, and the terms of the payment contract will take much of stakeholders' time and energy. In this respect, multiyear contracts with standardized contract elements can prevent stakeholders from reinventing the wheel, signal trust and a shared long-term ambition, and provide time and (financial) room to get used to bearing financial risk and invest in improvements in the care process. A related relevant finding is that it has proven possible to experiment with alternative payment models (including global payments) without replacing current payment and billing systems. Implementing a spending target with end-of-period reconciliation with fee-for-service claims can be a practical first step in moving away from volume-based payment, without the regulatory and administrative burdens of replacing current systems.

## 4. SUGGESTIONS FOR FUTURE RESEARCH

In addition to the directions for future research suggested in chapters 2 to 6, we highlight three major avenues for further study. First, only a few of the VBP initiatives included in our systematic review have been rigorously evaluated. Although available results are promising, evidence about the effects of VBP on quality and spending is scarce. Moreover, evidence is lacking on other relevant outcomes such as changes in work processes and the impact of financial incentives on the intrinsic motivation of providers. An important but understudied question in this regard is how to pass the financial incentives for value along from the provider entity receiving the payment, to the affiliated providers and individual professionals on the work floor. Furthermore, most evaluations focus on the short-term effects of payment models, while the benefits of VBP are likely to emerge after a long period of time (e.g., benefits from investments in prevention). Finally, there is still a lot to learn about barriers and facilitators to successful implementation of VBP models in various contexts. Implementation of VBP should therefore be accompanied with rigorous evaluation comprising both quantitative and qualitative methods and with careful documentation of design choices and tradeoffs made.

Second, risk adjustment and risk sharing are crucial measures to mitigate unintended effects of confronting providers with financial risk. An important direction for further research is how risk adjustment can be tailored to the specific purpose of provider payment. We have shown that risk-adjustment models used in VBP programs in practice typically make use of algorithms that were originally developed for the purpose of health plan payment. Arguably, the risk characteristics in these models may not be appropriate for risk adjusting provider payments because providers may have other incentives and tools for risk selection and other undesired behaviors. On the other hand, ethical constraints tempering such behaviors may be stronger for providers than for health plans. Design of risk adjustment for provider payment requires a better understanding of the differences in incentives, tools, and possible mitigating factors between providers and health plans. In addition, in this research only two forms of risk-sharing were studied. Many other forms are possible, including risk corridors. The design and effects of these other forms in the context of provider payment is an interesting avenue for further research.

Finally, more insight is required in the practical consequences of shifting financial risk to providers. For example, when providers bear substantial financial risk, relevant regulatory bodies might consider them as insurers and confront them with similar solvency requirements. This is an important but understudied issue. The same holds for the possible practical consequences in terms of (violation of) competition and antitrust regulations when larger provider entities with larger patient panels are developed for risk-bearing purposes.

# References

Afendulis, C., M. Fendrick, Z. Song, B.E. Landon, D.G. Safran, R.E. Mechanic & M.E. Chernew. 2014. 'The impact of global budgets on pharmaceutical spending and utilization: Early experience from the Alternative Quality Contract.' *Inquiry* 51:1-7.

Azjen, I. 1991. 'The theory of planned behavior.' *Organizational behavior and human decision processes* 50(2):179-211.

Altman, D., D.M. Cutler & R.J. Zeckhauser. 2000. 'Enrollee mix, treatment intensity, and cost in competing indemnity and HMO Plans' *Journal of Health Economics* 22(1):23-45.

Anderson, G.F. & W.E. Weller. 1999. 'Methods of reducing the financial risk of physicians under capitation.' *Archives of Family Medicine* 8(2):149-155.

APMF FPT Work Group, Alternative Payment Model Framework and Progress Tracking Work Group. 2016. *Alternative payment model (APM) framework: Final white paper.* Retrieved from https://hcp-lan.org/workproducts/apm-whitepaper.pdf.

Arrow, K.J. 1963. 'Uncertainty and the welfare economics of medical care.' *American Economic Review* 53(5):941-973.

Arrow, K.J. 1986. 'Agency and the market.' In: K.J. Arrow & M.D. Intriligator (eds.), *Handbook of Mathematical Economics*, 1183-1195. Amsterdam: Elsevier Science Publishers.

Ash, A.S. & R.P. Ellis. 2012. 'Risk-adjusted payment and performance assessment for primary care.' *Medical Care* 50(8):643-653.

Baicker, K. & D. Goldman. 2011. 'Patient cost-sharing and healthcare spending growth.' *Journal of Economic Perspectives* 25(2):47-68.

Baicker, D., S.L. Taubman, H.L. Allen, M. Bernstein, J.H. Gruber, J.P. Newhouse, E.C. Schneider, B.J. Wright. A.M. Zaslavsky & A.M. Finkelstein. 2013. 'The Oregon experiment: Effects of Medicaid on clinical outcomes.' *The New England Journal of Medicine* 368(18):1713-1722.

Bakker, F.M. 1997. *Effecten van eigen betaling op premies van zorgverzekeringen (doctoral dissertation).* Rotterdam: Erasmus University Rotterdam.

Barros, P.P. 2003. 'Cream-skimming, incentives for efficiency and payment method.' *Journal of Health Economics* 22(3):419-443.

Barry, C.L., E.A. Stuart, J.M. Donohue, S.F. Greenfield, E. Kouri, K. Duckworth, Z. Song, R.E. Mechanic, M.E. Chernew & H.A. Huskamp. 2015. 'The early impact of the 'Alternative Quality Contract' on mental health service use and spending in Massachusetts.' *Health Affairs* 34(12):2077-2085.

Bazemore, A., R.L. Phillips, Jr., R. Glazier & J. Tepper. 2018. 'Advancing primary care through alternative payment models: Lessons from the United States & Canada.' *Journal of the American Board of Family Medicine* 31(3):322-327.

Beeuwkes-Buntin, M. & A.M. Zaslavsky. 2004. 'Too much ado about two-part models and transformation? Comparing methods of modeling Medicare expenditures.' *Journal of Health Economics* 23(3):525-542.

Behrend, C., S. Felder & R. Busse. 2007. 'Susceptibility to strategy of the drug component of the IPHCC+RxGroups classification system in a risk-adjusted morbidity compensation scheme: A conceptional and data-supported analysis.' *Das Gesundheitswesen* 69(1):1-10.

Berenson, R.A. 2010. 'Shared Savings program for Accountable Care Organizations: A bridge to nowhere?' *The American Journal of Managed Care* 16(10):721-726.

Berenson, R.A. & C. Rich. 2010. 'US approaches to physician payment: The deconstruction of primary care.' *Journal of General Internal Medicine* 25(6):613-618.

Bertsimas, D., M.V. Bjarnadóttir, M.A. Kane, J.C. Kryder, R. Pandey, S. Vempala & G. Wang. 2008. 'Algorithmic prediction of healthcare costs.' *Operations Research* 56(6):1382-1392.

Berwick, D.M. 2011. 'Launching Accountable Care Organizations: The proposed rule for the Medicare Shared Savings Program.' *The New England Journal of Medicine* 364(16):e32.

Berwick, D.M. & A.D. Hackbarth. 2012. 'Eliminating waste in US health care.' *The Journal of the American Medical Association* 307(14):1513-1516.

Berwick, D.M., T.W. Nolan & J. Whittington. 2008. 'The Triple Aim: Care, health, and cost.' *Health Affairs* 27(3):759-769.

Blomqvist, A. 1991. 'The doctor as double agent: Information asymmetry, health insurance, and medical care.' *Journal of Health Economics* 10(4):411-432.

Blough, D.K., C.W. Madden & M.C. Hornbrook. 1999. 'Modeling risk using generalized linear models.' *Journal of Health Economics* 18(2):153-171.

Bodenheimer, T. & L.B. Casalino. 1999. 'Primary care physicians should be coordinators, not gatekeepers.' *Journal of the American Medical Association* 281(21):2045-2049.

Borza, J., M.K. Oerline, T.A. Skolarus, E.C. Norton, J.B. Dimick, B.L. Jacobs, L.A. Herrel, C. Ellimoottil, J.M. Hollingsworth, A.M. Ryan, D.C. Miller, V.B. Shahinian & B.K. Hollenbeck. 2019. 'Association between hospital participation in Medicare Shared Savings Program accountable care organizations and readmission following major surgery.' *Annals of Surgery* 269(5):873-878.

Bovbjerg, R.R. 1992. 'Reform of financing for health coverage: What can reinsurance accomplish?' *Inquiry* 29(2):158-175.

Brilleman, S.L., H. Gravelle, S. Hollinghurst, S. Purdy, C. Salisbury & F. Windmeijer. 2014. 'Keep it simple? Predicting primary health care costs with clinical morbidity measures.' *Journal of Health Economics* 35(100):109-122.

Brot-Goldberg, Z.C., A. Chandra, B.R. Handel & J.T. Kolstad. 2015. 'What does a deductible do? The impact of cost-sharing on health care prices, quantities, and spending dynamics.' *NBER working paper series 21632*.

Burwell, S.M. 2015. 'Setting value-based payment goals: HHS efforts to improve U.S. health care.' *The New England Journal of Medicine* 372(10):897-899.

Busch, A.B., H.A. Huskamp & J.M. McWilliams. 2016. 'Early efforts by Medicare accountable care organizations have limited effect on mental illness care and management.' *Health Affairs* 25(7):1247-1256.

Busse, R. & J. Stahl. 2014. 'Integrated care experiences and outcomes in Germany, the Netherlands, and England.' *Health Affairs* 33(9):1549-1558.

Campbell, S.M., D. Reeves, E. Kontopantelis, B. Sibbald & M. Roland. 2009. 'Effects of pay for performance on the quality of primary care in England.' *New England Journal of Medicine* 361(4):368-378.

Carter, R.L. (ed.) .1983. *Reinsurance.* Dordrecht: Springer Science and Business Media.

Casalino, L. 2001. 'Canaries in a coal mine: California physician groups and competition.' *Health Affairs* 20(4): 97-108.

Casalino, L.P., N. Erb, M.S. Joshi & S.M. Shortell. 2015. 'Accountable Care Organizations and population health organizations.' *Journal of Health Politics, Policy and Law* 40(4):819-835.

Cattel, D., F. Eijkenaar & F.T. Schut. 2020a. 'Value-based provider payment: Towards a theoretically preferred design.' *Health Economics, Policy and Law* 15(1):94-112.

Cattel, D. & F. Eijkenaar. 2020b. 'Value-based provider payment initiatives combining global payments with explicit quality incentives: A systematic review.' *Medical Care Research and Review* 77(6):511-537.

Cattel, D. & F. Eijkenaar. 2020c. 'How to manage financial risk for capitated primary care providers? The impact of care package, risk adjustment, risk sharing, and patient panel size.' *Preparing for submission.*

Cattel, D., F. Eijkenaar, K. Ahaus & M. Van der Laar. 2021. 'Bundelbekostiging in de zorg mogelijk, ondanks belemmeringen.' *ESB* 106(4794):86-89.

CMS, Centers for Medicare and Medicaid Services. 2017. *Request for applications Next Generation ACO Model*. Retrieved from https://innovation.cms.gov/initiatives/Next-Generation-ACO-Model/Archived-Materials.html.

CMS, Centers for Medicare and Medicaid Services. 2018. *Next Generation ACO Model*. Retrieved from https://innovation.cms.gov/initiatives/Next-Generation-ACO-Model/.

CMS, Centers for Medicare and Medicaid Services. 2019. *HHS news: HHS to deliver value-based transformation in primary care.* Retrieved from https://www.cms.gov/newsroom/press-releases/hhs-news-hhs-deliver-value-based-transformation-primary-care.

CMS, Centers for Medicare and Medicaid Services. 2020. *Bundled Payments for Care Improvement (BPCI) Initiative: General information.* Retrieved from https://innovation.cms.gov/innovation-models/bundled-payments.

Chang, R.E., S.P. Lin & D.C. Aron. 2012. 'A pay-for-performance program in Taiwan improved care for some diabetes patients, but doctors may have excluded sicker ones.' *Health Affairs* 31(1):93–102.

Chee, T.T., A.M. Ryan, J.H. Wasfy & W.B. Borden. 2016. 'Current state of value-based purchasing programs.' *Circulation* 133(22):2197-2205.

Chen, T.T., K.P. Chung, I.C., Lin & M. Lai. 2011. 'The unintended consequence of a diabetes mellitus pay-for-performance program in Taiwan: Are patients with more comorbidities or more severe conditions likely to be excluded from the P4P-program?' *Health Services Research* 46:4-60.

Chernew, M.E., R.E. Mechanic, B.E. Landon & D.G. Safran. 2011. 'Private-payer innovation in Massachusetts: The Alternative Quality Contract.' *Health Affairs* 30(1):51-61.

Chernew, M.E., P.H. Conway & A.B. Frakt. 2020. 'Transforming Medicare's payment systems: Progress shaped by the ACA.' *Health Affairs* 39(3):413-420.

Chien, A.T., Z. Song, M.E. Chernew, B.E. Landon, B.J. McNeil, D.G. Safran& M. Schuster. 2014. 'Two-year impact of the Alternative Quality Contract on pediatric health care quality and spending.' *Pediatrics* 133(1):96-104.

Christianson, J.B. & D. Conrad. 2011. 'Provider payment and incentives.' In: S. Glied & P. Smith (eds.), *The Oxford handbook of economics*, 624-648. New York: Oxford University Press.

Colla, C.H., V.A. Lewis, L.S. Kao, A.J. O'Malley, C.H. Chang & E.S. Fisher. 2016. 'Association between Medicare accountable care organization implementation and spending among clinically vulnerable beneficiaries.' *The Journal of the American Medical Association Internal Medicine* 176(8):1167–1175.

Conrad, D.A. 2015. 'The theory of value-based payment incentives and their application to health care.' *Health Services Research* 50(Suppl 2):2057-2089.

Conrad, D.A., D. Grembowski, S.E. Hernandez, B. Lau & M. Marcus-Smith. 2014. 'Emerging lessons from regional and state innovation in value-based payment reform: Balancing collaboration and disruptive innovation.' *The Milbank Quarterly* 92(3):568-623.

Conrad, D.A. & L. Perry. 2009. 'Quality-based financial incentives in health care: Can we improve quality by paying for it?' *Annual Review of Public Health* 30(1):357–71.

Conrad, D.A., M. Vaughn, D. Grembowski & M. Marcus-Smith. 2016. 'Implementing value-based payment reform: A conceptual framework and case examples.' *Medical Care Research and Review* 73(4):437-457.

Cutler, D.M. & K. Ghosh. 2012. 'The potential for cost savings through bundled episode payments.' *The New England Journal of Medicine* 366(12):1075-1077.

Cutler, D.M. & R.J. Zeckhauser. 1998. 'Adverse selection in health insurance.' In: A.M. Garber (ed.), *NBER book series frontiers in health policy research volume 1*, 1-32. Cambridge, Massachusetts: MIT Press.

De Bakker, D.H., J.N. Struijs, C.B. Baan, J. Raams, J.E. de Wildt, H.J.M. Vrijhoef & F.T. Schut. 2012. 'Early results from adoption of bundled payment for diabetes care in the Netherlands show improvement in care coordination.' *Health Affairs* 31(2):426-433.

De Brantes, F. & A. Rastogi. 2008. 'Evidence-informed case rates: Paying for safer, more reliable care.' *Issue brief Commonwealth Fund* 1146(40):1-13.

De Brantes, F., M.B. Rosenthal & M. Painter. 2009. 'Building a bridge from fragmentation to accountability: The Prometheus Payment model.' *The New England Journal of Medicine* 361(11):1033-1036.

Deci, E.L., R. Koestner & R.M. Ryan. 1999. 'A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation.' *Psychological Bulletin* 125(6):627-668.

DeGruy, F.V. & R.S. Etz. 2010. 'Attending to the whole person in the patient-centred medial home: The case for incorporating mental health care, substance abuse care, and health behaviour change.' *Families, Systems & Health: The Journal of collaborative family healthcare* 28(4):298-307.

Donabedian, A. 1988. 'The quality of care: How can it be assessed?' *The Journal of the American Medical Association* 260(12):1743-1748.

Douven, R., T.G. McGuire & J.M. McWilliams. 2015. 'Avoiding unintended incentives in ACO payment models.' *Health Affairs* 34(1):143-149.

Dranove, D., D. Kessler, M. McClellan, M. & M. Satterthwaite. 2003. 'Is more information better? The effects of "report cards" on health care providers.' *The Journal of Political Economy* 111(3):555-588.

Drewes, H.W., N.J.E. van Vooren, B. Steenkamer, P.F. Kemper. R.J. Hendrikx & C.A. Baan. 2018. *Regio's in beweging naar een toekomstbestendig gezondheidssysteem: Landelijke Monitor Proeftuinen – reflectie op 5 jaar proeftuinen.* Retrieved from https://www.rivm.nl/bibliotheek/rapporten/2018-0140.pdf

Duan, N., W.G., Manning, C.N. Morris & J.P. Newhouse. 1983. 'A comparison of alternative models for the demand for medical care.' *Journal of Business and Economic Statistics* 1(2):115-126.

Dudley, R.A. & H.S. Luft. 2001. 'Managed care in transition.' *The New England Journal of Medicine* 344(14):1087-1092.

Eggleston, K. 2000. 'Risk selection and optimal health insurance-provider payment systems.' *The Journal of Risk and Insurance* 67(2):173-196.

Eggleston, K. 2005. 'Multitasking and mixed systems for provider payment.' *Journal of Health Economics* 24(1):211-223.

EIB, Evaluatiecommissie Integrale Bekostiging. 2012. *Eindrapport van de Evaluatiecommissie Integrale Bekostiging. Integrale bekostiging van zorg: Werk in uitvoering*. Den Haag: Evaluatiecommissie Integrale Bekostiging.

Eijkenaar, F. 2013a. *Pay-for-performance for healthcare providers: Design, performance measurement, and (unintended) effects (doctoral dissertation)*. Rotterdam: Erasmus University Rotterdam.

Eijkenaar, F. 2013b. 'Key issues in the design of pay-for-performance programs.' *European Journal of Health Economics* 14(1):117-131.

Eijkenaar, F., M. Emmert, M. Scheppach & O. Schöffski. 2013. 'Effects of pay-for-performance in health care: A systematic review of systematic reviews.' *Health Policy* 10(2-3):115-130.

Eijkenaar, F. & F.T. Schut. 2015. *Uitkomstenbekostiging in de zorg: Een (on)begaanbare weg?* Rotterdam: Erasmus Universiteit Rotterdam.

Eijkenaar, F. & R.C.J.A. van Vliet. 2014. 'Performance profiling in primary care: Does the choice of statistical model matter?' *Medical Decision Making* 34(2):192-205.

Eisenhardt, K.M. 1989. 'Agency theory: An assessment and review.' *Academy of Management Review* 14(1):57-74.

Ellis, R.P. 1998. 'Creaming, skimping and dumping: Provider competition on the intensive and extensive margins.' *Journal of Health Economics* 17(5):537-555.

Ellis, R. P. & T. G. McGuire. 1986. 'Provider behaviour under prospective reimbursement.' *Journal of Health Economics* 5(2):129-151.

Ellis, R.P. & T.G. McGuire. 1988. 'Insurance principles and the design of prospective payment methods.' *Journal of Health Economics* 7(3):215-237.

Ellis, R.P. & T.G. McGuire. 2007. 'Predictability and predictiveness in health care spending.' *Journal of Health Economics* 26(1):25-48.

Ellis, R.P. & M.M. Miller. 2008. 'Provider payment methods and incentives.' In: H.K. Heggenhougen & S.R. Quah (eds.), *International Encyclopaedia of Public Health*, 395-402. Amsterdam: Elsevier Inc.

Enthoven, A.C. 1988. *Theory and practice of managed competition in health care finance.* Amsterdam: Elsevier Science.

Enthoven, A.C. 2009. 'Integrated Delivery Systems: The cure for fragmentation.' *American Journal of Managed Care* 15(Suppl 10):S284-S290.

Epping-Jordan J.E., S.D. Pruitt, R. Bengoa & E.H. Wagner. 2004. 'Improving the quality of health care for chronic conditions.' *Quality and Safety in Health Care* 13(4):299-305.

European Commission. 2019. *Defining value in value-based healthcare.* Retrieved from https://ec.europa.eu/health/sites/health/files/expert_panel/docs/024_defining-value-vbhc_en.pdf.

Evans, R.G. 1974. 'Supplier-induced demand: Some empirical evidence and implications.' In: M. Perlman (ed.), *The Economics of Medical Care,* 162-173. London: Macmillan.

Feldstein, M.S. 1973. 'The welfare loss of excess health insurance.' *The Journal of Political Economy* 81(2):251-280.

Folland, S., A.C. Goodman & M. Stano. 2013. *The economics of health and health care.* Boston: Pearson Education.

Frakt, A.B. & R. Mayes. 2012. 'Beyond capitation: How new payment experiments seek to find the sweet spot in amount of risk providers and payers bear.' *Health Affairs* 31(9):1951-1958.

Frank, R.G. & J.R. Lave. 1989. 'A comparison of hospital responses to reimbursement policies for Medicaid psychiatric patients.' *The Rand Journal of Economics* 20(4):588-600.

Frank, R.G. & T.G. McGuire. 1998. 'The economic functions of carve outs in managed care.' *The American Journal of Managed Care* 4(25):31-39.

Frederick, S., G. Loewenstein & T. O'Donoghue. 2002. 'Time discounting and time preference: A critical review.' *Journal of Economic Literature* 40(1):351-401.

Friedberg, M.W., P.G. Chen, M.M. Simmons, T. Sherry, P. Mendel et al. 2020. 'Effects of health care payment models on physician practice in the United States: Follow-up study.' *RAND Health Quarterly* 9(1):1-96.

Frølich, A., J.A. Talavera, P. Broadhead & R.A. Dudley. 2007. 'A behavioral model of clinician responses to incentives to improve quality.' *Health Policy* 80(1):179-193.

Gaynor, M., J.B. Rebitzer & L.J. Taylor. 2004. 'Physician incentives in health maintenance organizations.' *The Journal of Political Economy* 112(4):915-931.

Gaynor, M. & P. Gertler. 1995. 'Moral hazard and risk spreading in partnerships.' *The RAND Journal of Economics* 26(4):591-613.

Georgescu, I. & F.G.H. Hartmann. 2013. 'Sources of financial pressure and upcoding behavior in French public hospitals.' *Health Policy* 110:156-163.

Geruso, M. & T.J. Layton. 2015. *Upcoding: Evidence from Medicare on squishy risk adjustment* (NBER Working Paper 21222). Retrieved from http://www.nber.org/papers/w21222.pdf.

Geruso, M. & T.J. Layton. 2017. 'Selection in insurance markets and its policy remedies.' *Journal of Economics Perspectives* 31(4):23-50.

Geruso, M. & T.J. Layton. 2020. 'Upcoding: evidence from Medicare on squishy risk adjustment.' *Journal of Political Economy* 128(3):984-1026.

Geruso, M. & T.G. McGuire. 2016. 'Tradeoffs in the design of health plan payment systems: Fit, power and balance.' *Journal of Health Economics* 47:1-19.

Gilfillan, R.J., J. Tomcavage, M.B. Rosenthal, D.E. Davis, J. Graham, J.A. Roy, S.B. Pierdon, F.J. Bloom Jr., T.R. Graf, R. Goldman, K.M. Weikel, B.H. Hamory, R.A. Paulus & G.D. Steele Jr. 2010. 'Value and the medical home: Effects of transformed primary care.' *The American Journal of Managed Care* 16(8):607-614.

Ginsburg, P.B. & K.K. Patel. 2017. 'Physician payment reform: Progress to date.' *New England Journal of Medicine* 377(3):285-292.

Glickman, S.W., F. Ou, E.R. DeLong, M. Roe, B. Lytle, J. Mulgund, J. Rumsfeld, W. Gibler, E.M. Ohman, K. Schulman & E.D. Peterson. 2007. 'Pay for performance, quality of care, and outcomes in acute myocardial infarction.' *Journal of the American Medical Association* 297(21):2373-2380.

Gosden, T., F. Forland, I.S. Kristiansen, M. Sutton, B. Leese, A. Giuffrida, M. Sergison & L. Pedersen. 2000. 'Capitation, salary, fee-for-service and mixed systems of payment: Effects on the behaviour of primary care physicians.' *Cochrane Database on Systematic Reviews* 3:CD002215.

Harris, M. & A. Raviv. 1978. 'Some results on incentive contracts with applications to education and employment, health insurance and law enforcement.' *American Economic Review* 68(1):20-30.

Hart, O. 2003. 'Incomplete contracts and public ownership: Remarks and an application to public-private partnerships.' *The Economic Journal* 113(486):C69-C76.

Hartman, M., A.B. Martin, D. Lassman, A. Catlin & the National Health Expenditure Accounts Team. 2015. 'National health spending in 2013: Growth slows, remains in step with the overall economy.' *Health Affairs* 34(1):150-160.

Hayen, A.P., M.J. van den Berg, B.R. Meijboom, J.N. Struijs & G.P. Westert. 2015. 'Incorporating shared savings programs into primary care: From theory to practice.' *BMC Health Services Research* 15(580):1-15.

Hayen, A.P., M.J. van den Berg, J.N., Struijs & G.P. Westert. 2021. 'Dutch shared savings program targeted at primary care led to significant reduction in total medical spending.' *Health Policy* 125(4):489-494.

Higgins, J.P.T. & S. Green (eds.). 2011. *Cochrane handbook for systematic reviews of interventions*. West Sussex, England: John Wiley and Sons ltd.

Hildebrandt, H., C. Hermann, R. Knittel, M. Richter-Reichhelm, A. Siegel & W.S. Witzenrath. 2010. 'Gesundes Kinzigtal Integrated Care: Improving population health by a shared health gain approach and a shared savings contract', *International Journal of Integrated Care* 10(2): 1-15.

Hildebrandt, H., T. Schulte & B. Stunder. 2012. 'Triple aim in Kinzigtal, Germany: Improving population health, integrating health care and reducing costs of care – lessons for the UK?' *Journal of Integrated Care* 20(4):205-222.

Hofer, T.P., R.A. Hayward, S. Greenfield, E.H. Wagner, S.H. Kaplan & W.G. Manning. 1999. 'The unreliability of individual physician report cards for assessing the costs and quality of care of a chronic disease.' *Journal of the American Medical Association* 281(22):2098-2105.

Hsieh, H.M., S.L. Tsai, L.W. Mau & H.C. Chiu. 2016. 'Effects of changes in diabetes pay-for-performance incentive design on patient risk selection.' *Health Services Research* 51(2):667-686.

Holmstrom, B. & P. Milgrom. 1991. 'Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design.' *Journal of Law, Economics, and Organization* 7(1):24-52.

Huskamp, H.A., S.F. Greenfield, E.A. Stuart, J.M. Donohue, K. Duckworth, E.M. Kouri, Z. Song, M.E. Chernew & C.L. Barry. 2016. 'Effects of global payment and accountable care on tobacco cessation service use: An observational study.' *Journal of General Internal Medicine* 31(10):1134-11140.

Hussey, P.S., M.S. Ridgely & M.B. Rosenthal. 2011. 'The PROMETHEUS bundled payment experiment: Slow start shows problems in implementing new payment models.' *Health Affairs* 30(11):2116-2124.

IBM. 2017. *Bundled payment.* Retrieved from https://www.ibm.com/downloads/cas/YWBML8OR.

Iezzoni, L.I. (ed.). 2003. *Risk adjustment for measuring health care outcomes.* Chicago: Health Administration Press.

IOM, Institute of Medicine. 2001. *Crossing the quality chasm: A new health system for the 21st century.* Washington DC: National Academies Press.

Jegers, M., K. Kesteloot, D. de Graeve & W. Gilles. 2002. 'A typology for provider payment systems in health care.' *Health Policy* 60(3):255-273.

Jensen, M.C. & W.H. Meckling. 1976. 'Theory of the firm: Managerial behaviour, agency costs and ownership structure.' *Journal of Financial Economics* 3(4):305-360.

Kahneman, D. & A. Tversky. 1979. 'Prospect theory: An analysis of decision under risk.' *Econometrica* 47(2):263-292.

Kaufman, B.G., B.S. Spivack, S.C. Stearns, P.H. Song & E.C. O'Brien. 2019. 'Impact of accountable care organizations on utilization, care, and outcomes: A systematic review.' *Medical Care Research and Review* 76(3):255-290.

Kay, A. 2002. 'The abolition of the GP fundholding scheme: A lesson in evidence-based policy making.' *British Journal of General Practice* 52(475):141-144.

Keeler, E.B., J.P. Newhouse & C.E. Phelps. 1977. 'Deductibles and the demand for medical care services: The theory of a consumer facing a variable price schedule under uncertainty.' *Econometrica* 45(3):641 656.

Kelleher, K.J., J. Cooper, K. Deans, P. Carr, R.J. Brilli, S. Allen & W. Gardner. 2015. 'Cost savings and quality of care in a pediatric accountable care organization.' *Pediatrics* 135(3):e582-e589.

Kindig, D. A. 2007. 'Understanding population health terminology', *The Milbank Quarterly,* 85(1):139-161.

Kringos, D., W. Boerma, Y. Bourgueil, T. Cartier, T. Dedeu, T. Hasvold, A. Hutchinson, M. Lember, M. Oleszczyk, D. Rotar Pavlic, I. Svab, P. Tedeschi, S. Wilm, A. Wilson, A. Windak, J. van der Zee & P. Groenewegen. 2013. 'The strength of primary care in Europe: An international comparative study.' *British Journal of General Practice* 63(616):e742-750.

Laffont, J. & D. Martimort. 2002. *The theory of incentives: The principal-agent model.* Princeton: Princeton University Press.

Laffont, J. & J. Tirole. 1993. *A theory of incentives in procurement and regulation.* Cambridge, Massachusetts: MIT Press.

Lamers, L.M., R.C.J.A. van Vliet & W.P.M.M. van de Ven. 1999. *Farmacie Kosten Groepen: Een verdeelkenmerk voor normuitkeringen gebaseerd op medicijngebruik in het verleden (Pharmacy-Based Cost Groups: A risk adjuster for capitation payments based on the utilization of prescribed drugs).* Rotterdam: Erasmus University Rotterdam.

LHV, Landelijke Huisartsen Vereniging. 2018. *Feiten en cijfers huisartsenzorg.* Retrieved from https://www.lhv.nl/uw-beroep/over-de-huisarts/kerncijfers-huisartsenzorg.

Landon, B.E. 2014. 'Structuring payments to patient-centered medical home.' *Journal of the American Medical Association* 213(16):1633-1634.

Landon, B.E. & R.E. Mechanic. 2017. 'The paradox of coding – Policy concerns raised by risk-based provider contracts.' *The New England Journal of Medicine* 377(13):1211-1213.

Layton, T.J. & T.G. McGuire. 2016. 'Marketplace plan payment options for dealing with high-cost enrollees.' *American Journal of Health Economics* 3(2):165-191.

Layton, T.J., T.G. McGuire & A.D. Sinaiko. 2016. 'Risk corridors and reinsurance in health insurance marketplaces: Insurance for insurers.' *American Journal of Health Economics* 2(1):66-95.

Leijten, F.R.M., V. Struckmann, E. van Ginneken, T. Czypionka, M. Krauss, M. Reiss, A. Tsiachristas, M. Boland, A. de Bont, R. Bal, R. Busse & M.P.M.H. Rutten-van Mölken. 2017. 'The SELFIE framework for integrated care for multi-morbidity: Development and description.' *Health Policy* 122(1):12-22.

Lewis, R. 2004. *Practice-led commissioning: Harnessing the power of the primary care frontline.* Retrieved from https://www.kingsfund.org.uk/sites/default/files/field/field_publication_file/practice-led-commissioning-harnessing-power-primary-care-frontline-richard-lewis-kings-fund-1-june-2004.pdf.

Lewis, V.A., A.B. McGlurg, J. Smith, E.S. Fisher & J.P.W. Bynum. 2013. 'Attributing patients to Accountable Care Organizations: Performance year approach aligns stakeholders' interests.' *Health Affairs* 32(3):587-595.

Lindenauer, P.K., D. Remus, S. Roman, M.B. Rothberg, E.M. Benjamin, A. Ma & D.W. Bratzler. 2007. 'Public reporting and pay for performance in hospital quality improvement.' *The New England Journal of Medicine* 356(5):486-496.

Manning, W.G. & J. Mullahy. 2001. 'Estimating log models: To transform or not to transform?' *Journal of Health Economics* 20(4):461-494.

Markovitz, A.A., J.M. Hollingsworth, J.Z. Ayanian, E.C. Norton, N.M. Moloci, P.L. Yan & A. Ryan. 2019. 'Risk adjustment in Medicare ACO program deters coding increases but may lead ACOs to drop high-risk beneficiaries.' *Health Affairs* 38(2):253-261.

Marmor, T., J. Oberlander & J. White. 2011. 'Medicare and the federal budget: Misdiagnosed problems, inadequate solutions.' *Journal of Policy Analysis and Management* 30(4):928-934.

Marques, R.C. & S. Berg. 2011. 'Public-private partnership contracts: A tale of two cities with different contractual arrangements.' *Public Administration* 89(4):1585-1603.

Maskin, E. & J. Tirole. 1999. 'Unforeseen contingencies and incomplete contracts.' *Review of Economic Studies* 66(1):83-114.

McClellan, M., J. Kent, S. Beales, M. Macdonell, A. Thoumi & A. Shuttleworth. 2013. *Focusing accountability on the outcomes that matter.* Retrieved from http://www.wish-qatar.org/app/media/384.

McClellan, M., A.N. McKethan, J.L. Lewis, J. Roski & E.S. Fisher. 2010. 'A national strategy to put accountable care into practice.' *Health Affairs*, 29(5):982-990.

McConnell, K.J., A.M. Chang, D.J. Cohen, N. Wallace, M.E. Chernew, G. Kautz, D. McCarty, B. McFarland, B. Wright & J. Smith. 2014. 'Oregon's Medicaid transformation: An innovative approach to holding a health system accountable for spending growth.' *Healthcare* 2(3):163-167.

McDonald, R. & M. Roland. 2009. 'Pay for performance in primary care in England and California: Comparison of unintended consequences.' *Annals of Family Medicine* 7(2):121-127.

McGuire, T.G. 2000. 'Physician agency.' In: A.J. Culyer & J.P. Newhouse (eds.), *Handbook of Health Economics*, 461-536. Amsterdam: Elsevier Science B.V.

McGuire, T.G. 2011. 'Physician agency and payment for primary medical care.' In: S. Glied & P. Smith (eds.), *The Oxford Handbook of Health Economics*, 602-623. New York: Oxford University Press Inc.

McGuire, T.G., M. Chernew, M. McWilliams, T. Nham & S. Rose. 2021. 'Coding intensity as steady-state prevalence: Application to Medicare's Accountable Care Organizations.' *Preparing for submission.*

McGuire, T.G. & M.V. Pauly. 1991. 'Physician response to fee changes with multiple payers.' *Journal of Health Economics* 10(4):385-410.

McGuire, T.G., S. Schillo & R.C. van Kleef. 2020a. 'Reinsurance, repayments, and risk adjustment in individual health insurance: Germany, the Netherlands and the US Marketplaces.' *American Journal of Health Economics* 6(1):139-168.

McGuire, T.G., S. Schillo & R.C. van Kleef. 2020b. 'Very high and low residual spenders in private health insurance markets: Germany, The Netherlands and the U.S. Marketplaces.' *European Journal of Health Economics* 22:35-50.

McGuire, T.G. & R.C. van Kleef (eds.). 2018. *Risk adjustment, risk sharing and premium regulation in health insurance markets: Theory and practice.* London, United Kingdom: Academic Press.

McNeil, B.J., S.J. Pauker, H.C. Sox Jr & A. Tversky. 1982. 'On the elicitation of preferences for alternative therapies.' *The New England Journal of Medicine* 306(21):1259-1262.

McWilliams, J.M., E.C. Chernew, B.E. Landon & A.L. Schwartz. 2015. 'Performance differences in year 1 of Pioneer Accountable Care Organizations.' *The New England Journal of Medicine* 372(20):1927-1936.

McWilliams, J.M., L.G. Gilstrap, D.G. Stevenson, M.E. Chernew, H.A. Huskamp & D.C. Grabowski. 2017. 'Changes in postacute care in Medicare Shared Savings Program.' *The Journal of the American Medical Association Internal Medicine* 177(4):518-526.

McWilliams, J.M., L.A. Hatfield, M.E. Chernew, B.E. Landon & A. Schwartz, A. 2016. 'Early performance of accountable care organizations in Medicare.' *The New England Journal of Medicine* 374(24):2357-2366.

McWilliams, J.M., L.A. Hatfield, B.E. Landon & M.E. Chernew. 2020. 'Savings or selection? Initial spending reductions in the Medicare Shared Savings Program and considerations for reform.' *The Millbank Quarterly* 98(3):847-907.

McWilliams, J.M. L.A. Hatfield, B.E. Landon, P. Hamed & M.E. Chernew. 2018. 'Medicare spending after 3 years of the Medicare Shared Savings Program.' *The New England Journal of Medicine* 379(12):1139-1149.

McWilliams, J.M., B.E. Landon & M.E. Chernew. 2013. 'Changes in health care spending and quality for Medicare beneficiaries associated with a commercial ACO contract.' *The Journal of the American Medical Association* 310(8):829-836.

McWilliams, J.M., B.E. Landon, M.E. Chernew & A.M. Zaslavsky. 2014. 'Changes in patients' experiences in Medicare accountable care organizations.' *The New England Journal of Medicine* 371(18):1715-1724.

Mechanic, R.E. & S.H. Altman. 2009. 'Payment reform options: Episode payment is a good place to start.' *Health Affairs* 28(2):w286-w71.

Mechanic, R. & C. Tompkins. 2012. 'Lessons learned preparing for Medicare bundled payments.' *The New England Journal of Medicine* 367(20):1873-1875.

Mehrotra, A. & P. Hussey. 2015. 'Including physicians in bundled hospital care payments: Time to revisit an old idea?' *The Journal of the American Medical Association* 313(19):1907-1908.

Mendelson, A., K. Kondo, C. Damberg, A. Low, M. Motúapuaka, M. Freeman, M. O'Neil, R. Relevo & D. Kansagara. 2017. 'The effects of pay-for-performance programs on health, health care use, and processes of care: A systematic review.' *Annals of Internal Medicine* 166(5):341-353.

Miller, H.D. 2009. 'From volume to value: Better ways to pay for health care.' *Health Affairs* 28(5):1418-1428.

Miller, R.H. & H.S. Luft. 1997. 'Does managed care lead to better or worse quality of care?' *Health Affairs* 16(3):7-25.

Milstein, R. & J. Schreyögg. 2016. 'Pay for performance in the inpatient sector: A review of 34 P4P programs in 14 OECD countries.' *Health Policy* 120(10):1125-1140.

Moscucci, M., K.A. Eagle, D. Share, D. Smith, A.C. De Franco, M. O'Donnell, E. Kline-Rogers, S.M. Jani & D.L. Brown. 2005. 'Public reporting and case selection for percutaneous coronary interventions: an analysis from two large multicenter percutaneous coronary intervention databases.' *Journal of the American College of Cardiology* 45(11):1759-1765.

Mullen, K.J., R.G. Frank, M.B. Rosenthal. 2010. 'Can you get what you pay for? Pay-for-performance and the quality of healthcare providers.' *RAND Journal of Economics* 41(1):64-91.

NZa, Nederlandse Zorgautoriteit. 2012. *Marktscan huisartsenzorg.* Retrieved from https://puc.overheid.nl/nza/doc/PUC_3047_22/1/.

Newhouse, J.P. 1989. 'Do unprofitable patients face access problems?' *Health Care Financing Review* 11(2):33–42.

Newhouse, J.P. 1993. *Free for all?: Lessons from the RAND Health Insurance Experiment*. Cambridge: Harvard University Press.

Newhouse, J.P. 1996. 'Reimbursing health plans and health providers: Efficiency in production versus selection.' *Journal of Economic Literature* 34(3):1236-1263.

Newhouse, J.P., M.B. Buntin & J.D. Chapman. 1997. 'Risk adjustment and Medicare: Taking a closer look.' *Health Affairs* 16(5):26-43.

Newhouse, J.P. & D.J. Byrne. 1988. 'Did Medicare's prospective payment cause lengths of stay to fall?' *Journal of Health Economics* 7(4):413–416.

Nyman, J.A. 1999. 'The value of health insurance: The access motive.' *Journal of Health Economics* 18(2):141-152.

OECD. 2017. *Tackling wasteful spending on health.* doi: 10.1787/9789264266414-en.

Olson, M. 1965. *The logic of collective action: Public goods and the theory of groups.* Cambridge, Massachusetts: Harvard University Press.

Omachi, T.A., S.E. Gregorich, M.D. Eisner, R.A. Penaloza, I.V. Tolstykh, E.H. Yelin, C. Iribarren, R.A. Dudley & P.D. Blanc. 2013. 'Risk adjustment for health care financing in chronic disease. What are we missing by failing to account for disease severity?' *Medical Care* 51(8):740-747.

Pauly, M.V. 1968. 'The economics of moral hazard: Comment.' *The American Economic Review* 58(3):531-537.

Peikes, D., S. Dale, A. Ghosh, E.F. Taylor, K. Swankoski, A.S. O'Malley, T.J. Day, N. Duda, P. Singh, G. Angin, L.L. Sessums & R.S. Brown. 2018. 'The Comprehensive Primary Care Initiative: Effects on spending, quality, patients, and physicians.' *Health Affairs* 37(6):1-10.

Pham, H.H., P.B. Ginsburg, T.K. Lake & M.M. Maxfield. 2010. *Episode-based payments: Charting a course for health care payment reform* (Report No. Policy Analysis 1). Retrieved from National Institute for Health Care Reform website: http://nihcr.org/wp-content/uploads/2015/03/NIHCR_Policy_Analysis_No._1.pdf.

Phipps-Taylor, M. & S.M. Shortell. 2016. 'More than money: Motivating physician behavior change in Accountable Care Organizations.' *The Milbank Quarterly* 94(4):832-861.

Pimperl, A., T. Schulte, A. Mühlbacher, M. Rosenmöller, R. Busse, O. Groene, H.P. Rodriguez & H. Hildebrandt. 2017. 'Evaluating the impact of an accountable care organization in population health: The quasi-experimental design of the German Gesundes Kinzigtal.' *Population Health Management* 20(3):239-248.

Politzer, E. 2020. 'The impact of utilization thresholds in risk adjustment systems on fit and incentives for gaming.' *Preparing for submission.*

Pollack, D.A., L.E. Raney & E.R. Vanderlip. 2012. 'Integrated care and psychiatrists.' In: H.L. McQuistion, W.E. Sowers, J.M. Ranz & J.M. Feldman (eds.), *Handbook of Community Psychiatry*, 163-175. New York: Springer.

Pope, G.C., J. Kautter, R.P. Ellis, A.S. Ash, J.Z. Avanian, L.I. Iezzoni, M.J. Ingber, J.M. Levy & J. Robst. 2004. 'Risk adjustment of Medicare capitation payments using the CMS-HCC model.' *Health Care Financing Review* 25(4):119-141.

Pope, G.C., J. Kautter, M.J. Ingber, S. Freeman, R. Sekar & C. Newhart. 2011. *Evaluation of the CMS-HCC risk adjustment model: Final report.* Retrieved from https://www.cms.gov/Medicare/HealthPlans/MedicareAdvtgSpecRateStats/Downloads/Evaluation_Risk_Adj_Model_2011.pdf.

Porter, M.E. 2009. 'A strategy for health care reform: Toward a value-based system.' *The New England Journal of Medicine* 361(2):109-112.

Porter, M.E. 2010. 'What is value in health care?' *The New England Journal of Medicine* 363(26):2477-2481.

Porter, M.E. & M.S. Kaplan. 2016. 'How to pay for health care?' *Harvard Business Review* 94(7/8):88-102.

Porter, M.E. & E.O. Teisberg. 2006. *Redefining health care: Creating value-based competition on results.* Boston: Harvard Business Review.

Prendergast C. 1999. 'The provision of incentives in firms.' *Journal of Economic Literature* 37(1):7-63.

Pronovost, P.J. 2013. 'Enhancing physicians' use of clinical guidelines.' *The Journal of the American Medical Association* 310(23):2501-2502.

Qingyue, M., J. Liying & Y. Beibei. 2011. *Cost-sharing mechanisms in health insurance schemes: A systematic review.* Geneva: The Alliance for Health Policy and Systems Research, World Health Organization (WHO).

Quentin, W., A. Geissler, F. Wittenbecher, G. Ballinger, R. Berenson, K. Bloor, D.A. Forgione, P. Köpf, M. Kroneman, L. Serden, R. Suarez, J. W. van Manen & R. Busse. 2018. 'Paying hospital specialists: Experiences and lessons from eight high-income countries.' *Health Policy* 122(5):473-484.

Richardson, J. 1981. 'The inducement hypothesis: That doctors generate demand for their own services.' In: J. van der Graag & M. Perlmand (eds.), *Health, economics and health economics*, 189-214. Amsterdam: North-Holland Publishing Company.

Ridgely, M.S., D. de Vries, K.J. Bozic & P.S. Hussey. 2014. 'Bundled payment fails to gain a foothold in California: The experiment of the IHA bundled payment demonstration.' *Health Affairs* 33(8):1345-1352.

Robinson, J.C. 2001a. 'Theory and practice in design of physician payment incentives.' *Milbank Quarterly* 79(2):149-177.

Robinson, J.C. 2001b. 'The end of managed care.' *The Journal of the American Medical Association* 285(20):2622-2628.

Roblin, D.W. & M.L. Maciejewski. 2011. 'Repeat experience with the doughnut hole in Medicare Part D: When the doughnut hole becomes a tunnel. *Medical Care* 49(5):436-442.

Roland, M. & S. Campbell. 2014. 'Successes and failures of pay for performance in the United Kingdom.' *New England Journal of Medicine* 370(20):1944-1949.

Rose, S., A.M. Zaslavsky & J.M. McWilliams. 2016. 'Variation in Accountable Care Organization spending and sensitivity to risk adjustment: Implications for benchmarking.' *Health Affairs* 5(3):440-448.

Rosenthal, M.B. 2004. 'Doughnut-hole economics: Insurance often serves purposes other than risk protection.' *Health Affairs* 23(6):129-134.

Rosenthal, M.B. & R.A. Dudley. 2007. 'Pay-for-performance: Will the latest payment trend improve care?' *The Journal of the American Medical Association* 297(7):740-744.

Ross, S.A. 1973. 'The economic theory of agency: The principal's problem.' *American Economic Review* 62(2):134-139.

Ross, C. 2019. *U.S. health officials unveil experiment to overhaul primary care.* Retrieved from https://americanhealthcarechoices.org/administrative-actions/u-s-health-officials-unveil-experiment-to-overhaul-primary-care/.

Ryan, A.M., S. Krinsky, K.A. Maurer & J.B. Dimick. 2017. 'Changes in hospital quality associated with hospital value-based purchasing.' *The New England Journal of Medicine* 3769(24):2358-2366.

Ryan, A.M., S.M. Shortell, P.P. Ramsay & L.P. Casalino. 2015. 'Salary and quality compensation for physician practices participating in accountable care organizations.' *Annals of Family Medicine* 13(4):321-324.

Schillo, S., G. Lux, J. Wasem & F. Buchner. 2016. 'High cost pool or high cost groups: How to handle high(est) cost cases in a risk adjustment mechanism?' *Health Policy* 120(2):141-147.

Scott, A., M. Liu & J. Yong. 2018. 'Financial incentives to encourage value-based health care.' *Medical Care Research and Review* 75(1):3-32.

Sharfstein, J.M. 2016. 'Global budgets for rural hospitals.' *The Milbank Quarterly* 94(2):255-259.

Sharp, A.L., Z. Song, D.G. Safran, M.E. Chernew & M.A. Fendrick. 2013. 'The effect of bundled payment on emergency department use: Alternative Quality Contract effects after year one.' *Academic Emergency Medicine* 20(9):961-964.

Shen, Y. 2003. 'Selection incentives in a performance-based contracting system.' *Health Services Research* 38(2):535-552.

Shortell, S.M. 2013. 'Bridging the divide between health and health care.' *The Journal of the American Medical Association* 309(11):1121-1122.

Shortell, S.M. & L.P. Casalino. 2010. 'Implementing qualifications criteria and technical assistance for accountable care organizations.' *The Journal of the American Medical Association* 303(17):1747-1748.

Shortell, S.M., F.M. Wu, V.A. Lewis, C.H. Colla & E.S. Fisher. 2014. 'A taxonomy of accountable care organizations for policy and practice.' *Health Services Research* 49(6):1883-1899.

Silberberg, E. 1990. *The structure of economics: A mathematical analysis.* San Francisco: McGraw-Hill.

Simborg, D.W. 1981. 'DRG creep: a new hospital-acquired disease.' *The New England Journal of Medicine* 304(26):1602-1604.

Singh, J., S. Dahrouge & M.E. Green. 2019. 'The impact of the adoption of a patient rostering model on primary care access and continuity of care in urban family practices in Ontario, Canada.' *BMC Family Practice* 20(52):1-14.

Song. Z. 2014. *Payment reform in Massachusetts: Health care spending and quality in Accountable Care Organizations 4 years into global payment (doctoral dissertation).* Boston: Harvard Medical School.

Song, Z., A.M. Fendrick, D.G. Safran, B.E. Landon & M.E. Chernew. 2013. 'Global budgets and technology-intensive medical services.' *Healthcare* 1(1-2):15-21.

Song, Z., S. Rose, M.E. Chernew & D.G. Safran. 2017. 'Lower- versus higher-income populations in the Alternative Quality Contract: Improved quality and similar spending.' *Health Affairs* 36(1):74-82.

Song, Z., S. Rose, D.G. Safran. 2014. 'Changes in health care spending and quality 4 years into global payment.' *The New England Journal of Medicine* 371(18):1704-1714.

Song, Z., D.B. Safran, B.E. Landon, Y. He, R.P. Ellis, R.E. Mechanic, M.P. Day & M.E. Chernew. 2011. 'Health care spending and quality in year 1 of the Alternative Quality Contract.' *The New England Journal of Medicine* 365(10):909-918.

Song, Z., D.B. Safran, B.E. Landon, M.B. Landrum, Y. He, R.E. Mechanic, M.P. Day & M.E. Chernew. 2012. 'The Alternative Quality Contract in Massachusetts, based on global budgets, lowered medical spending and improved quality.' *Health Affairs* 31(8):1885-1894.

Song, S., B.A. Ynan Ji, D.G. Safran & M.E. Chernew. 2019. 'Health care spending, utilization, and quality 8 years into global payment.' *The New England Journal of Medicine* 381(13):252-263.

Sood, N. 2011. 'Medicare's bundled payment pilot for acute and postacute care: Analysis and recommendations on where to begin.' *Health Affairs* 30(9):1708-1717.

Spector, J.M., C. Gusland & C. Kim. 2018. *Insurance risk and its impact on provider shared risk payment models.* Retrieved from https://www.soa.org/globalassets/assets/files/resources/research-report/2018/insurance-risk-impact.pdf.

Spector, J.M., B. Studebakker & E.J. Menges. 2015. *Provider payment arrangements, provider risk, and their relationship with the cost of health care.* Retrieved from https://www.soa.org/globalassets/assets/Files/Research/Projects/research-2015-10-provider-payment-report.pdf.

Spence, A.M. & R.J. Zeckhauser. 1971. 'Insurance, information and individual action.' *American Economic Review* 61(2):380-387.

Silberberg, E. 1990. *The structure of economics: A mathematical analysis.* San Francisco: McGraw-Hill.

Stabile, M., S. Thomson, S. Allin, S. Boyle, R. Busse, K. Chevreul, G. Marchildon & Elias Mossialos. 2013. 'Health care cost containment strategies used in four other high-income countries hold lessons for the United States.' *Health Affairs* 32(4):643-652.

Starfield, B., L. Shi & J. Macinko. 2005. 'Contribution of primary care to health systems.' *The Milbank Quarterly* 83(3):457-502.

Steel, N., S. Maisey, A. Clark, R. Fleetcroft & A. Howe. 2007. 'Quality of clinical primary care and targeted incentive payments: An observational study.' *British Journal of General Practice* 57(539):449-454.

Steinbusch, P., J. Oostenbrink, J. Zuurbier & F. Schaepkens. 2007. 'The risk of upcoding in casemix systems: A comparative study.' *Health Policy* 81(2-3):289–299.

Stokes, J., V. Struckmann, S.R. Kristensen, S. Fuchs, E. van Ginneken, A. Tsiachristas, M. Rutten Rutten-van Mölken & M. Sutton. 2018. 'Towards incentivizing integration: A typology of payments for integrated care.' *Health Policy* 122(9):963-969.

Struijs, J., A. Hayen, A. & K. van der Swaluw. 2018. 'When designing bundled payments, don't ignore the lessons of behavioural economics.' *Health Affairs Blog*, doi: 10.1377/hblog20180420.640240.

Struijs, J.N., E.F. de Vries, C.A. Baan, P. van Gils & M.B. Rosenthal. 2019. 'Bundled-payment models around the world: How they work and what their impact has been.' *Issue brief Commonwealth Fund,* April 2020. doi: 10.26099/936s-0y65.

Stuart, E.A., C.L. Barry, J.M. Donohue, S.F. Greenfield, K. Duckworth, Z. Song, R. Mechanic, E.M. Kouri, C. Ebnesajjad, M.E. Chernew & H.A. Huskamp. 2017. 'Effects of accountable care and payment reform on substance use disorder treatment: Evidence from the initial three years of the Alternative Quality Contract.' *Addiction* 112(1):124-133.

Thaler, R.H. 1981. 'Some empirical evidence on dynamic inconsistency.' *Economic Letters* 8(3):201-207.

Town, R., D.R. Wholey, J. Kralewski & B. Dowd. 2004. 'Assessing the influence of incentives on physicians and medical groups.' *Medical Care Research and Review* 61(3):80S-118S.

Tsiachristas, A. 2015. *Payment and economic evaluation of integrated care (doctoral dissertation).* Rotterdam: Erasmus University Rotterdam.

Tsiachristas, A., C. Dikkers, M. Boland & M.P.M.H. Rutten-van Mölken. 2013. 'Exploring payment schemes used to promote integrated chronic care in Europe.' *Health Policy* 113(3):296-304.

Van Barneveld, E.M. 2000. *Risk sharing as a supplement to imperfect capitation: A tradeoff between selection and efficiency (doctoral dissertation).* Rotterdam: Erasmus University Rotterdam.

Van Barneveld, E.M., L.M. Lamers, R.C.J.A. van Vliet & W.P.M.M. van de Ven. 1998. 'Mandatory pooling as a supplement to risk-adjusted capitation payments in a competitive health insurance market.' *Social Science & Medicine* 47(2):223-232.

Van de Ven, W.P.M.M. 2014. 'Risk Equalization and risk adjustment, the European perspective.' In: A.J. Culyer (ed.), *Encyclopedia of Health Economics*, 281-288. San Diego: Elsevier.

Van de Ven, W.P.M.M. & R.P. Ellis. 2000. 'Risk adjustment in competitive health plan markets.' In: A.J. Culyer & J.P. Newhouse (eds.), *Handbook of Health Economics*, 1003-1092. Amsterdam: Elsevier Science.

Van de Ven, W.P.M.M. & F.T. Schut. 2010. 'Is de Zorgverzekeringswet een succes?' *TPE Digitaal* 4(1):1-24.

Van der Velden, L.F.J., A. Kasteleijn & R.J. Kenens. 2017. Cijfers uit de registratie van huisartsen: Peiling 2016. Retrieved from https://www.nivel.nl/sites/default/files/cijfers-uit-de-registratie-van-huisartsen-peiling-januari-2016.pdf.

Van Exel, N.J., M.A. Koopmanschap, W. Scholte op Reimer, L.W. Niessen & R. Huijsman. 2005. 'Cost-effectiveness of integrated stroke services.' *QJM* 98(6):415-425.

Van Kleef, R.C., F. Eijkenaar, R.C.J.A. van Vliet & W.P.M.M. van de Ven. 2018. 'Health plan payment in the Netherlands.' In: T.G. McGuire & R.C. van Kleef (eds.), *Risk adjustment, risk sharing and premium regulation in health insurance markets: Theory and practice*, 397-429. London: Academic Press.

Van Kleef, R.C., W.P.M.M. van de Ven & R.C.J.A. van Vliet. 2009. 'Shifted deductibles for high risks: More effective in reducing moral hazard than traditional deductibles.' *Journal of Health Economics* 28(1):198 209.

Van Kleef, R.C., W.P.M.M. van de Ven & R.C.J.A. van Vliet. 2010. 'Een solidair eigen risico in de zorg.' *ESB* 95(4592):522-524.

Van Kleef, R.C., W.P.M.M. van de Ven & R.C.J.A. van Vliet. 2011. 'How can we bend the cost curve? Risk-adjusting the doughnut hole to improve efficiency and equity.' *Inquiry* 48(4):313-321.

Van Kleef, R.C. & R.C.J.A. van Vliet. 2021. 'How to deal with persistently low/high spenders in health plan payment?' *Preparing for submission.*

Van Veen, S.H.C.M. 2016. *Evaluating and improving the predictive performance of risk equalization models in health insurance markets (doctoral dissertation).* Rotterdam: Erasmus University Rotterdam.

Van Veen, S.H.C.M., R.C. van Kleef, W.P.M.M. van de Ven & R.C.J.A. van Vliet. 2015a. 'Improving the prediction model used in risk equalization: Cost and diagnostic information from multiple prior years.' *European Journal of Health Economics* 16(2):201-218.

Van Veen, S.H.C.M., R.C. van Kleef, W.P.M.M. van de Ven & R.C.J.A. van Vliet. 2015b. 'Is there one measure-of-fit that fits all? A taxonomy and review of measures-of-fit for risk equalization models.' *Medical Care Research and Review* 72(2):220-243.

Van Vliet, R.C.J.A. 1992. 'Predictability of individual health care expenditures.' *The Journal of Risk and Insurance* 59(3):443-461.

Van Vliet, R.C.J.A. 2004. 'Deductibles and health care expenditures: Empirical estimates of price sensitivity based on administrative data.' *International Journal of Health Care Finance and Economics* 4(4):283–305.

Van Winssen, K.P.M., R.C. van Kleef & W.P.M.M. van de Ven. 2015. 'How profitable is a voluntary deductible in health insurance for the consumer?' *Health Policy* 119(5):688-695.

Vermaas, A. 2006. *Agency, managed care and financial risk sharing in general medical practice (doctoral dissertation).* Rotterdam: Erasmus University Rotterdam.

Versteeg, S. & R. Batenburg. 2017. *Cijfers uit de registratie van huisartsen: Peiling 2017.* Retrieved from https://www.nivel.nl/sites/default/files/pdf/Cijfers-uit-de-registratie-huisartsen-2017.pdf.

Vlaanderen, F.P., M.A. Tanke, B.R. Bloem, M.J. Faber, F. Eijkenaar, F.T. Schut & P.P.T. Jeurissen. 2019. 'Design and effects of outcome-based payment models in healthcare: A systematic review.' *European Journal of Health Economics* 20(2):217-232.

Von Eije, J.H. 1989. *Reinsurance management: A financial exposition (doctoral dissertation).* Rotterdam: Erasmus University Rotterdam.

Welch, W.P. 1999. 'Bundled Medicare payment for acute and postacute care' *Health Affairs* 17(6):69–81.

Werbeck, A., A. Wübker & N.R. Ziebarth. 2020. *Cream skimming by health care providers and inequality in health care access: Evidence from a randomized field experiment*. IZA Institute of Labor Economics Discussion Paper No. 13100. Retrieved from at http://ftp.iza.org/dp13100.pdf.

Westert, G.P., Jeurissen, P.T. & W.J.J. Assendelft. 2014. 'Why Dutch general practitioners do not put the squeeze on access to hospital care?' *Family Practice* 31(5):499-501.

Wilensky, G.R. 2014. 'Medicare physician payment reform in 2014 is looking unlikely.' *The Milbank Quarterly* 92(2):182-185.

Williams, T. & J. Yegian. 2014. Bundled payment: Learning from our failure. *Health Affairs Blog*, doi: 10.1377/hblog20140805.040596.

Winblad, U., V. Mor, J.P. McHugh & M. Rahman. 2017. 'ACO-affiliated hospitals reduced rehospitalizations from skilled nursing facilities faster than other hospitals.' *Health Affairs* 36(1):67-73.

Wynia, M.K. 2009. 'The risks of rewards in health care: How pay-for-performance could threaten, or bolster, medical professionalism.' *Journal of General Internal Medicine* 24(7):854-859.

Wynia, M.K., D.S. Cummins, J.B. van Geest & I.B. Wilson. 2000. 'Physician manipulation of reimbursement rules for patients: between a rock and a hard place.' *Journal of the American Medical Association* 283(14):1858-1865.

Young, G.J., M. Meterko, H. Beckman, E. Baker, B. White, K.M. Sautter, R. Greene, K. Curtin, B.G. Bokhour, D. Berlowitz & J.F. Burgess Jr.. 2007. 'Effects of paying physicians based on their relative performance for quality.' *Journal of General Internal Medicine* 22(6):872-876.

Zare, H. & G. Anderson. 2013. 'Trends in cost sharing among selected high-income countries: 2000-2010.' *Health Policy* 112(1):35-44.

Zhang, Y., J.M. Donohue, J.P. Newhouse & J.R. Lave. 2009. 'The effects of the coverage gap on drug spending: A closer look at Medicare Part D.' *Health Affairs* 28(2):317-325.

Zweifel, P. & W.G. Manning. 2000. 'Moral hazard and consumer incentives in health care.' In: A.J. Culyer & J.P. Newhouse (eds.), *Handbook of Health Economics Volume 1A*, 409-459. Amsterdam: Elsevier.

# Summary

Despite substantial contributions of healthcare systems to life expectancy and quality of life, in many countries it is widely recognized that there remains considerable room for improvements in the efficiency, quality, and outcomes of health care. In the context of ever-increasing healthcare expenditures, realizing more 'value' in health care has therefore increasingly become a focal point in health policy. An essential element in the transition towards more value is restructuring the financial incentives embedded in consumer and provider payment systems. There are at least two reasons for this. First, financial incentives have been convincingly shown to influence behavior. Second, predominant payment systems are ill-aligned with value. What alternative systems should look like and what this would entail in practice, however, remains poorly understood. Therefore, the main aim of this dissertation is to provide insights into key issues in the design of alternative, value-based payment incentives for consumers and providers, and in associated tradeoffs and incentive effects. In doing so, we contribute to the body of knowledge concerning smarter choices in payment system design.

**Chapter 1** introduces the topic of this dissertation and describes the relevance of focusing on the design of value-based payment incentives for consumers and providers. Incentives for (un)desired consumer and provider behavior are positioned within the theory of agency and an overview of common consumer cost sharing and provider payment models is presented. This chapter also introduces the objectives and research questions that are addressed in this dissertation.

In **chapter 2** we present our research on consumer cost-sharing methods to counteract moral hazard in health insurance markets. In many countries, policymakers are faced with choices on the specific design of cost sharing. An important factor in this decision-making process is which design is expected to lead to the strongest incentives for cost-conscious behavior. This chapter focuses on the design of one of the most popular methods of cost sharing: the deductible. As explained in this chapter, the common deductible design does not provide effective incentives for cost-conscious behavior to consumers with high expected healthcare expenses. Using rich administrative individual-level data, a simulation model is developed to compare the incentives for cost-conscious behavior for different groups of consumers under various deductible designs. We show that different designs result in different incentives for cost-conscious behavior and that these incentives may differ across risk-groups. An important finding is that a deductible with an adjustable starting point based on people's expected healthcare costs not only results in more effective incentives for cost-conscious behavior, but also to a more equal distribution of out-of-pocket payments across consumers with low and high expected healthcare costs.

In **chapter 3** we turn to the question what the 'optimal' provider payment system in theory looks like given a five-dimensional definition of value in health care: high quality of care, cost-consciousness, well-coordinated care, cost-effective innovation, and prevention. Based on synthesis of theoretical and empirical studies on provider behavior and payment incentives, we conclude that given this definition of value, a payment model ideally consists of two components: (1) a relatively large base payment that implicitly stimulates value and (2) a relatively small payment that explicitly rewards quality. Being the largest component, the base payment and its design is

crucial. We explain why this base payment ideally (1) is a single payment to a multidisciplinary group of providers, (2) pertains to a comprehensive set of care activities for a predefined population, (3) is fixed for a defined period, (4) is adjusted according to the risk profile of the population, and (5) is accompanied by risk-mitigating measures.

In **chapter 4** we turn our attention from theory to practice. Based on a systematic literature review we identify 18 provider payment reform initiatives from practice that come close to the theoretically 'optimal' design (i.e., a global base payment combined with explicit quality incentives). We describe how these payment reform initiatives are operationalized, and their effects on spending and quality. The initiatives are quite heterogenous regarding the operationalization of the two payment components and associated design features. Main commonalities between initiatives are a strong emphasis on primary care, the use of virtual spending targets with risk-sharing arrangements built on existing (often fee-for-service like) payment systems, and the application of risk adjustment and other risk-mitigating measures. Evaluated initiatives generally show promising results in terms of spending and quality, although in general methodologically sound research is scarce.

In **chapter 5** we focus on the question how a global payment can be designed such that financial risk is kept manageable for providers and unintended consequences are prevented as much as possible. In a simulation study using rich administrative data, we examine the relative impact on primary care providers' financial risk of four key determinants of that risk related to the design of the global base payment: the scope of the care package covered by the payment, the sophistication of risk adjustment, the presence or absence of risk sharing for high-cost cases, and patient panel size. We show that in our data the scope of the care package is the most important determinant of financial risk. In addition, irrespective of panel size, more sophisticated risk adjustment and applying full risk sharing for the 1% most costly cases sharply decreases risk, particularly for more comprehensive care packages. Finally, to bring financial risk for providers to appropriate levels, sufficiently large patient populations are required.

In **chapter 6** we examine an innovative form of risk sharing: residual-based risk sharing. Despite its potential, this form of risk sharing has not been studied in the context of provider payment. In this chapter we provide insight into the incentive effects and tradeoffs associated with the design of residual-based risk sharing in the presence of morbidity-based risk adjustment. Using rich administrative data, we simulate the effects of various modalities of residual-based risk sharing for primary care providers under global payment on cost-control incentives, risk selection incentives, upcoding incentives, and excessive provider-level losses/profits. We show that to substantially reduce the risk of unwanted effects through residual-based risk sharing, a sacrifice of incentives for cost control is required. Though small levels of risk sharing can achieve much in terms of less risk on unwanted effects, an acceptable reduction of that risk still requires a sizeable sacrifice of cost-control incentives. We conclude that residual-based risk sharing is a promising design feature of global provider payment models and that it is up to decision makers to make the unescapable incentive tradeoffs, given context-specific preferences.

In **chapter 7** the main findings of this dissertation are discussed and implications for policy and practice and suggestions for future research are presented. We emphasize the importance of paying sufficient attention to the design of the global base payment. Stakeholders should adjust the payment design to the specific context, while accounting for local economic and societal preferences. This is a process fraught with complex tradeoffs that should not be underestimated and requires a step-by-step approach and long-term vision. The most important suggestion for further research is to accompany implementation of VBP with rigorous evaluation comprising both quantitative and qualitative methods.

# Samenvatting

Hoewel de gezondheidszorg substantieel heeft bijgedragen aan een hogere levensverwachting en betere kwaliteit van leven, is duidelijk dat er veel ruimte voor verbetering is op het gebied van doelmatigheid, kwaliteit en uitkomsten van zorg. Het realiseren van meer 'waarde' in de zorg is de afgelopen jaren dan ook een belangrijk speerpunt geworden in gezondheidszorgbeleid. Een essentieel onderdeel hierbij is het hervormen van de financiële prikkels voor zorgconsumenten en zorgaanbieders, vanwege twee redenen. Ten eerste beïnvloeden financiële prikkels gedrag, ook in de zorg. Ten tweede wordt 'waarde' door de huidige vormgeving van financiële prikkels niet gestimuleerd, integendeel. Er is echter weinig bekend over hoe alternatieve financiële prikkels er idealiter uit zouden moeten zien. Daarom beoogt dit proefschrift inzicht te geven in belangrijke kwesties in de vormgeving van alternatieve, 'waardegedreven' financiële prikkels voor consumenten en zorgaanbieders en in de bijbehorende afwegingen en prikkelwerking. Zodoende dragen we bij aan de kennis over slimmere keuzes in de wijze van betaling voor de zorg.

**Hoofdstuk 1** biedt achtergrondinformatie over de vormgeving van waardegedreven financiële prikkels voor consumenten en zorgaanbieders. Het probleem van ongewenst consument- en zorgaanbiedergedrag wordt gepositioneerd binnen de principaal-agenttheorie. Daarnaast wordt een overzicht gegeven van de meest gangbare wijzen waarop de financiële prikkels voor consumenten en zorgaanbieders zijn vormgegeven. Ten slotte introduceert dit hoofdstuk de doelstellingen en onderzoeksvragen die in dit proefschrift aan de orde komen.

In **hoofdstuk 2** presenteren we ons onderzoek naar financiële prikkels voor consumenten in de vorm van eigen betalingen om moreel risico op zorgverzekeringsmarkten tegen te gaan. In veel landen staan beleidsmakers voor de keuze hoe eigen betalingen specifiek vorm te geven. Een van de overwegingen in dit besluitvormingsproces is welke vormgeving naar verwachting zal leiden tot de sterkste prikkels voor kostenbewust gedrag. In dit hoofdstuk staat de vormgeving van een populaire vorm van eigen betalingen centraal: het eigen risico. Uiteengezet wordt dat de huidige vormgeving van het eigen risico geen effectieve prikkels voor kostenbewust gedrag geeft aan consumenten met hoge verwachte zorgkosten. Op basis van administratieve data ontwikkelen en illustreren we een simulatiemodel waarmee prikkels voor kostenbewust gedrag kunnen worden vergeleken bij een verschillende vormgeving van het eigen risico. We laten zien dat verschillende manieren waarop het eigen risico kan worden vormgegeven resulteert in verschillende prikkels voor kostenbewust gedrag voor verschillende groepen consumenten. Een belangrijke conclusie is dat een eigen risico met een variabel startpunt dat afhankelijk is van de verwachte zorgkosten van de consument, niet alleen leidt tot effectievere prikkels voor kostenbewust gedrag, maar ook tot een meer solidaire verdeling van eigen betalingen tussen consumenten met lage en hoge verwachte zorgkosten.

In **hoofdstuk 3** gaan we in op de vraag hoe een theoretisch 'optimale' bekostiging van zorgaanbieders er uit zou zien, gegeven een vijf-dimensionale definitie van 'waarde' in de zorg: goede kwaliteit van zorg, kostenbewust gedrag, goed afgestemde zorg, kosteneffectieve innovatie en preventie. Op basis van de bevindingen van eerdere theoretische en empirische studies op het gebied van het zorgaanbiedersgedrag en financiële prikkels concluderen we dat – gegeven deze definitie

van waarde – een bekostigingsmodel idealiter bestaat uit (1) een relatief grote basiscomponent met impliciete prikkels voor waarde en (2) een relatief kleine aanvullende component met expliciete prikkels voor kwaliteit. Gezien de omvang van de basiscomponent, is juist de vormgeving van deze component cruciaal. Wij laten zien dat de basisbekostiging idealiter (1) één bedrag omvat voor een multidisciplinaire groep van zorgverleners dat (2) vast is voor een bepaalde periode, (3) betrekking heeft op een uitgebreid zorgpakket voor een bepaalde patiëntenpopulatie, (4) wordt aangepast aan het specifieke risicoprofiel van de populatie en (5) risicobeperkende maatregelen bevat. Een zodanig vormgegeven basiscomponent wordt ook wel populatiebekostiging genoemd.

In **hoofdstuk 4** verleggen we de focus van theorie naar praktijk. Op basis van een systematische review van de literatuur identificeren we 18 initiatieven met alternatieve bekostiging die overeenkomen met de theoretisch optimale vormgeving (dat wil zeggen populatiebekostiging gecombineerd met expliciete financiële prikkels voor kwaliteit). We beschrijven hoe de bekostigingsmodellen in deze initiatieven worden geoperationaliseerd en wat het effect is op kosten en kwaliteit. De initiatieven verschillen sterk in de uitwerking van de twee bekostigingscomponenten en in de specifieke kenmerken van het ontwerp. De belangrijkste overeenkomsten tussen de initiatieven zijn een sterke nadruk op eerstelijnszorg, het gebruik van normatieve prestatiedoelen met verrekening achteraf op basis van gedeclareerde kosten, een correctie voor verschillen in ziekterisico tussen patiënten of verzekerden en de toepassing van andere risicobeperkende maatregelen. Geëvalueerde initiatieven laten over het algemeen veelbelovende resultaten zien in termen van kosten en kwaliteit van zorg, hoewel het aantal beschikbare evaluatiestudies van hoge kwaliteit beperkt is.

In **hoofdstuk 5** richten we ons op de vraag hoe populatiebekostiging zodanig kan worden ontworpen dat het financiële risico voor zorgaanbieders behapbaar blijft en de kans op ongewenste neveneffecten beperkt is. In een simulatiestudie op basis van administratieve data onderzoeken we de relatieve invloed van vier determinanten van financieel risico in relatie tot de vormgeving van het populatiebekostigingsmodel op het financiële risico van eerstelijnszorgaanbieders: de omvang van het zorgpakket waarop de bekostiging betrekking heeft, de verfijndheid van het model voor risicocorrectie, volledige risicodeling voor patiënten met zeer hoge kosten, en de omvang van de patiëntenpopulatie. We laten zien dat in onze data de omvang van het zorgpakket de grootste impact heeft op het financiële risico. Daarnaast constateren we dat, ongeacht de omvang van de patiëntenpopulatie, het verbeteren van de risicocorrectie en het toepassen van hoge kosten compensatie dit risico aanzienlijk verlaagt, met name in geval van omvangrijke zorgpakketten. Ten slotte laten we zien dat omvangrijke patiëntenpopulaties nodig zijn om het financiële risico voor zorgaanbieders op een acceptabel niveau te brengen.

In **hoofdstuk 6** richten we ons op een innovatieve vorm van risicodeling in de aanwezigheid van geavanceerde risicocorrectie, te weten risicodeling op basis van residuele kosten. Ondanks de potentiële voordelen is de impact van deze specifieke vorm van risicodeling nog niet onderzocht in de context van bekostiging van zorgaanbieders. In dit hoofdstuk geven we inzicht in de (uitruil in termen van) prikkelwerking in relatie tot de vormgeving van deze specifieke vorm van risicode-

ling. Op basis van administratieve data simuleren we het effect van diverse varianten van risicodeling op basis van residuele kosten op doelmatigheidsprikkels, selectieprikkels, prikkels voor 'upcoding' en het risico op excessieve verliezen/winsten voor eerstelijnszorgaanbieders. We laten zien dat een opoffering in termen van doelmatigheidsprikkels nodig is om de kans op ongewenste neveneffecten te verminderen. Hoewel beperkte risicodeling de kans op deze ongewenste effecten al sterk doet verminderen, is een aanzienlijke opoffering van doelmatigheidsprikkels nodig om de kans te reduceren tot een acceptabel niveau. We concluderen dat risicodeling op basis van residuele kosten een veelbelovende optie is bij de vormgeving van het populatiebekostigingsmodel en dat het aan relevante beleidsmakers is om de onontkoombare voor- en nadelen in termen van prikkelwerking af te wegen, gegeven de context-specifieke voorkeuren.

In **hoofdstuk 7** worden de belangrijkste bevindingen van dit proefschrift bediscussieerd en worden de implicaties voor beleid en praktijk en suggesties voor vervolgonderzoek gepresenteerd. We benadrukken het belang van een zorgvuldige vormgeving van de basisbekostigingscomponent in plaats van enkel te focussen op de kwaliteitsbeloning. Beleidsmakers zouden bij de vormgeving van de financiële prikkels rekening moeten houden met de specifieke context, gegeven de lokale economische en maatschappelijke voorkeuren. Dit complexe proces moet niet onderschat worden en vraagt om een stapsgewijze aanpak en een langetermijnvisie. De belangrijkste suggestie voor vervolgonderzoek is om de implementatie van waardegedreven bekostiging gepaard te laten gaan met diepgaande kwantitatieve en kwalitatieve evaluatiestudies.

# Dankwoord

Promoveren is solowerk maar zonder steun kom je er niet. Daarom wil ik de volgende pagina's graag gebruiken om een aantal personen in het bijzonder te bedanken.

Allereerst Erik, mijn promotor en sectieleider van Health Systems & Insurance. Onze inhoudelijke gesprekken over mijn onderzoek zijn van grote waarde geweest voor het welslagen van dit promotieonderzoek. Het is een kunst om als expert op het gebied van de gezondheidseconomie de grote lijnen te blijven zien en jij bent op dit vlak wat mij betreft koning. Dit is echter niet waar ik je vooral erkentelijk voor ben. Mijn allereerste zin als kind was "meisje zelf doen" en eigenlijk beschrijft deze uitspraak nog steeds perfect hoe ik in elkaar steek. Jij hebt me de voor mij o zo belangrijke ruimte gegeven mijn promotietraject op eigen wijze vorm te geven en mijn eigen koers te bepalen. Dit is geen 'standaard' vierjarig promotietraject geworden met een focus op het schrijven van wetenschappelijke papers maar een langer traject waarbij ik onder andere veel onderwijs heb mogen geven en de link met 'de praktijk' heb kunnen opzoeken. Deze afwisselende werkzaamheden hebben ervoor gezorgd dat ik een breed scala aan kennis en vaardigheden heb opgedaan en hebben gemaakt dat ik mijn promotietraject met zo veel plezier heb doorlopen. Ik zal je altijd dankbaar zijn voor de vrijheid die jij mij hebt gegeven en voor jouw vertrouwen in mij – in woord en daad. Jouw (digitale) deur stond altijd open en op de momenten dat de onzekerheid mij bij de keel greep was jij in staat én bereid om rust te brengen. Dank ook voor je humor en je altijd aanwezige enthousiasme als ik weer eens met een nieuw plan aan jouw bureau stond. Een promotor zoals jij is een zeldzaamheid.

Frank, wat heb ik het met jou als copromotor getroffen! Van alle collega's ben jij verreweg de belangrijkste geweest gedurende mijn promotietraject. Uiteraard wil ik je in de eerste plaats bedanken voor jouw uitgebreide, constructieve feedback en scherpe oog voor detail. Keer op keer heb jij mijn stukken met volle aandacht gelezen en becommentarieerd en je hebt altijd ruim de tijd genomen voor onze gesprekken, hoe druk je het zelf ook had. Je hebt mij gemotiveerd om zo veel mogelijk uit de papers en analyses te halen. Naast jouw uitstekende inhoudelijke bijdragen aan mijn proefschrift wil ik je bedanken voor jouw prettige, informele wijze van begeleiden. Wat heb ik moeten lachen om jouw verhalen over de baldadige acties van Ruben en Tess. Je bent integer, benaderbaar en altijd belangstellend. Als mijn hoofd weer eens op hol sloeg, wist jij mij gerust te stellen zonder mijn zorgen te bagatelliseren. Je hebt mij op hele organische wijze begeleid naar een rol van zelfstandig onderzoeker en door de jaren heen heb ik ervaren dat we steeds gelijkwaardiger zijn gaan optrekken. Dank ook dat je, met name in de laatste fase van mijn promotietraject, zo met me mee hebt gedacht over mijn wensen voor de toekomst en loopbaan. Het is ontzettend waardevol om met iemand te kunnen sparren die je kunt vertrouwen en het beste met je voor heeft. Ik vind het een feest om onze samenwerking op het prachtige vakgebied van de bekostiging van zorgaanbieders voort te kunnen zetten!

De leden van de beoordelingscommissie – Kees Ahaus, Caroline Baan en Maureen Rutten-van Mölken – en de promotiecommissie – Arthur Hayen, Jeroen Struijs, Wynand van de Ven en Hester Lingsma – wil ik bedanken voor het beoordelen van mijn proefschrift en voor het opponeren tijdens de verdediging.

Ik wil René en Richard bedanken voor hun bijdragen aan de hoofdstukken van mijn proefschrift. René, jouw kennis en kunde op het gebied van kwantitatief onderzoek en SAS zijn ronduit indrukwekkend. Op jou kun je bouwen; afspraak is afspraak en de geleverde kwaliteit is altijd uitmuntend. Ook heb ik als collega genoten van jouw droge humor en directe wijze van communiceren. Tijdens vergaderingen spreekt jouw gezicht regelmatig boekdelen, wat deelnemen aan overleggen altijd nét wat leuker maakt. Richard, wat een plezier om mijn promotietraject met jou te mogen starten en ook weer af te mogen sluiten. Jouw enthousiasme werkt aanstekelijk en je bent voor mij een hele waardevolle collega. Ik ken maar weinig mensen die er zo goed in slagen om complexe materie op een eenvoudige en aansprekende wijze uit te leggen en hun gedachtegang zo helder kunnen verwoorden. Na afloop van onze overleggen had ik altijd meteen weer zin om aan de slag te gaan met de feedback.

Voor de empirische analyses in dit proefschrift heb ik gebruik gemaakt van de WOR-dataset en de Achmea Health Database. Dank aan het Ministerie van Volksgezondheid, Welzijn en Sport, Zorgverzekeraars Nederland en Zilveren Kruis voor het beschikbaar stellen van de data en aan de Begeleidingscommissie voor de waardevolle discussies over conceptartikelen.

Ook wil ik graag alle (oud-)collega's bij HSI, ESHPM en uit mijn bredere netwerk bedanken. Ik heb enorm geboft met alle fijne mensen om mij heen. Dank voor de prettige werksfeer, jullie hulp bij onderzoek en onderwijs en natuurlijk de gezelligheid. Een aantal (oud-)collega's wil ik graag in het bijzonder noemen.

Lieve Timo, lieve Wout, jullie zijn mijn steun en toeverlaat geweest gedurende dit promotietraject. We hebben de broodjes zalm bij New Fork duurder zien worden, maar jullie waren altijd mijn stabiele factor. In onze biotoop J8-03 deelden we dezelfde kantoorhumor (lang leve de kneuterigheid!), onze successen en ergernissen op het gebied van onderzoek en onderwijs maar ook de bijzonder mooie én moeilijke momenten in ons privéleven. Ik prijs mijzelf gelukkig dat ik jullie als vrienden heb en kan me geen betere paranimfen wensen.

Ik wil het ESHPM-projectteam risicoverevening bedanken dat ik onderdeel heb mogen uitmaken van deze geoliede machine. In tegenstelling tot het schrijven van wetenschappelijke artikelen vergt projectwerk een andere aanpak waarbij snel schakelen essentieel is. Een welkome afwisseling en leerzame ervaring! Dank gaat daarnaast uit naar alle leden van de Werkgroep Onderzoek Risicoverevening (WOR) en de Werkgroep Besluitvorming Risicoverevening (WBR) voor de prettige samenwerking.

Wynand, jouw bevlogenheid als docent en later als collega heeft indruk op mij gemaakt. Door jouw colleges waarin wetenschap en beleid samenkwamen, ben ik als student serieus gaan nadenken over een carrière bij de universiteit. Ik had nooit durven dromen dat ik een baan zou vinden die zo goed bij mij past. Daarnaast wil ik je bedanken voor alle vragen en opmerkingen bij mijn presentaties en conceptartikelen. Je beschikt over een ongekende hoeveelheid kennis over het zorgstelsel en de zeldzame kwaliteit om snel en scherp te kunnen reageren, immer positief en opbouwend verwoord. Mijn proefschrift heeft hier zonder twijfel baat bij gehad.

Marco, het gezamenlijk ontwerpen, coördineren en geven van het onderwijs in Blok 6 was voor mij even leuk als leerzaam. Op het gebied van onderwijs is er geen betere leermeester dan jij! Bedankt dat je mij wegwijs hebt gemaakt in deze wondere wereld. Ik hoop op nog vele mooie jaren van samenwerking.

Anne-Fleur, ik heb je leren kennen als een enorm warm persoon met oprechte interesse in anderen; een hele mooie eigenschap die ik koester in vriendschappen. Bedankt voor jouw belangstelling en gezelligheid. Als we afspreken vliegt de tijd voorbij (zélfs in de stromende regen) en voel ik me standaard een giechelende bakvis. Genieten! Onze gesprekken over onze loopbaan waren voor mij waardevoller dan die met de carrièrecoach. Ik ben enorm trots op wie je bent en wat je doet!

Brenda, jij bent goud! Ik ben oprecht dankbaar dat ik jou heb leren kennen. We zijn ongeveer tegelijkertijd gestart met ons promotietraject en steeds verder naar elkaar toe getrokken. Zo fijn om op het werk een vriendin te hebben met wie ik alle promotieperikelen, maar ook de bijzondere momenten in het leven thuis kan delen. Ik kan enorm genieten van je Rotterdamse directheid. Tegelijkertijd weet ik ook wat voor enorm lieve en zachte vrouw je bent. Een prachtige combinatie die maar weinigen gegund is. De wereld ligt aan je voeten!

Kayleigh, ik kijk met veel plezier terug op de tijd dat wij beiden bij HSI werkten en kamergenoten waren. Jij bezig met de afronding van jouw proefschrift, ik net gestart aan het monstertraject. Al snel zijn wij naast collega's vriendinnen geworden en hebben we lief en leed gedeeld. Ik vond het ontzettend bijzonder jouw paranimf te mogen zijn bij jouw promotie en samen met Jelte aanwezig te mogen zijn bij jouw prachtige huwelijk met Mart. Naast dat je simpelweg een fantastisch hartelijk mens bent, bewonder ik jouw ongekende drive. Ik kijk uit naar onze volgende lunches en goede gesprekken over feminisme.

Anja, bedankt voor onze gezellige telefoontjes de laatste maanden. Erg leuk om onze ervaringen rondom zwangerschap te delen. Ik wens jou, Rik en jullie kindje veel geluk toe en ik kijk er naar uit binnenkort jouw proefschrift in handen te hebben!

John, jij hebt er vanaf mijn allereerste werkdag mede voor gezorgd dat ik mij thuis heb gevoeld bij ESHPM. Ik hoor je nog "juf" roepen door de gangen van J7. Jouw positieve, betrokken houding is van onschatbare waarde voor onze organisatie, zowel voor alle studenten als voor de medewerkers. Ik gun je alle goeds van de wereld.

Ook veel dank aan mijn collega's bij jongBMG (Brenda, Dara, Iris, Kirti en Lotte) met wie ik me sterk heb gemaakt voor de positie van promovendi binnen onze School (en met wie ik heel veel pizza's heb gegeten bij het Paviljoen). Werner dank ik in dit kader voor zijn openheid als prodecaan en de constructieve samenwerking.

Dank gaat uit naar de werkgroep Bekostiging van het Linnean Initiatief, in het bijzonder naar Arie Franx voor de uitnodiging zitting te nemen in de werkgroep, Marloes Zuidgeest voor haar inspanningen als projectleider en Mart van de Laar voor zijn voorzitterschap. Ik heb het als zeer waardevol ervaren om kennis en ervaringen uit te kunnen wisselen met andere stakeholders op het gebied van waardegedreven zorg en bekostiging.

Arthur Hayen, bedankt voor jouw drive, enthousiasme en bereidheid om open na te denken over een samenwerking tussen Menzis en ESHPM. Fantastisch dat de Evidence-Based Health Insurer en BUNDLE nu vliegen en ik deel uit mag maken van het geweldige projectteam vol 'vakidioten'!

A special thanks to Thomas McGuire, Randy Ellis, and the other members of the Risk Adjustment Network for your advice regarding papers, career, and life.

Ik wil alle studenten bedanken die ik de afgelopen jaren onderwijs heb mogen geven. Het doceren heeft me zó ongelooflijk veel plezier gebracht! Zonder jullie had ik het maar een saaie boel gevonden. De collega's bij het Risbo, en dan in het bijzonder Alice, wil ik bedanken voor de nuttige trainingen en begeleiding gedurende mijn 'University Teaching Qualification' traject.

Ik bedank mijn lieve familie, schoonfamilie en vrienden voor de steun en interesse de afgelopen jaren. Margherita en Cees, onze gezellige etentjes waren een welkome afleiding tijdens dit promotietraject.

Maril, alweer 23 jaar zijn we vriendinnen (man, wat worden we oud!) en jij kent mij door-en-door. Je bent attent en niet op je mondje gevallen en dat waardeer ik zo ontzettend in je. Ik ben er heel trots op dat je het avontuur in Gambia bent aangegaan. In plaats van genoegen te nemen met de situatie waar je eigenlijk niet gelukkig mee was, ben jij in het diepe gesprongen. Een boel ervaringen rijker ben je nu gelukkig weer in het land zodat we weer in het 'echie' bij kunnen kletsen onder het genot van sushi en een flesje wijn (waar ik ten tijde van dit schrijven extreem naar uitkijk!).

Annemieke, wat ben ik blij dat ik jou ben tegengekomen in Groningen! Ik had die maandagavonden met soep-uit-zak in combinatie met Euroshopper pistolets en onze nachten in de Negende (gut-o-gut!) voor geen goud willen missen. Hoewel ons leven er inmiddels een stuk minder studentikoos uit ziet, is er in onze vriendschap niets veranderd. Ik ken niemand die mij zo goed begrijpt als jij en we hebben aan een blik genoeg. Bedankt dat je de afgelopen jaren zo belangstellend was en oprecht blij als ik weer een stapje verder was gekomen met mijn proefschrift. Je bent m'n allerliefste makker voor het leven!

Lieve pap en mam, bedankt voor jullie onvoorwaardelijke liefde en steun. De hechte en warme band die wij hebben koester ik. Ik kan jullie niet genoeg bedanken voor de fantastische basis die jullie mij hebben gegeven. Van kinds af aan heb ik de vrijheid en het vertrouwen gevoeld te doen waar mijn hart ligt. Dat heeft ervoor gezorgd dat ik nu zo 'op m'n plek ben' en weet waar ik voor sta. De jeugd die ik heb gehad, gun ik ook onze eigen kleine. Tijdens mijn promotietraject leefden jullie ontzettend mee en waren jullie er altijd voor mij met een luisterend oor (hoe lang de verhitte monoloog van mijn kant soms ook was…). Ik wil jullie ook bedanken voor de spiegel die jullie mij zo nu en dan voorhouden. Hoewel ik dit niet altijd even makkelijk vond, hebben jullie vragen me op belangrijke momenten aan het denken gezet en ervoor gezorgd dat de Truus de Mier in mij niet de overhand kreeg. Pap, mijn voorliefde voor grondige analyse, structuur en nauwkeurigheid heb ik van jou. Wat heb ik hier in mijn leven én in mijn promotietraject al een boel profijt van gehad! Ook jouw wijze lessen zoals 'denk in belangen' en 'op z'n janboerenfluitjes blijven nadenken' hebben mij ontzettend veel gebracht, precies op de momenten dat ik dacht

dat de situatie eigenlijk te complex was om te overzien. Mam, van jou heb ik geleerd dat er een enorme kracht schuilt in je kwetsbaar durven opstellen. Je hebt me laten zien dat je zakelijkheid kunt verenigen met een zacht en verzorgend karakter en dat je als powervrouw moet durven varen op je gevoel. Ook heb ik via jou, opa en oma de liefde voor koken, servies en tafelen meegekregen. Na een lange werkdag vond ik rust in het maken van een heerlijk bord dampende pasta of een gezonde salade.

Tenslotte wil ik het woord tot jou – Jelte – richten. Je bent de liefde van mijn leven. Ik ben dankbaar voor alle grootse en kleine dingen die we samen beleven. Tijdens mijn promotietraject was jij er om de mijlpalen mee te vieren (Calamaretti! Vongole! Cava!) maar ook om bij weg te kruipen als het even tegenzat. Jij maakt mijn wereld zo veel mooier en bent mijn veilige haven. Bedankt voor je luisterend oor als ik weer eens worstelde met alle ballen in de lucht houden. Jij hebt mij geleerd dat goed soms ook goed genoeg is. Hoewel ik hier nog wel eens aan herinnerd moet worden, helpt deze les me te relativeren en de soms o zo nodige rust te pakken. Ik ben dankbaar dat je me vrij laat mijn ambities na te streven en te doen waar ik gelukkig van word. Jouw aanmoediging en steun hebben gemaakt dat ik mijn eigen pad durf te kiezen en de opgetrokken wenkbrauwen en impertinente vragen van anderen beter naast me neer weet te leggen. Het is toch wij tegen de wereld! Ik kan niet wachten op mijn toekomst samen met jou en onze kleine. Later als ik groot ben, is in ene aangebroken!

Daniëlle Cattel
Nieuwerkerk a/d IJssel, juni 2021

# PhD portfolio

Name D. (Daniëlle) Cattel
Department: Erasmus School of Health Policy and Management (ESHPM)
PhD period: November 2014 – March 2021
Promotor: Prof.dr. F.T. Schut
Copromotor: dr. F. Eijkenaar

| Education | Year |
|---|---|
| Bachelor of Science (cum laude) in 'Beleid en Management voor de Gezondheidszorg', Erasmus University Rotterdam, the Netherlands | 2013 |
| Master of Science in 'Health Economics, Policy and Law', specialization in 'Health Economics', Erasmus University Rotterdam, the Netherlands | 2014 |
| University Teaching Qualification (in Dutch: Basis Kwalificatie Onderwijs), Risbo, the Netherlands (2018) | 2018 |

| International peer-reviewed publications (published or preparing for submission) | Year |
|---|---|
| Cattel, D., R.C. van Kleef & R.C.J.A. van Vliet. 2016. 'A method to simulate incentives for cost containment under various cost sharing designs: An application to a first-euro deductible and a doughnut hole.' *European Journal of Health Economics* 18: 987–1000 | 2016 |
| Cattel, D., F. Eijkenaar & F.T. Schut. 2020. 'Value-based provider payment: Towards a theoretically preferred design.' *Health Economics, Policy and Law* 15(1): 94-112. | 2020 |
| Cattel, D. & F. Eijkenaar. 2020. 'Value-based provider payment initiatives combining global payments with explicit quality incentives: A systematic review.' *Medical Care Research and Review* 77(6): 511-537. | 2020 |
| Cattel, D. & F. Eijkenaar. 2020. 'How to manage financial risk for capitated primary care providers? The impact of care package, risk adjustment, risk sharing, and patient panel size.' *Preparing for submission.* | 2020 |
| Cattel, D., F. Eijkenaar & R.C. van Kleef. 2021. 'Getting the incentives right: Simulating the effects of residual-based risk-sharing for primary care providers under global payment.' *Preparing for submission.* | 2021 |

| Dutch publications | Year |
|---|---|
| Cattel, D., F. Eijkenaar & F.T. Schut. 2019. 'Betalen voor waarde in de zorg.' *We are finance* 3:32-34. | 2019 |
| Cattel, D., F. Eijkenaar, K. Ahaus & M. Van der Laar. 2021. 'Bundelbekostiging in de zorg mogelijk, ondanks belemmeringen.' *ESB* 106(4794):86-89. | 2021 |
| Qruxx. 2021. 'Interview Daniëlle Cattel over bundelbekostiging.' Publication *Qruxx.nl* in April. | 2021 |

| Contributions to Dutch policy research projects | Year |
|---|---|
| Cattel, D., F. Eijkenaar, R.C. van Kleef, en R.C.J.A. van Vliet (2016). "Onderzoek risicoverevening 2017: Berekening normbedragen" Rapport iBMG, Rotterdam: Erasmus Universiteit. | 2016 |
| Cattel, D., F. Eijkenaar, R.C. van Kleef, en R.C.J.A. van Vliet (2016). "Onderzoek risicoverevening 2017 Overall Toets" Rapport iBMG, Rotterdam: Erasmus Universiteit. | 2016 |
| Cattel, D., F. Eijkenaar, R.C. van Kleef, en R.C.J.A. van Vliet (2016). "Onderzoek risicoverevening 2017: Gegevensfase" Rapport iBMG, Rotterdam: Erasmus Universiteit. | 2016 |
| Cattel, D., F. Eijkenaar, R.C. van Kleef, R.C.J.A. van Vliet en A.A. Withagen-Koster (2017). "Onderzoek risicoverevening 2018: Berekening Normbedragen" Rapport iBMG, Rotterdam: Erasmus Universiteit. | 2017 |

| | |
|---|---|
| Cattel, D., F. Eijkenaar, R.C. van Kleef, R.C.J.A. van Vliet en A.A. Withagen-Koster (2017). "Onderzoek risicoverevening 2018: Overall Toets" Rapport iBMG, Rotterdam: Erasmus Universiteit. | 2017 |
| Cattel, D., F. Eijkenaar, R.C. van Kleef, R.C.J.A. van Vliet en A.A. Withagen-Koster (2017). "Onderzoek risicoverevening 2018: Gegevensfase" Rapport iBMG, Rotterdam: Erasmus Universiteit. | 2017 |
| Cattel, D., Eijkenaar F, R.C. van Kleef, R.C.J.A. van Vliet en A.A. Withagen-Koster (2017). "Onderzoek risicoverevening 2018: Pre-Overall Toets" Rapport iBMG, Rotterdam: Erasmus Universiteit. | 2017 |
| Cattel, D., F. Eijkenaar, R.C. van Kleef, R.C.J.A. van Vliet en A.A. Withagen-Koster (2018). "Onderzoek risicoverevening 2019: Berekening Normbedragen" Rapport ESHPM, Rotterdam: Erasmus Universiteit. | 2018 |
| Cattel, D., F. Eijkenaar, R.C. van Kleef, R.C.J.A. van Vliet en A.A. Withagen-Koster (2018). "Onderzoek risicoverevening 2019: Overall Toets" Rapport ESHPM, Rotterdam: Erasmus Universiteit. | 2018 |
| Cattel, D., F. Eijkenaar, R.C. van Kleef, R.C.J.A. van Vliet en A.A. Withagen-Koster (2018). "Onderzoek risicoverevening 2019: Gegevensfase" Rapport ESHPM, Rotterdam: Erasmus Universiteit. | 2018 |
| Cattel, D., F. Eijkenaar, R.C. van Kleef, R.C.J.A. van Vliet en A.A. Withagen-Koster (2019). "Onderzoek risicoverevening 2020: Berekening Normbedragen" Rapport ESHPM, Rotterdam: Erasmus Universiteit. | 2019 |
| Cattel, D., F. Eijkenaar, R.C. van Kleef, R.C.J.A. van Vliet en A.A. Withagen-Koster (2019). "Onderzoek risicoverevening 2020: Overall Toets" Rapport ESHPM, Rotterdam: Erasmus Universiteit. | 2019 |
| Cattel, D., F. Eijkenaar, R.C. van Kleef, R.C.J.A. van Vliet en A.A. Withagen-Koster (2019). "Onderzoek risicoverevening 2020: Gegevensfase" Rapport ESHPM, Rotterdam: Erasmus Universiteit. | 2019 |
| Cattel, D., F. Eijkenaar, M. Oskam, R.C. van Kleef, R.C.J.A. van Vliet en A.A. Withagen-Koster (2020). "Onderzoek risicoverevening 2020: Berekening Normbedragen" Rapport ESHPM, Rotterdam: Erasmus Universiteit. | 2020 |
| Cattel, D., F. Eijkenaar, M. Oskam, R.C. van Kleef, R.C.J.A. van Vliet en A.A. Withagen-Koster (2020). "Onderzoek risicoverevening 2020: Overall Toets" Rapport ESHPM, Rotterdam: Erasmus Universiteit. | 2020 |
| Cattel, D., F. Eijkenaar, M. Oskam, R.C. van Kleef, R.C.J.A. van Vliet en A.A. Withagen-Koster (2020). "Onderzoek risicoverevening 2020: Gegevensfase" Rapport ESHPM, Rotterdam: Erasmus Universiteit. | 2020 |
| Cattel, D., F. Eijkenaar, M. Oskam, R.C. van Kleef, R.C.J.A. van Vliet en A.A. Withagen-Koster (2020). "Onderzoek risicoverevening 2021: pre Overall Toets" Rapport ESHPM, Rotterdam: Erasmus Universiteit. | 2020 |

| **PhD training** | **Year** |
|---|---|
| 'Ready within four years', Erasmus University Rotterdam | 2015 |
| 'Academic writing in English', Erasmus University Rotterdam | 2015 |
| 'Innovative teaching in health economics', international Health Economics Association | 2015 |
| 'Modelling healthcare costs and counts', international Health Economics Association | 2015 |
| 'Teaching plenary sessions', Erasmus University Rotterdam | 2015 |
| 'Keeping students' attention', Erasmus University Rotterdam | 2015 |
| 'Presenting and networking', Erasmus University Rotterdam | 2016 |
| 'SAS base programming', SAS institute Huizen | 2016 |
| 'Personal branding and networking', Erasmus University Rotterdam | 2016 |
| 'Presentation skills', Erasmus University Rotterdam | 2016 |
| 'Effective communication', Erasmus University Rotterdam | 2017 |

| | |
|---|---|
| 'Active blended learning', Erasmus University Rotterdam | 2017 |
| 'A new digital learning environment: Canvas', Erasmus University Rotterdam | 2018 |
| 'Online interactive tool experience', Erasmus University Rotterdam | 2020 |

| **Conferences and seminars** | **Year** |
|---|---|
| International Health Economics Association, Milan | 2015 |
| Risk Adjustment Network, Solothurn | 2015 |
| Risk Adjustment Network, Berlin | 2016 |
| Annual conference of the Dutch Association for Health Economics, Utrecht | 2016, 2020 |
| ESHPM/EUR research seminars, Rotterdam | 2016-2020 |
| International Health Policy Conference, London | 2017 |
| Seminar on 'Population-based payments' by the Dutch Ministry of Health, Welfare and Sport, Den Haag | 2017 |
| LoLa Health Economics Study Group, Rotterdam | 2017 |
| Conference on 'Purchasing health care' by Zorgvisie, Utrecht | 2018 |
| Bi-annual national meetings Linnean Initiative, the Netherlands | 2018-2020 |
| Conference on 'Affordable networks and population-based payments' by the Rijksacademie, Den Haag | 2019 |
| Conference 'Smarter Choices for Better Health' by the Erasmus Initiative, Rotterdam | 2019 |
| Risk Adjustment Network, digital meeting | 2020 |

| **Invited conference and seminar presentations** | **Year** |
|---|---|
| ESHPM, Rotterdam | 2015 |
| International Health Economics Association, Milan | 2015 |
| ESHPM, Rotterdam | 2016 |
| Risk Adjustment Network, Berlin | 2017 |
| Dutch Ministry of Health, Welfare and Sport, The Hague | 2017 |
| ESHPM, Rotterdam | 2017 |
| International Health Policy Conference, London | 2017 |
| Dag van de zorginkoop, Utrecht | 2018 |
| Dutch Ministry of Health, The Hague | 2018 |
| ESHPM, Rotterdam | 2019 |
| Symposium 'Betaalbare netwerken en populatiebekostiging', Den Haag | 2019 |
| ESHPM, Rotterdam | 2019 |
| Smarter Choices for Better Health, Rotterdam | 2019 |
| Vereniging voor Gezondheidseconomie, Utrecht | 2020 |

| **Projects and committees** | **Year** |
|---|---|
| Member of 'jongBMG', ESHPM's PhD association | 2015-2017 |
| Co-organizer of the post-academic course 'Risk equalization: what, why, and where do we stand?' at ESHPM, Rotterdam | 2016, 2018 |

| | |
|---|---|
| Member of the project team that calculates the risk equalization payments for health insurers in the Netherlands | 2016-present |
| Co-organizer of a multi-day in-house training for data-analysts working for a Dutch health insurer on statistics and the use of quality data | 2017 |
| Member of the external sounding board of the Dutch Health Authority on their long-term vision on the remuneration of hospital care | 2017 |
| Member of the steering committee 'Quality in education' | 2017 |
| Principal teacher quantitative research during the reform of the bachelor's program | 2017-2020 |
| Member of the Linnean Initiative workgroup 'Provider payments' | 2018-2021 |
| Member of the strategic workgroup 'Future-oriented education' | 2019 |

| **Supervising and teaching experience** | **Year** |
|---|---|
| Mentor of first-year bachelor students | 2015 |
| Tutor and supervisor in the bachelor course 'Stage: werken in de zorg' | 2015 |
| Supervisor in the premaster course 'Kwantitatief Leeronderzoek' | 2015-2016 |
| Tutor in the bachelor course 'Algemene Economie van de Gezondheidszorg' | 2015-2017 |
| Coordinator and tutor in the bachelor course 'Multivariate analyse' | 2015-2019 |
| Lecturer in the master course 'Economics and Financing of Healthcare Systems' | 2017-present |
| Coordinator, lecturer, and tutor in the bachelor course 'Marktordening in de zorg' | 2019-present |
| Bachelor thesis supervisor | 2020-present |

# About the author

Daniëlle Cattel (1990) completed a BSc. in Health Sciences in 2013 (with distinction) and obtained her master's degree in Health Economics, Policy and Law in 2014 at the Erasmus University Rotterdam. She then became a PhD student at Erasmus School of Health Policy and Management (ESHPM). Her research focusses on the design of innovative payments models for consumers and providers to enhance value in health care. The results of her research are published in several peer-reviewed scientific as well as professional journals. In addition, she presented her work to a wide range of audiences, including fellow researchers, policymakers, and students. While working on her dissertation, Daniëlle was a member of the project team that calculates the risk equalization payments for health insurers in the Netherlands and of the Linnean Initiative workgroup on value-based provider payments. As a teacher and coordinator, Daniëlle was involved in nine different courses of the bachelor's, pre-master's, and master's program of ESHPM.

Hoewel de gezondheidszorg substantieel heeft bijgedragen aan een hogere levensverwachting en betere kwaliteit van leven, is duidelijk dat er veel ruimte voor verbetering is op het gebied van doelmatigheid, kwaliteit en uitkomsten van zorg. Het realiseren van meer 'waarde' in de zorg is de afgelopen jaren dan ook een belangrijk speerpunt geworden in gezondheidszorgbeleid. Een essentieel onderdeel hierbij is het hervormen van de financiële prikkels voor zorgconsumenten en zorgaanbieders. Er is echter weinig bekend over hoe alternatieve financiële prikkels er idealiter uit zouden moeten zien. Dit proefschrift beoogt inzicht te geven in belangrijke kwesties in de vormgeving van innovatieve, 'waardegedreven' financiële prikkels voor consumenten en zorgaanbieders en in de bijbehorende afwegingen en prikkelwerking.