Let's talk about value

de Vries, E. F.

Publication date:
2020

Document Version
Publisher's PDF, also known as Version of record

LET'S TALK ABOUT VALUE

Grasping the concept of value in a population health management context

Eline Frouke de Vries

Tilburg University

541371-L-bw-deVries
Processed on: 24-2-2020

Cover idea: Eline Frouke de Vries
Cover design: Roy Hendrikx, royhendrikx.nl
Layout and design: Renske Hortensius, persoonlijkproefschrift.nl
Printed by Ipskamp Printing, proefschriften.net
ISBN/EAN: 978-94-028-1911-3

LET'S TALK ABOUT VALUE

Grasping the concept of value in a population health management context

Proefschrift ter verkrijging van de graad van doctor aan Tilburg University, op gezag van de rector magnificus, prof. dr. K. Sijtsma, in het openbaar te verdedigen ten overstaan van een door het college voor promoties aangewezen commissie in de Aula van de Universiteit
op dinsdag 21 april 2020 om 13.30 uur

door

Eline Frouke de Vries

geboren te Haarlem.

Promotiecommissie

Promotor
Prof. Dr. C.A. Baan

Copromotores
Dr. R. Heijink
Dr. J.N. Struijs

Overige leden
Prof. Dr. D.M.J. Delnoij
Prof. Dr. N.J.A. van Exel
Prof. Dr. P.T.T. Jeurissen
Prof. Dr. A. de Jonge
Prof. Dr. J.J. Polder

The research reported in this thesis was financially supported by the National Institute for Public Health and the Environment (Rijksinstituut voor Volksgezondheid en Milieu), Strategic Program RIVM (SPR) 2015-2019, grant number S/133002.

TABLE OF CONTENTS

SUMMARY

General introduction

Over the past decades, the ageing of the population with a changed demand for healthcare and advanced medical technology has increasingly pressured the sustainability of healthcare systems in Western countries. The demand has changed from patients with acute diseases requiring timely care from a single provider to persons with chronic diseases requiring highly coordinated care from providers over the entire continuum of care, cure and prevention. Yet, current health systems are fragmented and inefficient to meet the demand and are increasingly challenged to improve value.

Value is defined as the ratio between outputs (such as health outcomes) and inputs (such as spending) of health services. That implies that value can be increased by improving health outcomes while maintaining (or increasing) spending levels, for example by stimulating high-value services utilization such as care coordination pathways. Value can also be improved by maintaining (or improving) health outcomes while lowering spending levels, for example by reducing low-value service utilization such as the use of antibiotics in non-bacterial infections.

In an attempt to meet the changed demand, population health management (PHM) initiatives have emerged in a number of countries which aim to enhance value from a network perspective, for example in Germany (Gesundes Kinzigtal) and in the United States of America (Accountable Health Communities). Commonly, PHM initiatives aim to improve value by integration of services across care, cure and prevention within regional networks of providers, municipalities and insurers.

This thesis aims to contribute to the existing literature on PHM in two ways. First, in Part A, we explore how to measure value in a PHM context in order to make informed budget allocation decisions and monitor interventions across and within PHM regions. Second, in Part B, we gain insight into the different types of alternative payment models (APMs) and their effects on value and what experiences are with the implementation of APMs, in order to assess their role in PHM. This is important as many PHM initiatives are experimenting with payment reforms, but research on the effects of APMs on value and empirical research on the implementation of APMs is still lacking.

This thesis uses the context of PHM initiatives within the Dutch health care system. PHM initiatives emerged as a increased decentralization led to a larger role for regions and municipalities. Along with the developments that led to Dutch PHM initiatives, the Dutch government stimulated the integration of care for subpopulations including chronic diseases and episodes of care among which diabetes mellitus type 2 and

maternity care. Since Januari 2017 it is possible for maternity care providers organized in regional Maternity Care Networks (MCNs) to adopt bundled payment contracts. These bundled payment contracts are interventions that may be adopted in a PHM setting.

Part A: Measuring value in a population health management context

Part A explored how to measure the concept of value by reviewing the current state of low-value service indicators (chapter 2) and how to operationalize the concept of value within observational datasets (chapter 3 and chapter 4).

Low-value services are defined as services that provide no benefit to patients or can even cause harm. **Chapter 2** showed that the majority of low-value care indicators are in medical (primary or secondary) care (87 out of the 115 listed low-value care indicators). The remaining indicators were found in prevention (n=25) and in long-term care (n=3). No indicators were found in social care. Three indicators were assigned the highest level of evidence as they were underpinned by both guidelines and evidence from the literature. Other indicators were underpinned by clinical guidelines or Choosing Wisely recommendations. Despite the fact that several indicators are used in APMs, no information on the validity of the indicators was found in the literature.

Chapter 3 analyzed the drivers of regional variation in medical spending by looking at subgroups (i.e. individuals with diabetes and depression) in addition to the total population. Heterogeneity issues with regard to case-mix were aimed to overcome by using an extensive dataset (secondary health survey data linked with claims data, healthcare supply data and municipality registration data), in addition to the selection of subgroups. The results showed that PHM regions with above (or below) average spending for the general population mostly showed above (or below) average spending for diabetes and depression as well. Individual demand variables explained around 62% of the total variance. Less than 1% of the total variance was attributed to the regional level. Yet, the drivers of the variation at the regional level varied between subgroups. Demand factors explained nearly all variation across regions for depression but explained 88% of the variation for diabetes. The variation left unexplained (12% for diabetes) indicates differences across regions due to inefficiencies. This suggests that the extent to which regional variation in medical spending can be considered as inefficiency may differ between regions and subgroups.

In a first attempt to measure the value of maternity care, **chapter 4** described the variation of six proposed value-indicators for maternity care across Dutch MCNs. Specifically, we used the association between low-value services utilization (inputs) and

maternal and neonatal health outcomes (outputs) as measures of value across the MCNs. For example, the association of the rate of caesarean sections in low-risk pregnancies and the rate of Apgar score lower than 7 after 5 minutes at the regional (MCN) level. We found substantial variation across MCNs for the six value-indicators. The additional analyses showed that the inputs, i.e. the low-value care indicators, may have captured a part of the concept of value. However, despite the use of many case-mix variables, we could not rule out that these findings were due to population heterogeneity.

Part B: Alternative payment models in population health management
Part B gained insight into the current state of APMs using international literature (chapter 5) and experiences of PHM stakeholders in the Netherlands (chapter 6).

Chapter 5 reviewed what types of APMs have been implemented in maternity care in Western countries. Seventeen initiatives employed APMs in the United States (n=13), the United Kingdom (n=2), New Zealand (n=1) and in the Netherlands (n=1). Within these initiatives, pay-for-performance models (n=2), shared savings models (n=7) and bundled payment models (n=8) were found. Key design elements (such as eligible population, episode time span, care providers that participated in the model, care activities covered by the model, risk mitigation strategies) varied highly. Key terms describing the type of payment model (e.g. shared savings and bundled payments) were used interchangeably. APMs that shifted more financial accountability toward providers tended to include more strategies that mitigated financial risks. The first evaluations (n=4) on the effects of APMs are tentatively positive on different indicators of health and spending. Two studies found a positive association between the APM and health outcomes and two studies found a reduction in medical spending.

Chapter 6 investigated what types of payment models were implemented in the Dutch PHM regions and what barriers were experienced towards payment reform**.** After three years of PHM**,** shared savings models for pharmaceutical care (n=4) and extensions of the included services existing (bundled) payment models adding providers into the model (n=5) were adopted. The experienced barriers included information asymmetry between providers and insurers, worsening reputation of insurers, lack of trust as a result of failed reform attempts, misaligned incentives in the hospital settings, hesitation to accept financial accountability and lack of start-up funding, leadership and intrinsic motivation to reform. According to the interviewees, these experienced barriers were partly due to a lack of a sense of urgency.

General discussion

Finally, **chapter 7** discussed several issues that were encountered in the exploration to measure value and the effort to gain insight in the use and effects of APMs in a PHM context. One of the issues that were addressed was the need to develop comprehensive sets of value-indicators which capture the full continuum of cure, care and prevention. Currently, there are gaps in the availability of indicators of value. For example, low-value care indicators were mainly found in cure, while research has shown that low-value care is also present outside of cure and PHM initiatives increasingly develop interventions outside the cure sector. Another gap is the lack of indicators that reflect experiences that are reported by the population.

Another methodological issue is that operationalizing the spending part of the value-equation is problematic due to endogeneity. The problem is that spending encompasses (among others) the treatment and the complications that may be caused by the treatment itself. Moreover, we found that despite access to many case-mix variables, health spending may reflect the health status of the population rather than the quality of care at the regional level. We aimed to bypass this problem by using low-value care indicators as inputs instead of spending. This approach seemed to be able to capture a part of the value concept, but also raised new questions regarding the optimal amount of low-value care indicators to include, and other dimensions of value that are important (for instance high-value care). Therefore, in the future efforts to design a comprehensive set of value indicators, there also should also be attention for the development of high-value indicators and the development of guidance for deciding which and how many indicators should be included in which situation. Depending on the (level of the) question at hand the set of indicators may vary in number and contents.

Furthermore, progress to validly measure value is essential for the design and adoption of APMs. Only if value is measured validly, we are able to identify which key elements of APMs are essential in the progression towards value. We also discussed that adoption of APM is complicated by the lack of a common language for the APM typology. It is important to establish a common language, because APMs are complex and they operate in complex care settings; key elements vary highly among models and initiatives. In the meantime, APMs should be carefully designed to context at hand, while acknowledging the level of risk providers are able and willing to bear. In addition, it is important to be transparent about the stage of development of the APM, available evidence and have realistic expectations. Still many steps have to be taken to fully grasp the concept of value in a PHM context. This thesis hopes to have contributed to that goal.

SAMENVATTING

Algemene introductie

In de afgelopen decennia is de houdbaarheid van zorgstelsels in Westerse landen onder druk komen te staan. Als verklaring hiervoor worden doorgaans drie aspecten genoemd: de vergrijzing, een veranderde zorgvraag en de steeds grotere beschikbaarheid van geavanceerde medische technologie. De zorgvraag is verschoven van patiënten met een acute zorgvraag naar chronisch zieken met een vraag naar (complexe) geïntegreerde zorg. Bij geïntegreerde zorg staat de patiënt centraal en wordt samengewerkt tussen aanbieders over het gehele continuüm van zorg, welzijn en preventie. De huidige zorgstelsels in Westerse landen zijn echter gefragmenteerd en kunnen niet op een efficiënte manier tegemoetkomen aan de veranderingen in de zorgvraag. Daarnaast wordt een steeds groter deel van het Bruto Nationaal Product uitgegeven aan de zorg en verwacht men een verdere stijging van de zorgkosten in de toekomst. Westerse landen zijn daarom steeds meer op zoek naar manieren om de waarde *(value)* van zorgstelsels te verhogen.

Waarde is gedefinieerd als de ratio tussen wat we uitgeven aan de zorg (bijvoorbeeld in euro's) en wat de uitkomsten zijn van zorg (bijvoorbeeld gezondheid). De waarde kan verhoogd worden door uitkomsten te verbeteren en de uitgaven aan zorg gelijk te houden (of iets te verhogen). Dit kan bijvoorbeeld door het stimuleren van het gebruik van zorg die bijdraagt aan de gezondheid van de populatie (hoog-waarde zorg of *high-value care*), zoals de hielprikscreening bij pasgeboren baby's. De waarde kan ook verhoogd worden door de uitkomsten gelijk te houden en de zorguitgaven te verlagen. Dit kan bijvoorbeeld door het reduceren van het gebruik van zorg die niet bijdraagt aan gezondheid (laag-waarde zorg of *low-value care*), zoals het gebruik van antibiotica bij niet-bacteriële infecties.

Eén van de ontwikkelingen, om de veranderende zorgvraag en de houdbaarheid van het systeem het hoofd te bieden, is het initiëren van populatiemanagement (PM). PM-initiatieven streven naar het verhogen van waarde door, afgestemd op de behoefte van de populatie, zorg over de domeinen van zorg, welzijn en preventie te integreren. PM-initiatieven zijn regionale netwerken van aanbieders, gemeenten en zorgverzekeraars. Bekende voorbeelden zijn *Gesundes Kinzigtal* in Duitsland en de *Accountable Health Communities* in de Verenigde Staten.

Dit proefschrift draagt op twee manieren bij aan de huidige literatuur over PM. In Deel A is verkend hoe waarde gemeten kan worden in een PM-context. Het doel was om goed geïnformeerde beslissingen te kunnen nemen over nieuwe interventies voor specifieke subpopulaties en om de voortgang van de PM-initiatieven te kunnen meten. In Deel B is onderzocht wat de mogelijke rol van alternatieve bekostigingsmodellen

in PM kan zijn. Dat is belangrijk, omdat veel PM-initiatieven experimenteren met alternatieve bekostigingsmodellen, terwijl onderzoek naar de effecten van alternatieve bekostigingsmodellen op waarde nog ontbreekt. Daarnaast is er nog weinig bekend over de ervaringen met het implementeren van alternatieve bekostigingsmodellen.

Dit proefschrift is geschreven met behulp van data uit de literatuur (review) en twee landelijke studies: de landelijke monitor proeftuinen (2013-2018), waarin negen regio's zijn gevolgd in hun ontwikkeling naar PM en de monitor integrale bekostiging geboortezorg (2015-2020), waarin de ontwikkeling van integrale bekostigingscontracten wordt gevolgd en wat de effecten zijn van integrale bekostiging op kwaliteit, toegankelijkheid en betaalbaarheid van de geboortezorg.

Deel A: Het meten van waarde in populatiemanagement
In Deel A is gekeken naar de wijze waarop waarde gemeten kan worden.

Allereerst is in de literatuur (**hoofdstuk 2**) gekeken naar indicatoren voor laag-waarde zorg. Dat wil zeggen, welke indicatoren worden gebruikt voor het meten van het gebruik van zorg (of welzijn of preventie) die niet bijdraagt aan de gezondheid van de patiënt, en deze zelfs schade kan toebrengen. De systematische review laat zien dat het merendeel van de indicatoren voor laag-waarde zorg in het domein van zorg (eerste lijn of tweede lijn) vallen (87 van de 115 gevonden indicatoren). De overige indicatoren gaan over preventie (n=25) of langdurige zorg (n=3). Er zijn geen indicatoren voor laag-waarde zorg in het domein van welzijn gevonden. Daarnaast bleek dat slechts drie van de 115 indicatoren zijn ontwikkeld op basis van bewijs in de literatuur en in richtlijnen, en zijn daarom aangemerkt als 'goed ondersteund door bewijslast'. Andere indicatoren zijn ontwikkeld op basis van informatie uit richtlijnen en *Choosing Wisely*-aanbevelingen. Ondanks het feit dat meerdere indicatoren worden gebruikt in de zorgcontractering door zorgverzekeraars, vonden wij geen informatie over de validiteit van deze indicatoren in de wetenschappelijke literatuur.

In **hoofdstuk 3** is onderzocht welke factoren een rol spelen bij regionale variatie in zorguitgaven door niet alleen de totale populatie te analyseren, maar ook twee subpopulaties (diabetes en depressie). Er is hierbij onderscheid gemaakt tussen variatie op individueel niveau en variatie op regionaal niveau. Voor populatieheterogeniteit is rekening gehouden door te controleren voor een uitgebreide set aan variabelen waaronder leeftijd, geslacht, sociaal economische status, diagnose-kosten groepen, zelfgerapporteerde gezondheidsstatus en afstand tot verschillende zorgaanbieders. Hiertoe is een dataset gecreëerd waarbij bestaande registraties en monitors werden

gekoppeld (een vragenlijst over gezondheid (de Gezondheidsmonitor), declaratiegegevens van zorgverzekeraars (Vektis) en data over locaties van zorginstellingen en de Gemeentelijke Basisadministratie). De resultaten laten zien dat PM-regio's die voor de totale populatie meer (of minder) zorguitgaven hebben dan gemiddeld, ook voor de subgroepen (diabetes en depressie) meer (of minder) zorguitgaven hebben. Ongeveer 62% van de variatie op individueel niveau kon worden toegeschreven aan de zorgbehoefte, waarbij zelfgerapporteerde gezondheidsstatus de belangrijkste was (28%). Minder dan 1% van de totale variatie kon worden toegeschreven aan het regionale niveau. Echter, factoren die een rol spelen bij het verklaren van de variatie op het regionale niveau bleken te variëren tussen de subgroepen. Bij depressie konden factoren die te maken hebben met de zorgbehoefte nagenoeg alle variatie op regionaal niveau verklaren; bij diabetes is dat 88%. Dat betekent dat bij diabetes 12% van de variatie op regionaal niveau niet verklaard kon worden. Dit wijst op mogelijke verschillen in efficiency, zowel tussen regio's als tussen subgroepen.

Vervolgens is in **hoofdstuk 4** de waarde van geboortezorg onderzocht op basis van gegevens over de variatie tussen verloskundige samenwerkingsverbanden (VSV's) (uitgedrukt in zes indicatoren voor waarde). Er werd gekeken naar associaties tussen het gebruik van laag-waarde zorg en maternale en neonatale gezondheidsuitkomsten als mogelijke operationalisatie van waarde. Er is, onder andere, gekeken naar de associatie tussen de proportie keizersneden in laag-risico zwangerschappen en de proportie Apgar score lager dan zeven na vijf minuten op regionaal niveau (VSV). Voor alle zes de voorgestelde indicatoren voor waarde is substantiële variatie gevonden tussen VSV's. Additionele analyses wijzen erop dat de indicatoren die gebruikt werden om het gebruik van laag-waarde zorg te meten, daadwerkelijk (een deel van) het concept van waarde meten. Ondanks het gebruik van een veelheid aan variabelen om voor populatieverschillen te controleren, is het mogelijk dat onze resultaten beïnvloed zijn door verschillen die in werkelijkheid toe te schrijven zijn aan de populatie (in plaats van aan de VSV's).

Deel B: Alternatieve bekostigingsmodellen in populatiemanagement

In Deel B van dit proefschrift is een overzicht gegeven van typen en effecten van alternatieve bekostigingsmodellen en ervaringen met het implementeren van dergelijke modellen.

Eerst is in **hoofdstuk 5** geanalyseerd welke typen alternatieve bekostigingsmodellen in de geboortezorg tot op heden werden geïmplementeerd in Westerse landen. Er zijn zeventien initiatieven gevonden waarbij alternatieve bekostigingsmodellen werden

geïmplementeerd. Dertien van deze initiatieven zijn gevonden in de Verenigde Staten, twee in Engeland, één in Nieuw Zeeland en één in Nederland. Er zijn *pay-for-performance* modellen (n=2), *shared savings* modellen (n=7) en integrale bekostigingsmodellen (n=8) geïmplementeerd. Het ontwerp van belangrijke elementen in de modellen (zoals de in aanmerking komende populatie, tijdspanne, (type) aanbieders die deelnemen aan het model, zorgactiviteiten die binnen het model vallen, risicoverzachtende strategieën) varieerde tussen de initiatieven. Daarnaast is gevonden dat belangrijke termen die gebruikt werden om het type model te beschrijven, door elkaar zijn gebruikt (bijvoorbeeld *shared savings* en integrale bekostiging). Alternatieve bekostigingsmodellen met (intrinsiek) meer financieel risico voor aanbieders hadden ook meer strategieën om dit risico te verminderen. De eerste empirische evaluaties (n=4) over de effecten van alternatieve bekostigingsmodellen waren voorzichtig positief op verschillende indicatoren van gezondheid en zorguitgaven. Twee studies vonden een positief verband met gezondheidsuitkomsten en twee studies vonden een reductie van zorguitgaven.

In **hoofdstuk 6** is onderzocht welke typen alternatieve bekostigingsmodellen in de Nederlandse PM-regio's zijn geïmplementeerd en wat ervaren barrières waren om te komen tot bekostigingshervorming. Na drie jaar PM in Nederland zijn er *shared savings* modellen geïmplementeerd in de farmaceutische zorg (n=4) en zijn bestaande alternatieve bekostigingsmodellen uitgebreid met nieuwe (typen) zorg en aanbieders in het model (n=5). Er werden meerdere barrières om te komen tot bekostigingshervorming genoemd door de geïnterviewden. De meest relevante waren een informatieasymmetrie tussen aanbieders en zorgverzekeraars, verslechtering van de reputatie van zorgverzekeraars, gebrek aan vertrouwen door mislukte pogingen tot bekostigingshervorming, tegengestelde prikkels in de ziekenhuissetting, terughoudendheid met het accepteren van financieel risico en een gebrek aan de initiële investeringen tijdens de opstartfase, leiderschap en intrinsieke motivatie. Volgens de geïnterviewden zijn deze ervaren barrières deels het gevolg van een gebrek aan gevoel van urgentie.

Algemene discussie

Tot slot zijn in **hoofdstuk 7** een aantal aspecten besproken die aan het licht zijn gekomen bij de verkenning naar manieren om waarde te meten en in het verkrijgen van inzicht in alternatieve bekostigingsmodellen, beide in een PM-context.

Eén van die aspecten is dat het van belang is om een set van indicatoren voor waarde te ontwikkelen die het volledige continuüm van zorg, welzijn en preventie beslaat. Op dit moment zijn er namelijk hiaten in de beschikbaarheid van indicatoren die waarde kunnen meten. Zo werden indicatoren voor laag-waarde zorg voornamelijk gevonden in het

zorgdomein, terwijl er ook bewijs is dat laag-waarde zorg bestaat buiten het zorgdomein. Een ander voorbeeld is dat landelijke registratiedata alleen routinematig verzamelde gegevens van het primaire zorgproces bevatten en bijvoorbeeld geen patiëntervaringen.

Een tweede belangrijk aspect is dat het gebruik van zorguitgaven bij het meten van waarde kan leiden tot bias. Deze bias ontstaat doordat zorguitgaven een optelsom is van (onder andere) de behandelingen, maar ook van de complicaties van die behandelingen. Daarbij is (onder andere) in dit proefschrift gevonden dat, zelfs als we uitgebreid corrigeren voor populatieverschillen, de zorguitgaven op regionaal niveau de gezondheidstoestand van de populatie weergeeft, in plaats van kwaliteitsverschillen op regionaal niveau. Om dit probleem te omzeilen hebben we waarde geprobeerd te meten door indicatoren voor laag-waarde zorg te gebruiken in plaats van zorguitgaven. Ondanks dat met deze aanpak een deel van het concept van waarde gemeten kan worden, roept het ook nieuwe vragen op. Daarom is geadviseerd om bij de ontwikkeling van een set van indicatoren om waarde te kunnen meten ook te kijken naar de mogelijke bijdrage van indicatoren die hoog-waarde zorg meten. Daarnaast is het van belang om goed te bekijken welke indicatoren gebruikt kunnen worden om antwoord te geven op de voorliggende vraag. Afhankelijk van de precieze vraagstelling en het niveau van de vraag (op organisatieniveau, op regionaal niveau, op landelijk niveau), kan de inhoud en de omvang van de indicatorenset variëren.

Bij het ontwerp en de adoptie van alternatieve bekostigingsmodellen is het valide meten van waarde van groot belang. Alleen dan kunnen de prikkels van de bekostigingsmodellen zo vormgegeven worden dat ze ook daadwerkelijk tot meer waarde leiden, en kunnen de effecten van de alternatieve bekostigingsmodellen op waarde daadwerkelijk worden gemeten. Daarnaast is het belangrijk dat er eenduidigheid komt in het gebruik van taal om de typologie van bekostigingsmodellen te duiden. Een gemeenschappelijke taal kan de adoptie van alternatieve bekostigingsmodellen helpen te bevorderen, door het vergroten van begrip van de complexe modellen die opereren in een complex zorgsysteem. Verder is het verstandig om de alternatieve bekostigingsmodellen steeds af te stemmen naar de situatie in de praktijk, waarbij rekening wordt gehouden met het financiële risico dat zorgaanbieders willen en kunnen dragen. Daarbij is het belangrijk om transparant te zijn over de elementen in het model, wat het bewijs is voor de effecten van het model, en realistische verwachtingen te hebben over wat de uitkomsten kunnen zijn. Voordat het concept van waarde optimaal gebruikt kan worden in een PM-context moeten er nog veel stappen gezet worden. Dit proefschrift hoopt hieraan een bijdrage te hebben geleverd.

CHAPTER 1

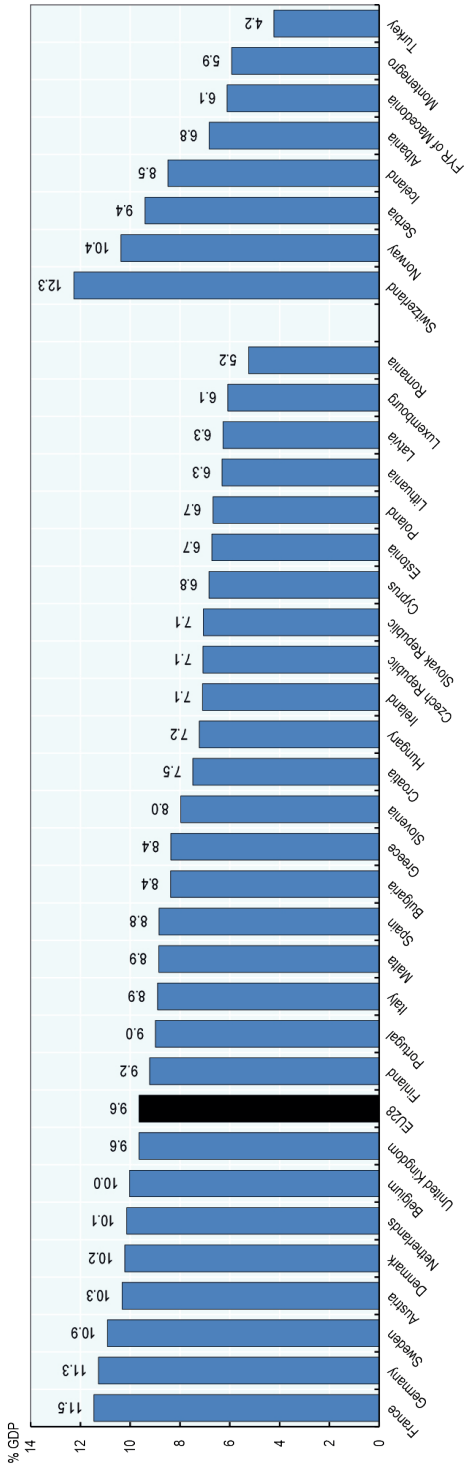
General introduction

Sustainability of health care systems in Western countries under pressure

Over the past decades, the sustainability of health care systems in Western countries is increasingly pressured. One of the main reasons was the reduction of the incidence of infectious diseases that resulted in an enormous increase in the average life expectancy [1]. In combination with problematic health factors such as obesity and smoking, this emerged in an era of chronic diseases after the 1950s. These chronic diseases, among which diabetes, chronic kidney disease, chronic vascular diseases, cancer and psychiatric disorders, are not only the main causes of death, but also have an enormous impact on the quality of life for an increasing amount of life years. Consequently, the demand for health care has changed from acute diseases requiring timely care from a single provider to chronic diseases requiring highly coordinated care from providers over the entire continuum of care, cure and prevention [2, 3]. Together with a decreasing birth rate, emerging technological innovations such as expensive innovative cancer therapies and high-tech diagnostic technologies, the changing demand for health care resulted in European spending levels of about 9.6% of the Gross Domestic Product (GDP) in 2017 (see Figure 1), with further growth expected in the future [4]. Therefore, countries are increasingly challenged to improve the value of their health systems by improving populations' health through meeting the changed demand for health care, while containing the spending levels [4, 5]. If these goals are not sufficiently achieved countries will need to increase taxes, shift resources from other public domains (e.g. infrastructure or education) or increase use of out-of-pocket payments [6].

Current health systems are fragmented and inefficient

Many health care systems are still largely organized to facilitate acute care delivery [2, 4]. Acute diseases, such as infectious diseases or fractured arms, require high-responsive timely care from single providers, meaning that care is organized, financed and delivered in specific and separate domains (e.g. primary care, specialty care, long-term care, social care, home care, pharmaceutical care; all separate). However, people with one or multiple chronic diseases need careful calibrated care from multiple providers that is guided by patient goals [2]. Both patients and providers struggle with the level of collaboration that is required for addressing their needs [7]. In daily practice, the current organization of care in sectors and domains require patients to explain their situation multiple times, to continuously communicate changes in treatment plans and to continuously keep each provider aligned to personal goals. Therefore, the fragmentation of care leads to high-pressure on the organizational capacity of patients. Additionally, the fragmentation of care may result in miscommunications that lead to medical errors or overutilization and payment models that incentivize production instead of value and high administrative burdens and other types of inefficiently organized, financed and



Source: OECD Health Statistics 2018; Eurostat Database; WHO Global Health Expenditure Database [4]

Figure 1: Health expenditures as a share of GDP, 2017 (or nearest year)

delivered care [4]. Recent estimates show that up to 20% of health spending in Europe [4, 5] and up to 30% of health spending in the United States is wasteful [8-10]. At the same time, there is growing evidence that low-value services (e.g. non-medically indicated caesarean sections or the use of antibiotics in viral infections) are increasingly performed (e.g. [11-15], while high-value services (e.g. addressing health-related social needs and targeting subpopulations that are being underserved) are underutilized [5]. Consequently, countries are increasingly searching for ways to improve the value of their health systems.

Population health management initiatives to achieve value

One of the responses to the sustainability problems health systems of Western countries are facing, is to improve health care delivery by integration of services across medical care, social care and public health within the region, which is called population health management (PHM) [16]. The ultimate goal for most PHM initiatives is to, simultaneously, attain better population health and experienced quality of care and a reduction of the per capita costs [9].

In a number of countries, networks of providers, payers and municipalities, within specific regions have adopted the PHM approach. Generally, PHM regions try to close the gap between health care and community services by addressing a number of themes. Common themes include building trust and increasing collaboration between all stakeholders and implementing interventions tailored for specific subgroups. Additionally, PHM regions try setting up data-infrastructure to share patient information between multiple care providers and make efforts to shift the financial accountability from payers towards groups of providers [17]. A well-known PHM initiative is Gesundes Kinzigtal (GK) in Germany [18, 19], which implemented a long-term shared savings contract to organize care across all domains for people of all ages and needs [20]. In the contract was specified that one organization (GK) is accountable for the care delivery and spending for the entire (insured) population [21]. GK has developed many interventions, which include evidence-based preventive programs that target common chronic diseases and supporting patients' self-management activities [20]. For example, if patients are identified to be at risk for a certain disease, individual treatment plans are developed by doctors and patients together. GK invests in training physicians in how to improve case management and shared-decision making. In addition, substantial investments were made to develop data-infrastructure to improve the collaboration between providers.

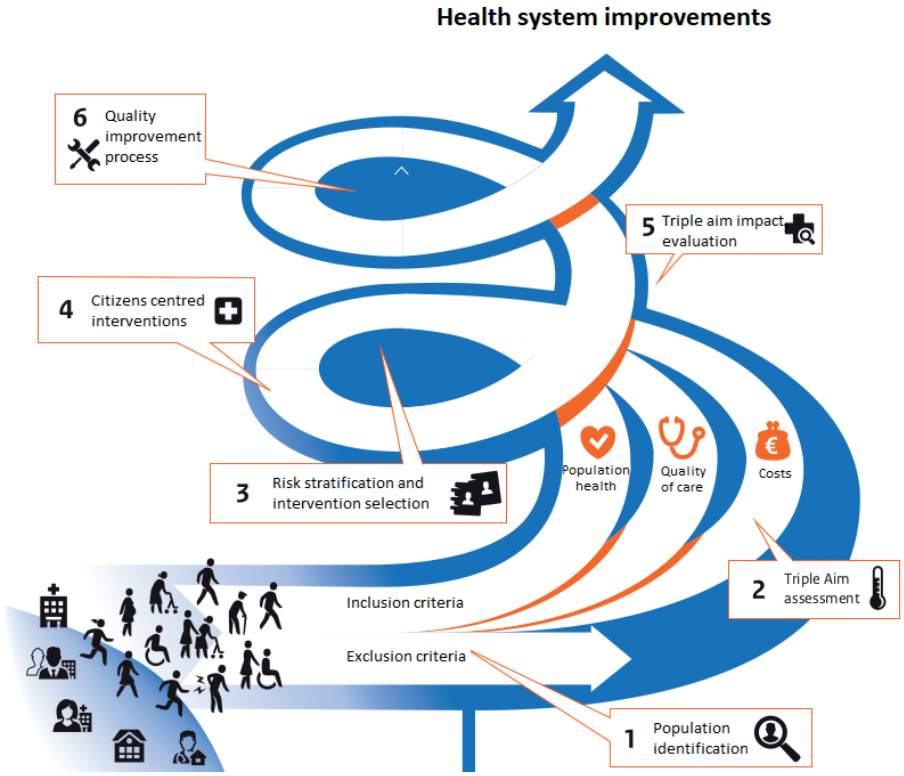
Other examples include Accountable Health Communities in the USA [22] and Dutch PHM initiatives [23]. These PHM initiatives employ interventions that, for instance,

aim to develop data-infrastructure for data sharing among providers. Other interventions include the substitution of low-complex hospital-based cardiovascular management care towards the primary care setting and implementing alternative payment models (e.g. shared savings in pharmaceutical care or bundled payments for maternity care) [22-28]. Although these PHM initiatives make efforts to experiment with payment reform, they struggle on how to successfully develop and implement alternative payment models [22, 23]. As such, Accountable Health Communities search how payment model incentives should be set to enhance social determinants, while improving financial sustainability [29] and the Dutch PHM initiatives are in search for the most appropriate payment model that is aligned with the PHM goals.

Research on population health management is starting to emerge

As research on PHM initiatives is just starting to emerge, evidence of the effects of PHM initiatives is scarce. Until thus far, two studies [19, 20] showed beneficial outcomes of Gesundes Kinzigtal; savings and a lower mortality rate compared to the control group. Additionally, patient and provider experiences were found to be positive in Gesundes Kinzigtal [30, 31]. Up until now, research in PHM has focused on defining the concept of PHM [16] and how to evaluate PHM initiatives [32]. Struijs and colleagues (2015) developed an analytical framework to evaluate PHM (Figure 2), that was based on the Care Continuum Alliance model [33]. The essence of this framework is that through a deep understanding of the needs of the population and continuously monitoring on PHM goals, resources may be allocated in such a way, that those needs are met in an efficient way. Based on Struijs' analytical framework, Hendrikx and colleagues worked on operationalizing PHM goals such as population health [34-36] and experienced quality of care [37]. Yet, the link between the interdependent PHM goals (i.e. population health, quality of care and spending) and the overarching goal of 'value' is currently lacking. This link is important to be able to monitor interventions and make informed budget allocation decisions in PHM regions. Therefore, the first aim of this thesis is to contribute to the existing PHM literature by exploring how to measure and operationalize 'value' in a PHM context.

Other research in PHM concentrated on how to successfully implement interventions, such as Steenkamer and colleagues [38] who proposed a set of guiding principles for how collaboration can be improved in pharmaceutical care, or whether existing prediction models are helpful to identify the needs of the population to optimally target interventions [39]. In addition, a few PHM interventions were studied, as for example the substitution of care from hospital to primary care [40, 41]. Although many PHM initiatives are experimenting with payment reforms, research on the implementation and



Source: Struijs et al 2015 [32]

Figure 2: Analytical framework for population health management

the effects of these models on value in PHM is still lacking [42]. Therefore, the second aim of this thesis is to gain insight into experiences with, and the types and effects of, alternative payment models in PHM.

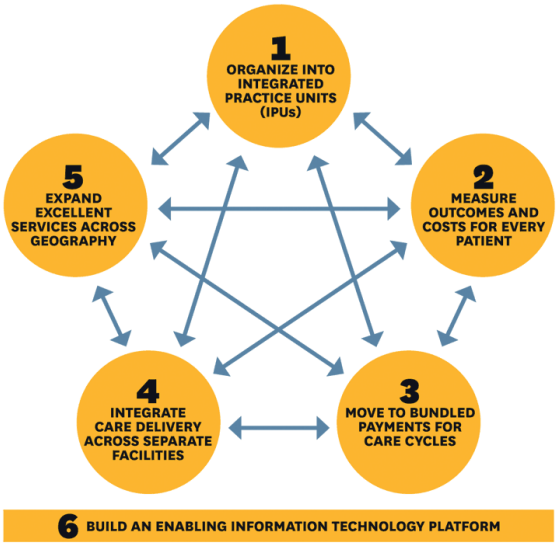
The following section elaborates on the background of the two thesis aims.

Theoretical considerations

Since the introduction of Porters' book 'Redefining health care' in 2006 [43] and the development of concepts such as 'value-based payment models' (or as is further referred to in this thesis: alternative payment models (APMs)), the concept of 'value' gained popularity, in particular in the health care domain. Porters' definition of value is health outcomes that matter to patients relative to its spending, where outcomes are patient reported and disease-specific and spending comprises the total spending for the full cycle of care for the disease [44]. Based on this concept, the Value-Based Health Care

model (Figure 3) was developed, and consists of six steps to transform care pathways such that the ratio between patients' health and euro's spent, increases [45]. The model was applied in many initiatives (mainly in hospitals) and can be compared to PHM in the sense that it aims to integrate services across (groups of) providers by using aligned and supportive payment models and data-infrastructure in order to improve overall value, but particularly designed for the health care domain.

Conceptually, it is important to acknowledge the difference between *value* - which seeks to optimally increase the ratio between inputs (i.e. spending, or other inputs such as utilization or labor) and outputs (i.e. health outcomes that matter to the patient), and *cost containment*, which only limits the inputs with no regard to changes in the output [46]. This implies that value is, partly, about reducing inefficiencies, such as reducing administrative excess, inefficient and duplicate use of resources and low-value care [5, 47]. *Low-value care* is generally defined as health services that provide no benefit or may even cause harm [5, 13]. Low-value services can be found at the entire care continuum of care [48, 49] and are, per definition, services for which the costs outweigh the benefits [50]. For example, overtesting with overdiagnosis as a result may seem harmless but have undesired effects when diagnoses causes stress among patients and family members and may even lead to more invasive procedures or medication, that, in turn, may have serious



Source: Porter and Lee 2013 [45]

Figure 3: The value agenda

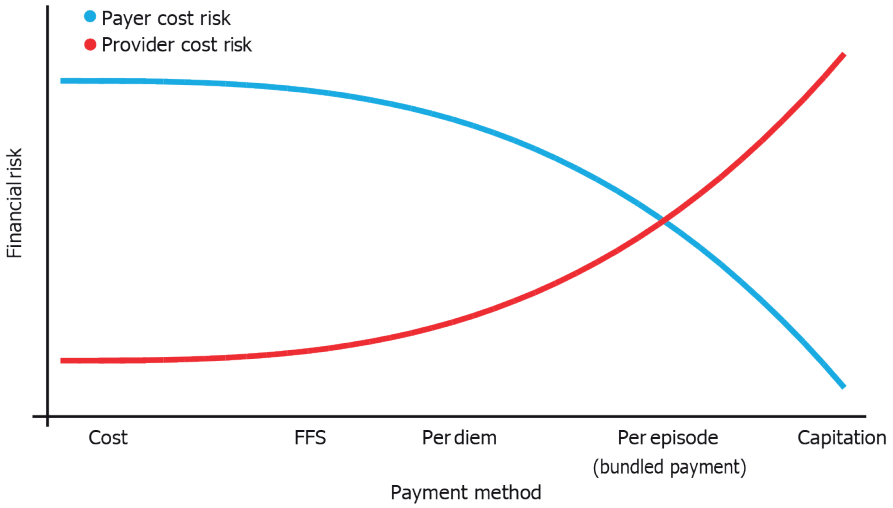
adverse outcomes or adverse effects [50]). Note that whether services are of high- or low-value may differ between groups of patients [51], e.g., the use of propranolol is of low-value for patients using it for anxiety issues, but is of high-value in the management of high blood pressure.

As is discussed in previous sections, PHM initiatives employ interventions that introduce mechanisms to reduce the utilization of low-value services and stimulate the utilization of high-value services. Examples of such interventions are substituting low-complex care to the primary care setting and implementing APMs. Commonly, PHM initiatives search to improve value of care, both at the level of individuals and the system. Therefore, it is important to move beyond the disease-specific definition from Porter. This thesis defines value as the ratio between outputs and inputs of health services specifically at the regional level and irrespective of domain (e.g. medical care, social care, prevention).

Alternative payment models in population health management

Theoretically, a precondition for improving value is to shift away from the traditional fragmented fee-for-service (FFS) payment models to more value-based payment models [2, 52, 53]. That is because FFS models are known to incentivize each provider to increase the amount of services produced (as long as price is above marginal cost) and they are designed for acute care specifically [54]. Therefore, FFS models seem to be misaligned with the PHM strategy that aims to integrate services over medical care, social care and prevention [55]. APMs such as shared savings arrangements, bundled payments or global payments may fit PHM better.

The promise is that APMs incentivize care coordination, stimulate high-value services utilization and discourage low-value services utilization through increased financial accountability for (groups of) providers [56]. Figure 4 shows the theoretical effects of APMs on the allocation of financial risks between provider and payer that have been discussed extensively in the literature [57-60]. In FFS, the financial risks are largely born by the payer. Because providers do not run financial risks in terms of volume and value of care, FFS incentivizes providers to increase volume, even when the services are of low value [52, 57-59]. APMs aim to shift a part of the financial risks for value (i.e. health outcomes and spending) towards providers [57]. The increased financial risk for providers incentivizes them to increase efficiency by avoiding low-value services utilization and overutilization [61], in addition to downward substitution of care, task reallocation and increased coordination of care. Also, APMs encourage the utilization of services for which the benefits exceed the costs (high-value services) [56]. The allocation of financial risk between provider and payer is determined by the scope and type of



Source: Frakt and Mayes 2012

Figure 4: (alternative) payment methods and provider/payer risk

APM. Ideally, only performance risk, which is the risk related to what the provider is able to influence [52], is allocated with the provider, and insurance risk, which is the risk related to patient case-mix [52], is allocated with the payer. The optimal allocation of risk is where provider risk is maximized and insurance risk is minimized for providers [60]. In search for the optimum allocation of risks, APMs may be augmented with bonuses and penalties for meeting certain (quality) targets (for example in shared savings) and strategies that mitigate the level of risk for providers (e.g. high-risk population exclusions, risk-adjustments, or a stop-loss provision, which is a threshold that caps the maximum for which the provider is at risk [62]).

In practice, many PHM initiatives struggle on how to successfully design and implement APMs [17, 22, 23] both at the intervention level and at the PHM region level.

Thesis objectives

This thesis aims to contribute to the existing literature on PHM in two ways. First, it explores how to measure the concept of value in PHM for monitoring and allocation decision-making purposes in PHM regions (Part A). Specifically, this part explores how to measure the concept of value by reviewing the potential of low-value care indicators in PHM and how to operationalize value within observational datasets. The second part (Part B) aims to gain insight into experiences with, and the types and effects of, APMs

in order to assess its possible role in PHM. Part B questions what types of APMs are currently implemented and what evidence is available on spending and health outcomes. It also investigates experiences with the implementation of APMs in PHM. To answer these questions, this thesis uses several types of data sources: international literature, nationwide datasets containing information from multiple data sources and interviews with various stakeholders from PHM regions.

Thesis context

This thesis uses the context of PHM initiatives within the Dutch health care system, which has a Bismarckian history of social health insurance. In 2006, managed competition was introduced in which the government adopted a more distant role as supervisor and facilitator of the health care markets (i.e. between patients and providers, between patients and insurers and between insurers and providers) [63]. In this context, the trend of increasing decentralization of medical care and social care had created a larger role for regions and municipalities. In 2013, the Dutch Ministry of Health, Welfare and Sport, designated nine regional partnerships to be monitored by the Dutch National Institute for Public Health and the Environment (RIVM) [23]. In the period 2013 through 2018, the RIVM monitored these regions using both qualitative and quantitative research approaches in the National Monitor Population management. This thesis uses interviews that were held periodically with both payers (insurers and municipalities) and providers (among which GPs, care groups, hospitals). In addition, this thesis uses existing nationwide linked dataset (claims data, health data and municipality registration data) that were acquired for the purpose of the National Monitor Population management. In the Netherlands, 19 PHM regions have been identified [64].

Along with the developments that led to Dutch PHM initiatives being emerged, the Dutch government actively stimulated the integration of care for subpopulations including both episodes of care and chronic disease. For instance, since January 2017 it is possible for maternity care providers to implement bundled payment contracts [27, 28, 65]. The adoption of these bundled payment contracts are interventions for subpopulations that are suitable (and may be adopted) in a PHM setting. The RIVM was asked to monitor and evaluate the implementation of the bundled payment model in maternity care. This thesis uses the individual level nationwide dataset that linked claims data with maternal and infant health outcomes and quality of care and multiple datasets containing case-mix.

Thesis outline

The remainder of this thesis is outlined as follows. Part A explores how to measure the concept of value in PHM using scientific literature and extensive nationwide observational data. Chapter 2 assesses whether low-value care indicators are suitable for use in PHM by reviewing the presence and the validity of indicators in the scientific literature that measure low-value care across the entire continuum of care in the international literature. Chapter 3 investigates how to relieve population heterogeneity in its effort to identify areas on where to improve value, by assessing differences across subgroups across Dutch PHM regions. The study uses a nationwide linked dataset from the National Monitor Population management. Chapter 4 explores how to measure value by using the association between the utilization of low-value care and maternal and neonatal health outcomes across Dutch Maternity Care Networks using nationwide observational data at the individual level that combined multiple data sources from the Monitor Bundled Payments for Maternity Care.

Part B gains in-depth insight into the current state of payment reform across PHM regions using international literature and the experiences of actual stakeholders. Chapter 5 reviews what types and key elements of APMs are currently implemented in one specific PHM intervention, maternity care, and what their effects are on health outcomes and spending. Chapter 6 identifies which APMs are currently implemented in Dutch PHM regions and what experiences are with implementing the APMs.

Finally, the main findings are discussed and reflected on in chapter 7.

REFERENCES

1. Van der Velden, J. & Mackenback, J.P., *Gezondheid en Ziekte in de Wereld,* in Mackenbach, J.P. & Stronks, K. *Volksgezondheid en Gezondheidszorg*, 2012, Amsterdam: Reed Business.
2. Nolte, E., Knai, C., McKee, M., *Managing chronic conditions: experience in eight countries*, in WHO. *European Observatory on Health Systems and Policies Series*, 2008.
3. Nolte, E., McKee, M., *Caring for people with chronic conditions: a health systems perspective*, in *European Observatory on Health Systems and Policies Series*, 2008.
4. Organisation for Economic Co-operation and Development (OECD) / European Union (EU), *Health at a Glance: Europe 2018: State of Health in the EU Cycle,* 2018, OECD Publising: Paris/EU, Brussels.
5. Organisation for Economic Co-operation and Development (OECD), *Tackling Wasteful Spending on Health*. 2017, OECD Publising: Paris.
6. Expert Group on Health System Performance Assessment. *Tools and methodologies to assess the efficiency of health care services in Europe: An overview of current approaches and opportunities for improvement*, European Commission, 2019, Publications Office of the European Union: Luxembourg.
7. Organisation for Economic Co-operation and Development (OECD), *Fiscal Sustainability of Health Systems: Bridging Health and Finance Perspectives*. 2015, OECD: Paris.
8. Berwick, D.M., Hackbarth, A.D., *Eliminating Waste in US Health Care.* JAMA, 2012. 307(14): p. 1513-1519.
9. Berwick, D.M., Nolan, T.W., and Whittington, J., *The triple aim: care, health, and cost.* Health Aff (Millwood), 2008. 27(3): p. 759-69.
10. Shrank, W.H., Rogstad, T.L., and Parekh, N., *Waste in the US Health Care System: Estimated Costs and Potential for Savings.* JAMA, 2019. 322(15): p. 1501-1509.
11. Schwartz, A.L., Landon, B.E., Elshaug, A.G., Chernew, M.E., McWilliams, J.M., *Measuring Low-Value Care in Medicare.* JAMA Intern Med, 2014. 174(7): p. 1067-1076.
12. Brownlee, S., et al., *Evidence for overuse of medical services around the world.* Lancet, 2017. 390(10090): p. 156-168.
13. Bhatia, R.S., et al., *Measuring the effect of Choosing Wisely: an integrated framework to assess campaign impact on low-value care.* BMJ Quality & Safety, 2015. 24(8): 523-31.
14. Elshaug, A.G., Moss, J.R., Littlejohns, P., Karnon, J., Merlin, T.L., Hiller, J.E., *Identifying existing health care services that do not provide value for money.* Med J Aust, 2009. 190(5): p. 269-273.
15. Elshaug, A.G., Watt, A.M., Mundy, L., Willis, C.D., *Over 150 potentially low-value health care practices: an Australian study.* Med J Aust, 2012. 197(10): p. 556-560.
16. Steenkamer, B.M., et al., *Defining Population Health Management: A Scoping Review of the Literature.* Popul Health Manag, 2017. 20(1): p. 74-85.
17. Struijs, J.N., Drewes H.W., and Stein K.V., *Beyond integrated care: challenges on the way towards population health management.* Int J Integr Care, 2015, 15: e043.
18. Hildebrandt, H., Schulte, T., and Stunder, B., *Triple Aim in Kinzigtal, Germancy: improving population health, integrating health care and reducing costs of care- lessons for the UK?* Journal of Integrated Care, 2013. 20(4): p. 205-222.
19. Pimperl, A., et al., *Evaluating the Impact of an Accountable Care Organization on Population Health: The Quasi-Experimental Design of the German Gesundes Kinzigtal.* Popul Health Manag, 2017. 20(3): p. 239-248.
20. Busse, R. and J. Stahl, *Integrated care experiences and outcomes in Germany, the Netherlands, and England.* Health Aff (Millwood), 2014. 33(9): p. 1549-58.
21. Hendrikx, R.J., et al., *Which Triple Aim related measures are being used to evaluate population management initiatives? An international comparative analysis.* Health Policy, 2016. 120(5): p. 471-85.

22. Alley, D.E., et al., *Accountable Health Communities--Addressing Social Needs through Medicare and Medicaid.* N Engl J Med, 2016. 374(1): p. 8-11.

23. Drewes, H.W., Struijs, J.N., and Baan, C.A., *How the Netherlands Is Integrating Health and Community Services.* NEJM Catalyst, 2016.

24. Drewes, H.W., et al., *National Monitor Populationmanagament. Part 1: first description of the pioneer sites.* (in Dutch) 2014, National Institute for Public Health and the Environment (RIVM): Bilthoven.

25. Drewes, H.W., Heijink, R., Struijs, J.N., Baan, C.A., *Working together towards sustainable care. National Monitor Population Management* (in Dutch). 2015, National Institute for Public Health and the Environment (RIVM): Bilthoven.

26. Drewes, H.W., et al., *Regions moving towards sustainable health systems: National Monitor Populationmanagement – Reflection on five years of pioneering* (in Dutch). 2018, National Institute of Public Health and the Environment (RIVM): Bilthoven.

27. Struijs, J.N., De Vries, E.F., Van Dorst, H.D.C.A., Over, E.A.B., Baan, C.A. *Insight in outcomes, utilization and medical spending of maternity care and first experiences with bundled payments* (in Dutch). 2018, Bilthoven: The Dutch National Institute for Public Health and the Environment (RIVM).

28. Struijs, J.N., De Bruin-Kooistra, M., Heijink, R., Baan, C.A. *Towards bundled payments for maternity care* (in Dutch). 2016. Bilthoven: The Dutch National Institute for Public Health and the Environment (RIVM).

29. Gottlieb, L., et al., *Evaluating the Accountable Health Communities Demonstration Project.* J Gen Intern Med, 2017. 32(3): p. 345-349.

30. Hildebrandt, H., et al., *Gesundes Kinzigtal Integrated Care: improving population health by a shared health gain approach and a shared savings contract.* Int J Integr Care, 2010. 10: p. e046.

31. Hildebrandt, H., Schmitt, G., Roth, M., Stunder, B., *Integrierte regionale Versorgung in der Praxis: Ein Werkstattbericht aus dem „Gesunden Kinzigtal".* Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen, 2011. 105(8): p. 585-589.

32. Struijs, J.N., et al., *How to evaluate population management? Transforming the Care Continuum Alliance population health guide toward a broadly applicable analytical framework.* Health Policy, 2015. 119(4): p. 522-529.

33. Care Continuum Alliance (CCA), *Outcomes Guidelines Report*. 2010.

34. Hendrikx, R.J.P., et al., *Comparing the Health of Populations: Methods to Evaluate and Tailor Population Management Initiatives in the Netherlands.* Popul Health Manag, 2018. 21(5): p. 422-427.

35. Hendrikx, R.J.P., et al., *Measuring Population Health from a Broader Perspective: Assessing the My Quality of Life Questionnaire.* Int J Integr Care, 2019. 19(2): p. 7.

36. Hendrikx, R.J.P., et al., *How to Measure Population Health: An Exploration Toward an Integration of Valid and Reliable Instruments.* Popul Health Manag, 2018. 21(4): p. 323-330.

37. Hendrikx, R.J.P., et al., *Harvesting the wisdom of the crowd: using online ratings to explore care experiences in regions.* BMC Health Serv Res, 2018. 18(1): p. 801.

38. Steenkamer, B., et al., *Population health management guiding principles to stimulate collaboration and improve pharmaceutical care.* J Health Organ Manag, 2018. 32(2): p. 224-245.

39. Elissen, A.M., et al., *Estimating community health needs against a Triple Aim background: What can we learn from current predictive risk models?* Health Policy, 2015. 119(5): p. 672-9.

40. Quanjel, T.C.C., et al., *Evaluating a Dutch cardiology primary care plus intervention on the Triple Aim outcomes: study design of a practice-based quantitative and qualitative research.* BMC Health Serv Res, 2017. 17(1): p. 628.

41. Quanjel, T.C.C., et al., *Shifting hospital care to primary care: An evaluation of cardiology care in a primary care setting in the Netherlands.* BMC Fam Pract, 2018. 19(1): p. 55.

42. Struijs, J.N., *Payment reform and integrated care: the need for evaluation.* Int J Integr Care, 2013. 13: p. e056.

43. Porter, M.E., Teisberg, E.A., *Redefining Health Care: creating value-based competition on results*, ed. Harvard Business Review Press. 2006.

44. Porter, M.E., *What Is Value in Health Care?* NEJM, 2010. 363(26): p. 2477-2481.

45. Porter, M.E., Lee, T.H., *The strategy that will fix health care.* 2013, Harvard Business Review: p. 50-70.

46. Palmer, S., Torgerson, D.J., *Economic notes: definitions of efficiency.* BMJ, 1999. 318(1136).

47. Bentley, T.G., Effros, R.M., Palar, K., & Keeler, E.B., *Waste in the U.S. Health Care System: A Conceptual Framework.* Millbank Q, 2008. 86(4): p. 629-659.

48. Baker, D.W., et al., *Design and Use of Performance Measures to Decrease Low-Value Services and Achieve Cost-Conscious Care.* Ann Intern Med, 2013. 158(1): p. 55-59.

49. The King's Fund, *Better value in the NHS: The role of changes in clinical practice.* 2015, The King's Fund: London.

50. Carroll, A.E., *The High Costs of Unnecessary Care.* JAMA, 2017. 318(18): p. 1748-1749.

51. Chernew, M.E., Rosen, A.B., and Fendrick, A.M., *Value-based insurance design.* Health Aff (Millwood), 2007. 26(2): p. w195-203.

52. Miller, H.D., *From volume to value: better ways to pay for health care.* Health Aff (Millwood), 2009. 28(5): p. 1418-1428.

53. Huerta, T.R., J.L. Hefner, and McAlearney, A.S., *Payment models to support population health management.* Adv Health Care Manag, 2014. 16: p. 177-83.

54. Busse R, Mays, N.B, *Paying for chronic disease care*, in *Caring for people with chronic conditions: a health system perspective.* 2008, European Observatory on Health Systems and Policies.

55. Tsiachristas, A., *Financial Incentives to Stimulate Integration of Care.* Int J Integr Care, 2016. 16(4): p. 8.

56. Baicker, K. and Chernew, M.E., *Alternative Alternative Payment Models.* JAMA Intern Med, 2017. 177(2): p. 222-223.

57. Frakt, A.B. and Mayes, R., *Beyond capitation: how new payment experiments seek to find the 'sweet spot' in amount of risk providers and payers bear.* Health Aff (Millwood), 2012. 31(9): p. 1951-8.

58. Organisation for Economic Co-operation and Development (OECD), *Better Ways to Pay for Health Care.* 2016, OECD Publising: Paris.

59. Robinson, J.C., *Theory and Practice in the Design of Physician Payment Incentives.* Mill Q, 2001. 79(2): p. 28.

60. Ruwaard, S., *Purchasing healthcare; Beyond setting the financial incentives right.* 2018, Tilburg University: Tilburg.

61. Mechanic, R.E. and Altman S.H., *Payment reform options: episode payment is a good place to start.* Health Aff (Millwood), 2009. 28(2): p. w262-71.

62. Porter, M. and Kaplan R., *How should we pay for healthcare?* 2015, Harvard Business School.

63. Kroneman, M., et al., *Netherlands: Health System Review.* Health Syst Transit, 2016. 18(2): p. 1-240.

64. Lemmens, L.C., et al., *Een populatiegerichte aanpak voor verbinding van preventie, zorg en welzijn: de beweging in beeld.* Ned Tijdschr Geneeskd, 2017(161: D849).

65. Struijs, J.N. and Hargreaves D.S., *Turning a crisis into a policy opportunity: lessons learned so far and next steps in the Dutch early years strategy.* Lancet Child Adolesc Health, 2019. 3(2): p. 66-68.


A

PART A

Measuring value in a population health management context


2

CHAPTER 2

Are low-value care measures up to the task? A systematic review of the literature

Published as
De Vries EF, Struijs JN, Heijink R, Hendrikx, RJP, Baan CA
Are low-value care measures up to the task? A systematic review of the literature
BMC Health Serv Res. 2016, 16:405

ABSTRACT

Background*:* Reducing low-value care is a core component of healthcare reforms in many Western countries. A comprehensive and sound set of low-value care measures is needed in order to monitor low-value care use in general and in provider-payer contracts. Our objective was **t**o review the scientific literature on low-value care measurement, aiming to assess the scope and quality of current measures.

Methods*:* A systematic review was performed for the period 2010 - 2015. We assessed the scope of low-value care recommendations and measures by categorizing them according to the Classification of Health Care Functions. Additionally, we assessed the quality of the measures by 1) analysing their development process and the level of evidence underlying the measures, and 2) analysing the evidence regarding the validity of a selected subset of the measures.

Results*:* Our search yielded 292 potentially relevant articles. After screening, we selected 23 articles eligible for review. We obtained 115 low-value care measures, of which 87 were concentrated in the cure sector, 25 in prevention and 3 in long-term care. No measures were found in rehabilitative care and health promotion. We found 62 measures from articles that translated low-value care recommendations into measures, while 53 measures were previously developed by institutions as the National Quality Forum. Three measures were assigned the highest level of evidence, as they were underpinned by both guidelines and literature evidence**.** Our search yielded no information on coding/ criterion validity and construct validity for the included measures. Despite this, most measures were already used in practice.

Conclusion*:* This systematic review provides insight into the current state of low-value care measures. It shows that more attention is needed for the evidential underpinning and quality of these measures. Clear information about the level of evidence and validity helps to identify measures that truly represent low-value care and are sufficiently qualified to fulfil their aims through quality monitoring and in innovative payer-provider contracts. This will contribute to creating and maintaining the support of providers, payers, policy makers and citizens, who are all aiming to improve value in health care.

BACKGROUND

The concept of low-value care, defined as services that provide no benefit to patients or can even cause harm [1, 2], has received much attention in recent years in Western countries. Reducing the use of low-value care is expected to contribute to cost containment and more efficiency in health care [3-5]. It leads to a reduction in medical spending without harming health outcomes and it may stimulate a reallocation of resources to high-value services [4]. In this way, measuring low-value care for which the non-effectiveness is proven provides information on a specific type of inefficiency, i.e. spending with no benefit, which can be used besides other, more indirect, types of efficiency analysis such as traditional cost-effectiveness studies or analyses of practice variation.

Internationally, several initiatives have been launched to reduce low-value service utilization, among which the Choosing Wisely (CW) campaign in the US. Similar initiatives have originated in 12 other countries including the United Kingdom, Canada, Australia and the Netherlands [4, 6]. In the CW campaign, participating specialty societies produce lists of recommendations that are to be discussed in the doctor's office, as for example, 'don't order diagnostic tests at regular intervals (such as every day), but rather in response to specific clinical questions' [7]. Ideally, these lists of recommendations would meet the CW criteria: 1) each of the services is within the specialty's purview, 2) each of the services is frequently used or costly, 3) each recommendation is based on sufficient evidence, and 4) the process for developing the recommendation list is documented and is made available to the public if requested [8]. In general, the recommendations aim to increase awareness among both doctors and patients [5] and subsequently influence the decision whether or not to use a specific service.

Besides these rather generic *recommendations*, studies have tried to assess the prevalence and geographic or practice variation in low-value care utilization (e.g. [9-12]) using direct *measures* of low-value care. The aim of the direct measures differs from the aim of recommendations. Where recommendations aim to create awareness among physicians and patients, low-value care measures may be widely used, for example in payer-provider contracts [13, 14] and for monitoring low-value care initiatives [4, 15].

To meet these aims, low-value care measures need to be methodologically sound [2, 16, 17]. Otherwise, using these measures might create misinterpretation, underuse of indicated services, patient selection or damage the patient-physician relationship [18]. To date, only one study [19] reviewed the state of low-value care measurement by

performing a scan of the published and grey literature. They found 37 specified measures and 123 services that may be developed into measures, covering mainly diagnostic or therapeutic areas. Furthermore, another study [20] identified a set of low-value services and demonstrated significant variance in its utilization between hospital referral regions in the US.

Still, major knowledge gaps exist in the literature on measuring low-value care. First, there is lack of knowledge regarding the validity of current low-value care measures [2, 16, 17]. As Baker et al. [15] pointed out earlier, low-value care measures must at least be rigorously evidence-based. In addition, they must be able to detect variation between providers, regions or countries, reflect actual cases of the concept of interest, be supported by correlations to other measures indicating the same concept, and not be subject to substantive systemic bias (i.e. importance, coding or criterion validity, construct validity and risk adjustment) [21]. Therefore, specific standards for how to develop and assess low-value care measures should be developed [15, 18]. Second, it is unclear whether current low-value care measures cover the whole continuum of care. This is important, because it was argued that low-value care use is present in all sectors along the care continuum [15, 22]. However, the low-value service recommendations from the CW initiative cover mainly specialist care in the cure sector [8].

In this study, we aimed to start filling these gaps by performing a systematic review of the recent scientific literature on low value care measurement. Our objective was twofold. Firstly, to assess the scope of low-value care *recommendations* and *measures* in the literature by categorizing them according to health care function (such as curative care, long-term care and rehabilitation). Secondly, to assess the quality of the *measures* by 1) analysing their development process and the evidence that underlies the measures and 2) analysing the evidence regarding the validity of a selection of the included measures.

METHODS

Study design and search strategy

A systematic review of the literature was performed, focusing on English-language articles published between January 2010 and January 2015. As recommended by Cochrane [23], we performed our search in multiple databases including EMBASE, Medline, SciSearch, BIOSIS Previews and GLOBAL Health. We developed a search strategy to identify articles matching a variation of the following search terms: 1) initiatives, design, measuring, indicators, instrument, identifying, index; 2) waste,

overuse, overutilization, misuse, low-value; and 3) health care, cure, care, prevention. Additional file 1 gives a detailed description of the search strategy.

Article selection

Two researchers (EFdV & RJPH) independently reviewed the relevance of the articles by screening titles and abstracts. As recommended by Cochrane [23], we included articles from peer-reviewed journals only. The full-text was retrieved when both researchers considered the paper relevant. Articles were eligible for review when they met the following predefined criteria: 1) the low-value service recommendation or measure in the paper matched the definition 'services that provide no benefit to patients or may even cause harm [1, 2]; 2) the low-value service recommendation or measure was described using clinical details such as diagnosis, patient population and treatment. We removed duplicate articles and replies or commentaries and theoretical or discussion articles that did not present any low-value service recommendations or measures. Any disagreement between the reviewers was resolved by discussion and consensus.

Data extraction

We extracted general characteristics of the articles (i.e. name of first author, year of publication, country, aim of the paper, methods) and the measures (i.e. the name of the measure, the numerator, the denominator, exclusion criteria and direction). In addition, we retrieved the original source or reference of the measure.

Recommendations versus measures

The literature search yielded both recommendations and measures for low-value care. We considered a description of low-value care as 'measure' when at least a numerator and denominator were specified as such. We identified the scope of both recommendations and measures, while the quality assessment was performed for the measures only.

Categorizing low-value care recommendations and measures by function in health care

All recommendations and measures were categorized using the Classification of Health Care Functions (ICHA-HC) as defined by the Organization for Economic Co-operation and Development (OECD), the World Health Organization (WHO) and Eurostat [24]. The ICHA-HC provides a framework to classify services according to their purpose or function and is commonly used to compare medical services internationally. It covers the entire continuum of the health system, i.e. curative care, rehabilitative care, long-term care and preventive care. We subcategorized curative care into general (i.e. primary) care and specialized care. General care involves basic care such as routine examinations,

basic maternity care, routine diagnosis and follow-up, prescriptions and vaccinations (unless they are covered under a preventive program) [24]. Specialized care involves more complex technology and is often a breakdown from the basic fields (e.g. neurosurgery or allergology) [24]. In addition, the measures were categorized according to the non-functional categories ancillary services (i.e. laboratory, imaging, transport), and medical goods (i.e. pharmaceutical and therapeutic appliances).

Assessing the quality of low-value care measures

We assessed the quality of the measures by 1) analysing their development process and the level of evidence underlying the measures, and 2) analyse the validity of a selection of the measures.

Development process and level of evidence

We distinguished two groups: A) articles that translated low-value service recommendations into low-value care measures, and B) articles that used measures previously developed by institutions. For both groups we reviewed how the measures were developed.

For group A, we searched for evidence underlying the recommendations. We categorized each measure based on the evidence, distinguishing three levels of evidence: 1) a combination of evidence from the literature (trial or review), guidelines and from CW, United States Preventive Services Task Force (USPSTF) or National Institute of Clinical Excellence (NICE) recommendations, 2) evidence from the literature (trial or review) or guidelines, and 3) evidence not found. As criteria for developing CW recommendations do not prescribe the level of evidence required [8] we labelled measures with CW, USPSTF or NICE evidence only, as 'unknown'. We valued the first level highest, and the third level lowest.

For group B, we distinguished the same levels of evidence. However, here we specifically searched for elements of a quality label indicating the soundness of the measure. A National Quality Forum (NQF) endorsement corresponds with the qualification of 'minor or no evidence gaps' [21]. Measures with such qualification have the strongest evidence base regarding importance, face validity, criterion validity, construct validity and risk adjustment [21]. Therefore, NQF endorsed measures were valued highest. The Agency of Healthcare Research and Quality (AHRQ) and the Centers of Medicare and Medicaid Services' (CMS) Quality provide information on the level of evidence by specifying the literature underpinning the measure. Therefore, measures from these sources were valued second best.

For both groups, our assessment was limited to the evidence provided in the reviewed article and the first document retrieved by reference tracking.

Validity

We selected a subset of five unique measures in order to gain insight in the quality of the low value care measures. Ideally, we would extensively assess each measure regarding their validity. However, for 115 measures this was beyond the scope of this review. Therefore, we chose five unique measures that appeared most frequently in the reviewed articles, assuming more information on validity to be available for these measures. For these five measures, we searched for evidence regarding the measures' validity by reviewing the original source and reference tracking. In addition, we performed a PubMed search using key words from the name of the measures (i.e. diagnosis and procedure) and "low-value" or "overuse", augmented with "validity". Specifically, we searched for studies that aimed to assess the validity of the selected low-value care measures. Hereby, we distinguished between the most commonly used types of validity (as seen in e.g. [21, 25, 26]): face validity, coding/criterion validity (i.e. reflect actual cases low-value care) and construct validity (i.e. supported by correlations to other measures indicating low-value care) [21]. Face validity refers to the empirical or clinical rationale of the measure, and therefore we used the information from Table 2 for this criterion.

RESULTS

Article retrieval

Our literature search yielded 292 potentially relevant articles (Figure 1). Based on titles and abstracts, 108 articles were selected for full-text retrieval and thorough screening. This screening process generated 23 articles that were eligible for review. Main reasons for exclusion were using a different definition of low-value care (n=138), for example articles on garbage, patient safety or drug abuse, or not providing clinical details (n=49). Figure 1 shows all reasons for exclusion.

Article characteristics

All articles were published after 2011 and the vast majority of the 23 included articles originated from the United States (n=22) (Table 1). Seven articles explicitly focused on low-value care measures. One of these reviewed the literature on low-value care measurement [19], and six were empirical studies measuring low-value care utilization [3, 9, 11, 12, 20, 27]. Low-value care recommendations were presented in 17 articles of which most were related to the CW campaign (n=12).

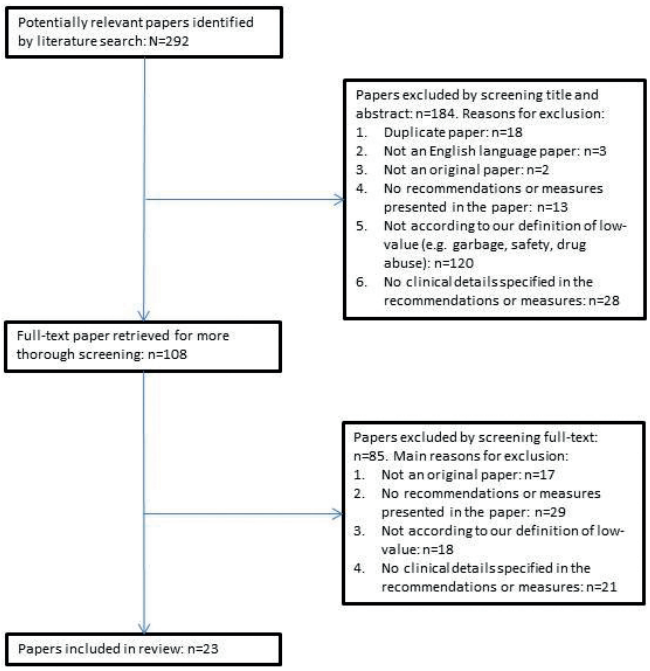


Figure 1: Flow chart summarizing article selection

Our search yielded 115 low-value care measures and 412 low-value care recommendations. Additional file 2 shows the characteristics of the 115 low-value care measures (i.e. containing a numerator and denominator). Out of these 115 measures, 42 contained exclusion criteria. For one of these measures (measure no. 72, Additional file 2), the direction of the measure was specified.

Table 1: General characteristics of the included articles (n=23)

First author	Year of publication	Country	Aim	Method	Number retrieved		Recommendation initiative
					Measures*	Recommendations	
AGS Choosing Wisely Workgroup [8]	2013	US	To identify five services that physicians and patients should question.	Review + Delphi / consensus	0	5	CW
AGS Choosing Wisely Workgroup [28]	2014	US	To identify another five services that physicians and patients should question.	Delphi / consensus	0	5	CW
Amos [29]	2015	US	To determine the prevalence of PIMs for older adults in Elimia-Romagna, Italy, using updated Maio criteria.	Empirical analysis	0	16	Other
Bulger [30]	2013	US	To identify five services that physicians and patients should question.	Review + Delphi / consensus	0	5	CW
Chan [19]	2013	US	To describe and critique the current state of overuse measurement.	Review	37	122	Other

Table 1: General characteristics of the included articles (n=23)

First author	Year of publication	Country	Aim	Method	Number retrieved		Recommendation initiative
					Measures*	Recommendations	
Colla [9]	2015	US	To develop claims-based algorithms to estimate the prevalence of Choosing Wisely services and to examine the demographic, health and health care system correlates of low-value care at a regional level.	Empirical analysis	11	0	N.A.
Elshaug [31]	2012	AUS	To develop and apply a novel method for scanning a range of sources to identify existing health care services (excluding pharmaceuticals) that have questionable benefit, and produce a list that warrant further investigation.	Review	0	174	Other
Halpern [32]	2014	US	To present the Critical Care Societies Collaborative top 5 list in Critical Care Medicine and describe its development.	Review + Delphi / consensus	0	5	CW
Hicks [33]	2013	US	To identify five services that physicians and patients should question.	Review + Delphi / consensus	0	5	CW

Table 1: General characteristics of the included articles (n=23)

First author	Year of publication	Country	Aim	Method	Number retrieved		Recommendation initiative
					Measures*	Recommen-dations	
Kale [27]	2013	US	The objective of this study was to determine whether the overuse and misuse of health care services in the ambulatory setting has decreased in the past decade.	Empirical analysis	13	0	N.A.
Keyhani [34]	2013	US	To compare rates of overuse in different health care systems and examine whether certain systems of care or insurers have lower rates of overuse of health care services.	Systematic review	0	7	Other
Korenstein [35]	2012	US	To perform an extensive search for studies of overuse of therapeutic procedures, diagnostic tests, and medications in the United States and describe the state of the literature.	Extensive search	0	33	Other
Mathias [11]	2012	US	To characterize performance on imaging-use measures, determine whether performance was consistent across measures, and identify hospital characteristics associated with highest-decile imaging use.	Empirical analysis	4	0	N.A.

Table 1: General characteristics of the included articles (n=23)

First author	Year of publication	Country	Aim	Method	Number retrieved		Recommendation initiative
					Measures*	Recommen-dations	
Morden [12]	2014	US	To measure the prevalence and describe the geographic variation of short-interval (repeated in under 2 years) DXAs among Medicare beneficiaries and estimated the cost of this testing and its responsiveness to payment change.	Empirical analysis	4	0	N.A.
Onuoha [36]	2014	US	To develop a top 5 list of unnecessary medical services in anesthesiology.	Review + Delphi / consensus	0	5	CW
Quinonez [37]	2013	US	To produce top 5 lists.	Review + Delphi / consensus	0	5	CW
Rouster-Stevens [38]	2014	US	To create a pediatric rheumatology Top 5 list as part of the American Board of Internal Medicine Foundation's Choosing Wisely campaign.	Review + Delphi / consensus	0	5	CW

Table 1: General characteristics of the included articles (n=23)

First author	Year of publication	Country	Aim	Method	Number retrieved		Recommendation initiative
					Measures*	Recommen-dations	
Schuur [39]	2014	US	To create a top-five list of tests, treatments, and disposition decisions that are of little value, are amenable to standardization, and are actionable by emergency medicine clinicians.	Delphi / consensus	0	5	CW
Schwartz [3]	2014	US	To develop claims-based measures of low-value services, examine service use (and associated spending) detected by these measures in Medicare, and determine whether patterns of use are related across different types of low-value services.	Empirical analysis	26	0	N.A.
Segal [20]	2014	US	To identify a set of possible indicators of overuse that can be operationalized with claims data and to describe variation in these indicators across the hospital referral regions (HRRs).	Empirical analysis	20	0	N.A.
Wiener [40]	2014	US	To create a top 5 list.	Review + Delphi / consensus	0	5	CW

Table 1: General characteristics of the included articles (n=23)

First author	Year of publication	Country	Aim	Method	Number retrieved		Recommendation initiative
					Measures*	Recommen-dations	
Williams [41]	2012	US	To present the final five Choosing Wisely Don't do recommendations, the rationale for these specific recommendations, and two other recommendations.	Delphi / consensus	0	5	CW
Wood [42]	2013	US	To report on the CW top 5 list.	Review + Delphi / consensus	0	5	CW

*at least a numerator and denominator was specified; AGS: American Geriatrics Society; AUS: Australia; CW: Choosing Wisely; N.A.: Not Applicable; PIM: Potentially Inappropriate Medications; US: United States

Low-value care recommendations and measures by function in health care

Figure 2 displays an overview of low-value care recommendations and measures categorized by health care function [24]. Here, we combined recommendations and measures covering the same combination of diagnosis and procedure. For instance, we found 8 measures for imaging in low back pain (measure no. 2-9, Additional file 2) using slightly varying exclusion criteria regarding e.g. age category (18-50 years versus 18-55 years) or intervention (imaging in general versus specific MRI). These eight measures were combined into a single group. In this manner, we found that 115 measures and 101 low-value care recommendations corresponded with 65 measure groups. The remaining recommendations (n=412-101=311) were aggregated into 241 new recommendation groups.

In the cure dimension we found 87 measures, which we further subdivided in general care (n=85) and specialized care (n=2). Most measures in the cure dimension were in imaging (n=50) or pharmaceutical goods (n=15). The remaining measures were categorized in long-term care (n=3) and secondary prevention (n=25).

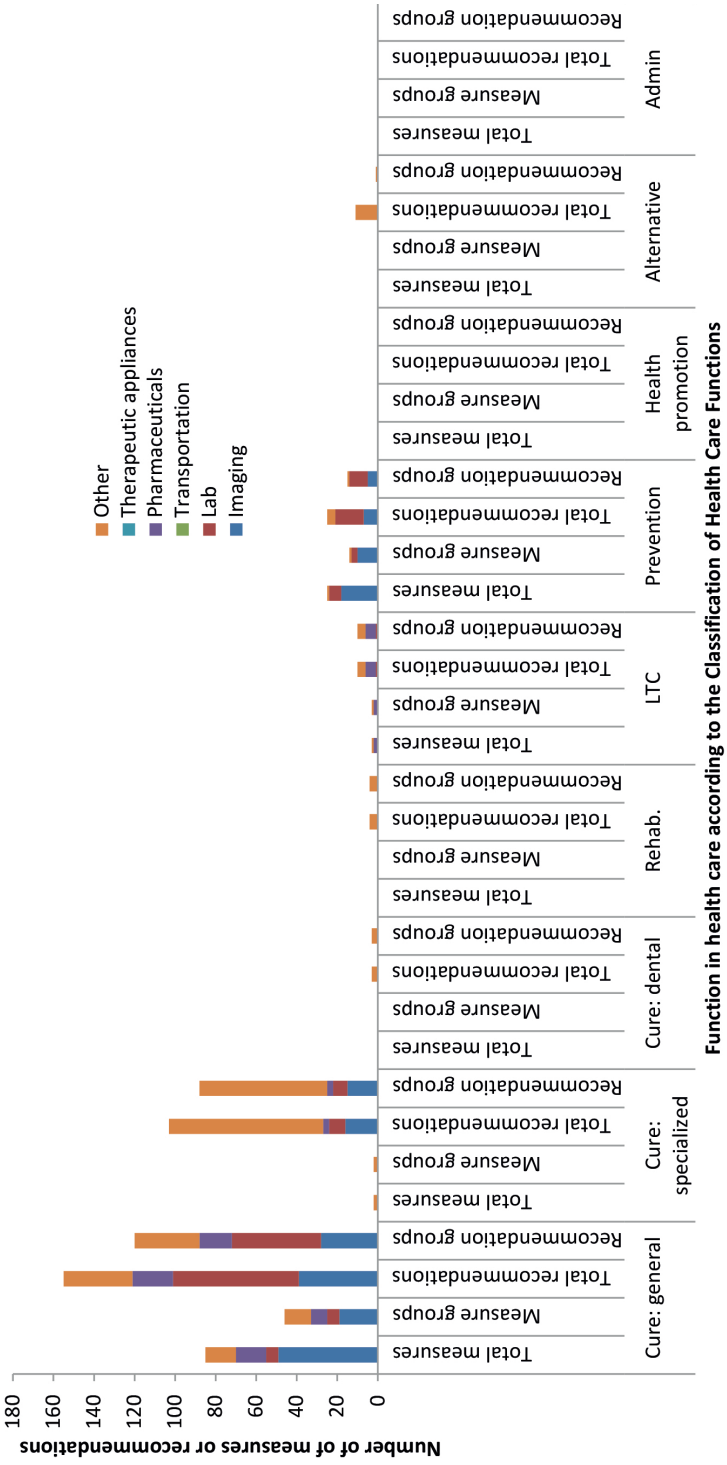
Quality of low-value care measures

Development process

Approximately half of the measures (n=62) originated from low-value care recommendations (group A). Although the authors of the articles [3, 9, 12, 20, 27] described the methods to operationalize the low-value care measures, it was not specifically described how each specific low-value care recommendation was translated into a measure, i.e. how the denominator, numerator, exclusion and direction were determined for the purpose of the study. We did find that the measures developed by clinicians (n=18) [9, 20] used (a combination of) International Classification of Diseases (ICD-9) and/or current procedural terminology (CPT) codes to construct the denominator [2, 9, 20, 27]. The other half of the measures (n=54) were developed by institutions (group B), including the NQF (n=25) [9, 19, 20], the AHRQ (n=10) [19, 20, 27], CMS QualityNet (n=16) [11, 19, 20] and Blue Cross Blue Shield (n=2) [20].

Level of evidence

Table 2 shows the level of evidence provided in the referenced sources for each measure. In group A, the recommendations were mainly derived from CW, USPSTF, and NICE (n=45). Other group A measures originated from guidelines, peer-reviewed literature or sources that summarized low-value services [43].



Admin.: Administrative; Alternative: Traditional, Complementary and Alternative Medicine; LTC: Long Term Care; Rehab.: Rehabilitative care; *We yielded 115 low-value care measures and 412 recommendations from the literature. Since 101 recommendations had a similar subjects as the measures, we subtracted these from 412 recommendations. That leaves 311 recommendations. Therefore, the total recommendations and measures in figure is 311+101=426.

Figure 2: Number of low-value care recommendations and measures categorized by the OECD/WHO/Eurostat Classification of Health Care Functions (n=426)*

Three measures (measure no. 39, 40, 46; Additional file 2) were assigned the highest level of evidence (1), as they were underpinned by guidelines and literature (trial or review) and recommendations. For most measures (n=33), however, we found guideline or literature evidence solely. For one measure (measure no. 101) the USPSTF considered the evidence for the underlying recommendation insufficient to assess the benefits and harms of the procedure, which we therefore assigned with the lowest level of evidence. At the time of our review, for 24 measures, we considered the level of evidence to be unknown. In group B, we found 19 measures [9, 19] supported by a quality label (NQF). For six measures the NQF endorsement was removed (n=4) or not found (n=2). Although the AHRQ website provides detailed information on the measures, we found no quality label, such as the NQF endorsement. We found seven measures (measure no. 1, 4, 55, 60, 69, 70, 86) displaying measurement characteristics (e.g. domain (process/outcome), description of denominator and numerator and target population) and evidence supporting the measure. The measures derived from QualityNet [44] were described in detail, however, no evidence supporting the description was provided.

Validity

Table 3 shows the validity of the five measures that were found most frequently (n=26). Two measures had the highest level of evidence. Our search yielded no information on coding/criterion validity and construct validity for the included measures, while four out of five measures are currently used in practice.

DISCUSSION

To the best of our knowledge, this is the first systematic literature review identifying, categorizing and assessing the scope and quality of low-value care measures. We obtained 115 low-value care measures from the literature. Out of these 115 measures, 87 focused on the cure sector (primary and specialized care), 25 on secondary prevention and 3 on long-term care. Most measures (n=62) originated from low-value care recommendations, while 53 were previously developed by institutions as the National Quality Forum. Three measures were assigned the highest level of evidence, as they were underpinned by both guidelines and literature evidence. For other measures, such a level of evidence was not transparently apparent. We do not conclude that these measures are invalid, because validity tests may not have been performed at all. Nevertheless, a lack of evidence is present at least. Our search yielded no information on coding/criterion validity and construct validity for the included subset of measures in this emerging field. Despite this, most measures are currently used in practice.

Table 2: Level of evidence of low-value care measures

Level of evidence	Group A: Recommendation source	Measure numbers*	Count
1	CW, NICE or USPSTF recommendations; Guideline; Literature evidence (review or clinical trial)	39, 40, 46	3
2	CW, NICE or USPSTF recommendations; Literature evidence (review or clinical trial)	13, 14, 19, 20, 22, 23, 24, 25, 26, 44, 48, 50, 55, 77, 80, 90, 95, 103, 112, 115	20
2	CW, NICE or USPSTF recommendations; Guideline	33, 53	2
2	Literature evidence (reviews or clinical trial)	3, 21, 58, 76, 78, 81, 82, 83, 85, 89	10
2	Guideline	54, 57	2
3	USPSTF concludes that evidence is insufficient	101	1
Unknown	Literature: other compiled low-value service lists	47, 49, 113	3
Unknown	USPSTF recommendation not found	104, 107	2
Unknown	CW, NICE or USPSTF recommendations	34, 38, 43, 45, 51, 52, 59, 61, 84, 92, 98, 102, 105, 106, 108, 109, 110, 111, 114	19
Level of evidence	**Group B: Institutional measure status**	**Measure numbers***	
1	NQF endorsed	5, 11, 16, 18, 41, 56, 62-67, 72, 73, 91, 93, 94, 96, 97	19
2	AHRQ measure supported by a clinical practice guideline or other peer-reviewed synthesis of clinical research evidence and one or more research studies published in a National Library of Medicine (NLM) indexed, peer-reviewed journal	60, 69, 70	3

Table 2: Level of evidence of low-value care measures (continued)

Level of evidence	Group B: Institutional measure status	Measure numbers*	
2	AHRQ measure supported by a clinical practice guideline or other peer-reviewed synthesis of the clinical research evidence	1, 4, 55, 86	4
2	CMS QualityNet	2, 7, 8, 9, 27, 28, 29, 30, 31, 32, 37, 42, 99, 100	14
3	NQF endorsement removed since April 2014	6, 10, 74, 75	4
unknown	NQF endorsement not found	17, 71	2
unknown	AHRQ measure/guideline not found	68, 79, 87	3
unknown	CMS QualityNet under revision	15	1
unknown	CMS not found	12	1
unknown	BCBS AQC measures not found	35, 36	2

*: measure numbers are in correspondence with Additional file 2; AHRQ: Agency for Healthcare Research and Quality; BCBS AQC: Blue Cross Blue Shield, The Alternative Quality Contract; CMS: Centers for Medicare & Medicaid Services; CW: Choosing Wisely; IOM: Institute of Medicine; NICE: National Institute for Clinical Excellence (UK): do not do recommendations; NQF: National Quality Forum; USPSTF: United States Preventive Services Task Force.

Table 3: Validity of the top five published low-value care measures

	Preoperative cardiac tests for non-cardiac low-risk surgery	**Antibiotics for upper respiratory tract infections**	**Imaging for low-back pain**	**Cervical cancer screening**	**Imaging for sinusitis diagnosis**
Number of measures included in review #	4 (measure no.: 42-44, 48)	7 (measure no.: 57-59, 62, 63, 65, 66)	8 (measure no.: 2-9)	3 (measure no.: 110-112)	4 (measure no.: 33, 35, 36, 59)
*Measure criteria **					
Face validity ^	Yes: level of evidence is 2	Yes, level of evidence is 1	Yes, level of evidence is 1	Yes, level of evidence is 2	Yes, level of evidence is 2
Coding / criterion validity	Not found	Not found	Not found	Not found	Not found
Construct validity	Not found	Not found	Not found	Not found	Not found
Used in practice	Yes, for payment determination (Hospital Outpatient Quality Reporting) [45] [46]	Yes, in Physician Quality Reporting System [46]	Yes, for payment determination (Hospital Outpatient Quality Reporting) [45] [46]	Yes, in Physician Quality Reporting System [46]	Not found

#: Measure numbers corresponding with Additional file 2 between brackets; *: Criteria for quality measures (AHRQ); ^: For level of evidence also see Table 2.

Low-value care measures have received increased attention and are now used for monitoring purposes, alignment of financial incentives [14, 47] and, in the foreseeable future, in shared saving programs such as the Alternative Quality Contract (AQC) [48]. In this manner, low-value care measurement may incentivize providers and insurers to shift resources from low-value services to high-value services [49]. Our findings show that more attention is needed for the evidential underpinning and quality of these measures. Otherwise, the lack of transparency and evidence will reduce acceptance of low-value care measures by its users. Additionally, using measures of low quality, might lead to negative consequences including underuse of indicated services, cost-shifting, damages to the patient-physician relationship, provider dissatisfaction, adverse health effects, or patient selection [18].

Our review showed that more than half of the low-value care measures originated from low-value service recommendations (i.e. CW, NICE, USPSTF). This implies that the empirical evidence of many low-value care measures is based on the evidence supporting the underlying low-value service recommendations. However, criteria for the development of recommendation lists remains rather vague in the CW initiative, as well as in other similar campaigns [8]. Therefore, more transparency regarding the evidential underpinning of the recommendations is needed. Next to the importance of evidence underlying both low-value service recommendations and measures, one should be aware that the aim of low-value service recommendations differs from the aim of low-value care measures. The aim of CW recommendations is patient and physician awareness, while the aim of low-value care measures in turn may be to inform decisions on several levels. Consequently, requirements for the quality and development of recommendations and measures approaches vary accordingly.

We found that most current low-value care measures are concentrated in the cure sector even though it was argued that low-value services are provided and used along the entire continuum of care [22]. For example, we only found four low-value care recommendations (that could possibly be transformed into low-value care measures) in rehabilitative care and none in the health promotion domain. This is probably the result of most measures originating from the CW initiative, which has its origin in the cure sector. While we acknowledge the emerging state of the field of research, we emphasize that similar consensus-based efforts are needed to stimulate the development of measures in other settings to broaden the scope and impact of the low-value care concept.

Given the potential impact of using low-value care measures, it is essential that guidelines for developing them be created by combined efforts of the involved parties: physicians,

citizens, government and insurers [18, 50]. We do not suggest creating an evidence base for *each* health care intervention demonstrating all circumstances in which it is *not* effective. This will prove an undoable exercise. Expert judgement by the clinician will always remain necessary to some degree. Therefore, other types of information, e.g. from studies on practice variation in procedure rates or cost-effectiveness studies, will remain necessary to identify inefficiencies in healthcare, especially when high quality low-value care measures are not available. We do propose using expert opinion from initiatives such as Choosing Wisely as a starting point for monitoring low-value care. These qualitative information sources can be complemented with new scientific insights. For example, the insight that certain genes predict the development of breast cancer, must be used to prevent a considerable amount of low-value care utilization. Still, as soon as we start measuring and monitoring low-value care in such areas, it will be of particular interest to fully specify and define all measurement information, such as exclusion criteria, direction and evidence supporting the measure, and to make this publicly available. Furthermore, low-value care measures should be extensively tested regarding their level of evidence and validity before implementing them for use in practice, and specifically for the measures that are already in use. Recently, articles started studying aspects that are closely related to validity. As for example, Schwartz et al. [3] who found that the sensitivity and specificity strongly depends on the definition of the measures. Notwithstanding the efforts already been made, we stress the importance of the validity of the measures specifically being studied. Another area of research would be to further standardize low-value care measures, which ideally would result in alignment of the low-value care metrics and determining specifically for what subgroup or population a service is of low-value [3, 51]. Moreover, the guidelines should take into account any differences between countries in terms of the availability and provision of healthcare services that are likely to occur due to cultural or economic differences.

Another important issue to pay further attention to is the data requirements. Measuring low-value care utilization requires information on services provided to patients in combination with diagnosis and possibly additional patient characteristics. It is not clear to which extent current data sources can provide this information [3, 4], since rather detailed data need to be registered and data sources, such as claims data and detailed (hospital) registration data need to be connected in order to retrieve the necessary information.

Limitations

Our study has two main limitations. First, we did not evaluate the quality of each individual measure. Ideally, we would extensively assess each measure regarding their

validity. To perform this task for 115 measures was, however, beyond the scope of this review. Nonetheless, we performed a first attempt in assessing the validity for the five measures that appeared most often in the literature and highlight several important general quality issues. Second, we did not include grey literature in our search. Therefore, we may have missed relevant measures. Nevertheless, for the purpose of our review, namely to systematically map the state of affairs of low-value care measurement, we are confident that the publications we did use provided sufficient evidence.

Conclusions

To conclude, our systematic review provides insight in the current state of low-value care measures. It shows that current low-value care measures only cover a selective part of the health care system. To achieve their full potential, future research should be focused on generating clear information about the level of evidence and validity to identify measures that truly represent low-value care in this emerging field of research. This will contribute to creating and maintaining the support of stakeholders who will use these measures for monitoring purposes and innovative insurer-provider contracts, all aiming to improve efficiency in health care with better health outcomes.

ACKNOWLEDGEMENTS

We are grateful to Margje H. Haverkamp, PhD (**M.H.Haverkamp@lumc.nl**) for helpful comments on an earlier version of this manuscript.

REFERENCES

1. Institute of Medicine (IOM). *Crossing the Quality Chasm: A New Health System for the 21st Century*. 2001, National Academy Press: Washington DC.
2. Schwartz, A.L., et al. *Measuring low-value care in Medicare.* JAMA Intern Med, 2014. 174(7): p. 1067-76.
3. Schwartz, A.L., Landon, B.E., Elshaug, A.G., Chernew, M.E., McWilliams, J.M., *Measuring Low-Value Care in Medicare.* JAMA Intern Med, 2014. 174(7): p. 1067-1076.
4. Bhatia, R.S., et al., *Measuring the effect of Choosing Wisely: an integrated framework to assess campaign impact on low-value care.* BMJ Qual Saf, 2015. 24(8): p. 523-31.
5. Levinson, W. and Huynh, T., *Engaging physicians and patients in conversations about unnecessary tests and procedures: Choosing Wisely Canada.* CMAJ, 2014. 186(5): p. 325-326.
6. Levinson, W., et al., *'Choosing Wisely': a growing international campaign.* BMJ Qual Saf, 2015. 24(2): p. 167-74.
7. Angus, D.C., et al., *Choosing wisely in critical care: maximizing value in the intensive care unit.* Crit Care Med, 2014. 42(11): p. 2437-8.
8. American Geriatrics Society Choosing Wisely Workgroup (AGSCW), *American Geriatrics Society identifies five things that healthcare providers and patients should question.* J Am Geriatr Soc, 2013. 61(4): p. 622-31.
9. Colla, C.H., et al., *Choosing wisely: prevalence and correlates of low-value health care services in the United States.* J Gen Intern Med, 2015. 30(2): p. 221-8.
10. Keyhani, S., et al., *The relationship between geographic variations and overuse of healthcare services: a systematic review.* Med Care, 2012. 50(3): p. 257-61.
11. Mathias, J.S., J. Feinglass, and D.W. Baker, *Variations in US hospital performance on imaging-use measures.* Med Care, 2012. 50(9): p. 808-14.
12. Morden, N.E., et al., *Overuse of short-interval bone densitometry: assessing rates of low-value care.* Osteoporos Int, 2014. 25(9): p. 2307-11.
13. Morden, N.E., et al., *Choosing wisely--the politics and economics of labeling low-value services.* N Engl J Med, 2014. 370(7): p. 589-92.
14. Shaw, D. and Melton, P., *Should GPs be paid to reduce unnecessary referrals?* BMJ, 2015. 351.
15. Baker, D.W., et al., *Design and use of performance measures to decrease low-value services and achieve cost-conscious care.* Ann Intern Med, 2013. 158(1): p. 55-9.
16. Rosenberg, A., et al., *Early Trends Among Seven Recommendations From the Choosing Wisely Campaign.* JAMA Intern Med, 2015: p. 1-9.
17. Struijs, J.N., Drewes, H.W., and Stein, K.V., *Beyond integrated care: challenges on the way towards population health management.* Int J Integr Care, 2015; 15:e043.
18. Mathias, J.S. and Baker, D.W., *Developing quality measures to address overuse.* JAMA, 2013. 309(18): p. 1897-8.
19. Chan, K.S., et al., *The state of overuse measurement: a critical review.* Med Care Res Rev, 2013. 70(5): p. 473-96.
20. Segal, J.B., et al., *Identifying possible indicators of systematic overuse of health care procedures with claims data.* Med Care, 2014. 52(2): p. 157-63.
21. Agency for Healthcare Research and Quality (AHRQ). *Guidance on Using the AHRQ QI for Hospital-Level Comparative Reporting.* AHRQ, 2009.
22. The King's Fund, *Better value in the NHS: The role of changes in clinical practice*. 2015, The King's Fund: London.
23. Higgins, J.P.T, Green, S. (editors). *Cochrane Handbook for Systematic Reviews of Interventions*. Version 5.1.0. [updated March 2011]. The Cochrane Collaboration, 2011.
24. Organisation for Economic Co-operation and Development (OECD) / World Health Organization (WHO) / Eurostat, *A System of Health Accounts*. 2011, OECD Publishing: Paris.

25. Kaplan, R.M., Bush, J.W., and Berry, C.C., *Health status: types of validity and the index of well-being.* Health Serv Res, 1976. 11(4): p. 478-507.

26. Winters, B.D., et al., *Validity of the Agency for Health Care Research and Quality Patient Safety Indicators and the Centers for Medicare and Medicaid Hospital-acquired Conditions: A Systematic Review and Meta-Analysis.* Med Care, 2016.

27. Kale, M.S., et al., *Trends in the overuse of ambulatory health care services in the United States.* JAMA Intern Med, 2013. 173(2): p. 142-8.

28. American Geriatrics Society Choosing Wisely Workgroup (AGSCW), *American Geriatrics Society identifies another five things that healthcare providers and patients should question.* J Am Geriatr Soc, 2014. 62(5): p. 950-60.

29. Amos, T.B., et al., *Inappropriate prescribing in a large community-dwelling older population: a focus on prevalence and how it relates to patient and physician characteristics.* J Clin Pharm Ther, 2015. 40(1): p. 7-13.

30. Bulger, J., et al., *Choosing wisely in adult hospital medicine: five opportunities for improved healthcare value.* J Hosp Med, 2013. 8(9): p. 486-92.

31. Elshaug, A.G., Watt, A.M., Mundy, L., Willis, C.D., *Over 150 potentially low-value health care practices: an Australian study.* Med J Aust, 2012. 197(10): p. 556-560.

32. Halpern, S.D., et al., *An official American Thoracic Society/American Association of Critical-Care Nurses/American College of Chest Physicians/Society of Critical Care Medicine policy statement: the Choosing Wisely Top 5 list in Critical Care Medicine.* Am J Respir Crit Care Med, 2014. 190(7): p. 818-26.

33. Hicks, L.K., et al., *The ASH choosing wisely® campaign: Five hematologic tests and treatments to question.* Blood, 2013. 122(24): p. 3879-3883.

34. Keyhani, S., et al., *Overuse and systems of care: a systematic review.* Med Care, 2013. 51(6): p. 503-8.

35. Korenstein, D., Falk, R., Howell, A.E., Bishop, T, Keyhani, S. , *Overuse of health care services in the United States: an understudied problem.* Arch Intern Med, 2012. 172(2): p. 171-178.

36. Onuoha, O.C., V.A. Arkoosh, and L.A. Fleisher, *Choosing wisely in anesthesiology: The gap between evidence and practice.* JAMA Internal Medicine, 2014. 174(8): p. 1391-1395.

37. Quinonez, R.A., et al., *Choosing wisely in pediatric hospital medicine: five opportunities for improved healthcare value.* J Hosp Med, 2013. 8(9): p. 479-85.

38. Rouster-Stevens, K.A., et al., *Choosing wisely: The american college of rheumatology's top 5 for pediatric rheumatology.* Arthritis Care and Research, 2014. 66(5): p. 649-657.

39. Schuur, J.D., et al., *A top-five list for emergency medicine: a pilot project to improve the value of emergency care.* JAMA Intern Med, 2014. 174(4): p. 509-15.

40. Wiener, R.S., et al., *An official American Thoracic Society/American College of Chest Physicians policy statement: the Choosing Wisely top five list in adult pulmonary medicine.* Chest, 2014. 145(6): p. 1383-91.

41. Williams, A.W., et al., *Critical and honest conversations: the evidence behind the "Choosing Wisely" campaign recommendations by the American Society of Nephrology.* Clin J Am Soc Nephrol, 2012. 7(10): p. 1664-72.

42. Wood, D.E., et al., *Choosing wisely: cardiothoracic surgeons partnering with patients to make good health care decisions.* Ann Thorac Surg, 2013. 95(3): p. 1130-5.

43. Qaseem, A., et al., *Appropriate use of screening and diagnostic tests to foster high-value, cost-conscious care.* Ann Intern Med, 2012. 156(2): p. 147-9.

44. Centers for Medicaid & Medicare Services (CMS). *Imaging Efficiency Measures*. 2011; Available from: https://www.qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetTier2&cid=1228695266120.

45. Centers for Medicaid & Medicare Services (CMS). *Imaging Efficiency Measures*. 2011.

46. National Quality Forum (NQF). *QPS Tool*. Available from: http://www.qualityforum.org/QPS/

47. McCarthy, M., *US Choosing Wisely campaign has had only modest success, study finds.* BMJ, 2015. 351: p. h5437.

48. Chernew, M.E., et al., *Private-payer innovation in Massachusetts: the 'alternative quality contract'.* Health Aff (Millwood), 2011. 30(1): p. 51-61.

49. Fendrick, A.M., Smith D.G., and, Chernew, M.E., *Applying value-based insurance design to low-value health services.* Health Aff, 2010. 29(11): p. 2017-21.

50. Willson, A., *The problem with eliminating 'low-value care'*. BMJ Qual Saf, 2015. 24(10): p. 611-4.

51. Elshaug, A.G., J. McWilliams, and B.E. Landon, *The value of low-value lists*. JAMA, 2013. 309(8): p. 775-776.

ADDITIONAL FILES

A1. Search strategy

Database: MEDLINE 1950 to present, MEDLINE In-Process & Other Non-Indexed Citations

Search Strategy:

---

1 *health services/ or *adolescent health services/ or *community health services/ or *child health services/ or *community health nursing/ or *community mental health services/ or *community pharmacy services/ or *community health centers/ or *home care services/ or *maternal health services/ or *occupational health services/ or *preventive health services/ or *preventive medicine/ or *dental health services/ or *emergency medical services/ or *emergency service, hospital/ or *hospitals/ or *health services for the aged/ or *mental health services/ or *nursing services/ or *personal health services/ or *pharmaceutical services/ or *rehabilitation/ or *reproductive health services/ or *rural health services/ or *suburban health services/ or *women's health services/ or *mass screening/ (295687)

2 *primary health care/ or *general practice/ or *family practice/ or *physicians, family/ or *professional practice/ or *physicians practice patterns/ or *comprehensive health care/ or *managed care programs/ or *delivery of health care/ or *delivery of health care, integrated/ or *patient care management/ or *nursing process/ or *nursing/ or *nurse practitioners/ or *telemedicine/ or *health services administration/ or *health services research/ or *translational medical research/ or *health facility administration/ or *health facilities/ or *health maintenance organizations/ or *health planning/ or *regional health planning/ or *community health planning/ or *regional medical programs/ or *health policy/ or *national health programs/ or *social work/ or *social welfare/ or *child welfare/ or *infant welfare/ or *maternal welfare/ or *government regulation/ or *government programs/ or *multi-institutional systems/ (344106)

3 (health system* or healthcare or health service* or care or (health and services) or health planning or health policy or health reform or welfare or preventive service* or screening).ti. (523842)

4 (family practic* or family physician* or general practi* or "gps" or "gp's" or nurse practitioners or community health center* or community health centre* or municipal health center* or municipal health centre* or health center* or health centre*).ti. (47193)

5 "quality of health care"/ or "quality indicators, health care"/ or health care costs/ (93908)

6 1 or 2 or 3 or 4 or 5 (1008916)

7 ((identif* or indicator* or correlat* or pattern* or predictor* or measur* or assess* or examin* or classify* or categoriz* or characterization or quantif* or variation*) adj5 (low value or "no value" or overuse or overutilization or overused or misuse or disuse or (wasteful adj3 "use") or (wasteful adj3 services))).tw. (1541)
8 ((estimat* or evaluat* or distinguishing or labe?ling or compar* or potential* or factors or prevalence or rate* or degree or evidence or understanding) adj5 (low value or "no value" or overuse or overutilization or overused or misuse or disuse or (wasteful adj3 "use") or (wasteful adj3 services))).tw. (2408)
9 ((identif* or indicator* or correlate* or pattern* or predictor* or measur* or assess* or examin* or classify* or categoriz* or characterization or quantif* or variation*) adj5 ((unsafe adj3 practice*) or (ineffective adj3 practice*) or (ineffective adj3 care) or (ineffective adj3 healthcare) or (inappropriat* adj3 practice*) or (inappropriat* adj3 "use") or (harmful adj3 practice*))).tw. (490)
10 ((estimat* or evaluat* or distinguishing or labe?ling or compar* or potential* or factors or prevalence or rate* or degree or evidence or understanding) adj5 ((unsafe adj3 practice*) or (ineffective adj3 practice*) or (ineffective adj3 care) or (ineffective adj3 healthcare) or (inappropriat* adj3 practice*) or (inappropriat* adj3 "use") or (harmful adj3 practice*))).tw. (810)
11 ((identif* or indicator* or correlate* or pattern* or predictor* or measur* or assess* or examin* or classify* or categoriz* or characterization or quantif* or variation) adj5 waste).tw. (2127)
12 ((estimat* or evaluat* or distinguishing or labe?ling or compar* or potential* or factors or prevalence or rate or degree or evidence or understanding) adj5 waste).tw. (2833)
13 ((index or instrument) and (low value or "no value" or overuse or overutilization or overused or misuse or disuse or (wasteful adj3 "use") or (wasteful adj3 services))).ti. (21)
14 ("choosing wisely" or "do not do recommendations").tw. (134)
15 6 and (7 or 8 or 9 or 10 or 11 or 12 or 13 or 14) (1126)
16 prescription drug misuse/ or (prescription misuse or substance misuse or drug misuse or alcohol or hazardous drinking or substance).ti. (100941)
17 (hospital waste or medical waste or biomedical waste or bio-medical waste or healthcare waste or health-care waste or care waste or genotoxic waste or plate waste or waste anesthetic gases or infectious waste or pharmaceutical waste or solid waste or liquid waste or aeration or waste disposal or waste treatment or electronic waste or waste waster* or waste materials or aeration).ti. (4893)
18 15 not (16 or 17) (927)
19 (practice* or low value or value or "no value" or overuse or misuse or overutilization or overused or inappropr* or ineffective or unsafe or "use" or waste).ti. or ("choosing wisely" or "do not do recommendations").tw. (664168)

20 18 and 19 (487)
21 20 and english.lg. (452)
22 limit 21 to yr=2010-2015 (227)
23 22 not (letter or comment or news).pt. (205)
24 remove duplicates from 23 (201)

A2. Full list of low-value care measures

Talbe A2. Low-value care measures including numerator, denominator, exclusion criteria, direction and measure source and reference specified by function according to the OECD/WHO/Eurostat Classification of Health Care Functions (n=115)

No	Measure	Measure details				Original source		Ref.	Function		Non-function
		Numerator*	Denominator*	Exclusion	Direction	Measure	Recommendation				
1	Barium swallow test for GERD	% patients ≥ 18 seen for an initial evaluation of GERD who did not have a barium swallow test ordered		-	higher rate is better performance	NGC	-	Chan [19]	Cure	General	Imaging
2	MRI Back	All patients undergoing an MRI of the lumbar spine with a diagnosis of low back pain without evidence of antecedent conservative therapy (e.g. physical therapy within 60 d of the MRI, chiropractic manipulation within 60 d of the MRI, or low back pain evaluation and management using CPT codes within 28 to 60 d of the MRI-during which time the patient is likely to have been educated or home exercises are likely to have been recommended).	All patients undergoing an MRI of the lumbar spine with a diagnosis of low back pain.	MRI's performed in patients with diagnosis of cancer, trauma, intravenous drug abuse, neurological impairment, immune deficiency, or intra spinal abscess.	-	CMS / QualityNet	-	Mathias [11]	Cure	General	Imaging
3	Back pain images for patients with nonspecific low back pain	Back imaging with a diagnosis of lower back pain	Patients with back pain	-	-	-	CW/ NICE / Lit	Schwartz [3]	Cure	General	Imaging
4	X-ray for back pain in adults aged 18-55 y	Visits by adults with acute back pain who received x-ray	Visits by adults with acute back pain	Visits by adults with malignancy, weight loss, fever, cachexia, or neurologic signs	-	NCQA	-	Kale [27]	Cure	General	Imaging

Table A2. Low-value care measures including numerator, denominator, exclusion criteria, direction and measure source and reference specified by function according to the OECD/WHO/Eurostat Classification of Health Care Functions (n=115) (continued)

No	Measure	Measure details				Original source		Ref.	Function		Non-function
		Numerator*	Denominator*	Exclusion	Direction	Measure	Recommendation				
5	Imaging for low back pain	% members aged 18-50 with negative diagnosis history who had outpatient or ER visit with primary diagnosis of low back pain who did not have an imaging study (plain x-ray, MRI CT scan) within 28 days of this visit. Members with any low back pain diagnosis during the 180 days of the index visit, who have a diagnosis of cancer, who had a diagnosis in the prior 12 months of recent trauma, intravenous drug abuse, r neurological impairment are excluded.	-	higher rates is better performance	NGC / NQF	-	Chan [19]	Cure	General	Imaging
6	Imaging for acute back pain	% patients with a diagnosis for back pain for whom the physician ordered imaging studies during the 6 weeks after pain onset, in the absence of 'red flags'.	-	lower rates is better performance	NQF	-	Chan [19]	Cure	General	Imaging
7	MRI lumbar spine for low back pain	% MRI of the lumbar spine studies with a diagnosis of low back pain on the imaging claim and for which the patient did not have prior claims-based evidence of antecedent conservative therapy	-	lower rates is better performance	QualityNet	-	Chan [19]	Cure	General	Imaging

Table A2. Low-value care measures including numerator, denominator, exclusion criteria, direction and measure source and reference specified by function according to the OECD/WHO/Eurostat Classification of Health Care Functions (n=115) (continued)

No	Measure	Measure details				Original source		Ref.	Function		Non-function
		Numerator*	Denominator*	Exclusion	Direction	Measure	Recommendation				
8	MRI lumbar spine for low back pain	MRI of the lumbar spine studies with a diagnosis of low back pain (from the denominator) without the patient having claims-based evidence of prior antecedent conservative therapy. CPT=72148, or 72149, or 72158 with no codes for 97110, 97112, 97113, 97124, 97140, 98940, 98941, 98942, 98943 in the 60 days preceding the MRI of the lumbar spine AND no codes for 99210-99205, 99211 -99215, 99241-99245, 99341-99345, 99347-99350, 99354-99357, 99385-99387, 99395-99397, 99401-99404, 99455-99456, 99499 between 28 and 60 days preceding the MRI of the lumbar spine	MRI of the lumbar spine studies with a diagnosis of low back pain on the imaging claim. CPT=72148, or 72149, or 72158 AND ICD-9: 721.3, 721.90, 722.10, 722.52, 722.6, 722.93, 724.02, 724.2, 724.3, 724.5, 724.6, 724.70, 724.71, 724.79, 738.5, 739.3, 739.4, 846.0, 846.1, 846.2, 846.3, 846.8, 846.9 , 847.2	Excluded from the denominator - CPT codes: 22010-22865 and 22899 in 90 days preceding MRI; ICD-9 codes: 140-208, 230-234, 235-239, 304.0X, 304.1X, 304.2X, 304.4X, 305.4X, 305.5X, 305.6X, 305.7X, 344.60, 344.61, 729.2, 042-044, 279.3 in preceding 365 days; 800-839, 850-854, 860-869, 905-909, 926.11, 926.12, 929, 952, 958-959 in preceding 45 days; 324.9, 324.1 on same claim as MRI	-	QualityNet	-	Segal [20]	Cure	General	Imaging

Table A2. Low-value care measures including numerator, denominator, exclusion criteria, direction and measure source and reference specified by function according to the OECD/WHO/Eurostat Classification of Health Care Functions (n=115) (continued)

No	Measure	Measure details				Original source		Ref.			
		Numerator*	Denominator*	Exclusion	Direction	Measure	Recommendation			Function	Non-function
9	Don't do imaging for low back pain when no red flags are present	Beneficiaries who received a low back x-ray, CT or MRI within six weeks of incident low back pain diagnosis	Beneficiaries with low back pain over age 65 without other imaging indication	Prior diagnosis of low back pain, trauma and neurological impairment, within previous 12 months and cancer at any point during study period: E code (external causes of injury) or trauma diagnosis on imaging event claim	-	NQF	CW	Colla [9]	Cure	General	Imaging
10	Repeat imaging studies for back pain	% patients who received inappropriate repeat imaging studies in the absence of red flags or progressive symptoms		-	lower rates is better performance	NQF	-	Chan [19]	Cure	General	Imaging
11	Cardiac stress imaging (routine testing after percutaneous coronary intervention, PCI)	% of all stress SPECT MPI, stress echo, CCTA, and CMR performed routinely after PCI, with reference to timing of test after PCI and symptom status.		-	lower rates is better performance	NQF	-	Chan [19]	Cure	General	Imaging
12	Overuse of stress testing	% patients ≥ 18 with a CAD diagnosis for ≥ 12 months but no documentation of AMI OR unstable angina OR referral for cardiac rehabilitation during measurement period who received ≥ 2 stress tests within a 12-month period.		-	lower rates is better performance	CMS	-	Chan [19]	Cure	General	Imaging

Table A2. Low-value care measures including numerator, denominator, exclusion criteria, direction and measure source and reference specified by function according to the OECD/WHO/Eurostat Classification of Health Care Functions (n=115) (continued)

No	Measure	Measure details				Original source		Ref.	Function		Non-function
		Numerator*	Denominator*	Exclusion	Direction	Measure	Recommendation				
13	Stress echocardiography in symptomatic or ischemic equivalent acute chest pain	Individuals with CPT codes as listed or HCPCS codes as listed for echocardiography	Individuals with a code for emergency visit* with any of the ICD-9 diagnoses OR individuals with a hospitalization with DRGs as listed, or primary or secondary diagnosis code during hospitalization for any of the ICD-9 diagnoses	-	-	-	CW / Lit	Segal [20]	Cure	General	Imaging
14	Stress testing for stable coronary disease	Stress testing for patients with an established diagnosis of ischemic heart disease or angina (≥6 months before the stress test) and thus not done for screening purposes	IHD patients	Test not associated with inpatient or emergency care, which might be indicative of unstable angina; only patients with a past diagnosis of myocardial infarction to exclude patients with a history of non-cardiac chest pain inaccurately coded as angina (i.e., those with no underlying ischemic heart disease who might benefit from screening and optimization of medical management).	-	-	CW / Lit	Schwartz [3]	Cure	General	Imaging

Table A2. Low-value care measures including numerator, denominator, exclusion criteria, direction and measure source and reference specified by function according to the OECD/WHO/Eurostat Classification of Health Care Functions (n=115) (continued)

No	Measure	Measure details				Original source		Ref.			
		Numerator*	Denominator*	Exclusion	Direction	Measure	Recommendation		Function		Non-function
15	CT for headache in emergency department	OF ED visits with primary diagnosis of headache, the number with a brain CT on the same day; patients admitted to the hospital or have secondary diagnoses on claims indicating clinical need are excluded		-	lower rates is better performance	CMS / Lewin Group	-	Chan [19]	Cure	General	Imaging
16	Head CT imaging for mild traumatic brain injury	% adult patients presenting within 24 hours of a non-penetrating head injury with a Glasgow coma score >13 and underwent head CT for trauma in the ED who have a documented evidence-based indication prior to imaging		-	higher rates is better performance	NQF	-	Chan [19]	Cure	General	Imaging
17	MRI in adults with mild traumatic brain injury	MRI on the same claim as diagnosis if outpatient or during hospitalization if inpatient	Patients with traumatic brain injury	-	-	NQF	-	Segal [20]	Cure	General	Imaging
18	Cervical spine imaging in trauma	% adult patients undergoing cervical spine radiographs or CT for trauma who fulfill NEXUS low-risk criteria for cervical spine injury or Canadian Cervical Spine Rule documented prior to imaging. Patients are excluded if they have communication difficulties, inadequate prior radiographs, or for whom further imaging is indicated.		-	higher rates is better performance	NQF	-	Chan [19]	Cure	General	Imaging
19	EEG for headaches	EEG with headache diagnosis in the claim (specified with claim codes)	Patients with headache diagnosis	-	-	-	CW / Lit	Schwartz [3]	Cure	General	Imaging
20	Head imaging for uncomplicated headache	CT or MR imaging of the head with a diagnosis of (non-thunderclap, non-posttraumatic) headache (specified with claim codes)	Patients with headache diagnosis	-	-	-	CW / Lit	Schwartz [3]	Cure	General	Imaging

Table A2. Low-value care measures including numerator, denominator, exclusion criteria, direction and measure source and reference specified by function according to the OECD/WHO/Eurostat Classification of Health Care Functions (n=115) (continued)

No	Measure	Measure details				Original source		Ref.	Function		Non-function
		Numerator*	Denominator*	Exclusion	Direction	Measure	Recommendation				
21	Bone mineral density testing at frequent intervals	Bone mineral density test less than 2 y after prior bone mineral density test (specified with claim codes)	Patients with osteoporosis	-	-	-	Lit	Schwartz [3]	Cure	General	Imaging
22	Don't routinely repeat dual-energy x-ray absorptiometry (DXA) scans more often than once every two years	DXA scans performed on female beneficiaries at low risk for fracture within 23 months of a previous scan	DXA scans performed on female beneficiaries over age 66 at low risk for fracture	Fragility fracture or cancer diagnosis within 23 months of the index DXA scan	-	-	CW	Colla [9]	Cure	General	Imaging
23	Short interval dual energy x-ray absorptiometry scans (DXAs): population rate	DXA scans performed on female beneficiaries over age 66 at low risk for fracture	Number of DXAs performed per 100 female beneficiaries over age 66	We excluded beneficiaries not continuously enrolled in fee-for-service Medicare Parts A and B in the 23 months prior to each DXA test identified. We conservatively excluded beneficiaries who were (i) diagnosed with any cancer, except non-melanoma skin cancer (using the Clinical Classifications [17] or (ii) diagnosed with fragility fracture in the 23 months.	-	-	CW	Morden [12]	Cure	General	Imaging

Table A2. Low-value care measures including numerator, denominator, exclusion criteria, direction and measure source and reference specified by function according to the OECD/WHO/Eurostat Classification of Health Care Functions (n=115) (continued)

No	Measure	Measure details				Original source		Ref.	Non-function		
		Numerator*	Denominator*	Exclusion	Direction	Measure	Recommendation			Function	
24	Short interval dual energy x-ray absorptiometry scans (DXAs): short interval rate	DXA scans performed on female beneficiaries over age 66 at low risk for fracture	Number of DXAs performed per within 23 months of a previous DXA per 100 female beneficiaries aged over 66	We excluded beneficiaries not continuously enrolled in fee-for-service Medicare Parts A and B in the 23 months prior to each DXA test identified. We conservatively excluded beneficiaries who were (i) diagnosed with any cancer, except non-melanoma skin cancer (using the Clinical Classifications [17] or (ii) diagnosed with fragility fracture in the 23 months.	-	-	CW	Morden [12]	Cure	General	Imaging

Table A2. Low-value care measures including numerator, denominator, exclusion criteria, direction and measure source and reference specified by function according to the OECD/WHO/Eurostat Classification of Health Care Functions (n=115) (continued)

No	Measure	Measure details				Original source		Ref.			
		Numerator*	Denominator*	Exclusion	Direction	Measure	Recommendation			Function	Non-function
25	Short interval dual energy x-ray absorptiometry scans (DXAs): proportion of inappropriate DXAs	DXA scans performed on female beneficiaries over age 66 at low risk for fracture	The proportion of all DXAs done at a inappropriately short interval (occurring within 23 months of a previous DXA)	We excluded beneficiaries not continuously enrolled in fee-for-service Medicare Parts A and B in the 23 months prior to each DXA test identified. We conservatively excluded beneficiaries who were (i) diagnosed with any cancer, except non-melanoma skin cancer (using the Clinical Classifications [17] or (ii) diagnosed with fragility fracture in the 23 months.	-	-	CW	Morden [12]	Cure	General	Imaging

Table A2. Low-value care measures including numerator, denominator, exclusion criteria, direction and measure source and reference specified by function according to the OECD/WHO/Eurostat Classification of Health Care Functions (n=115) (continued)

No	Measure	Measure details				Original source		Ref.	Function		Non-function
		Numerator*	Denominator*	Exclusion	Direction	Measure	Recommendation				
26	Short interval dual energy x-ray absorptiometry scans (DXAs): mean intertest time interval	DXA scans performed on female beneficiaries over age 66 at low risk for fracture	The mean inter-test time interval for DXAs done within 23 months of a previous DXA.	We excluded beneficiaries not continuously enrolled in fee-for-service Medicare Parts A and B in the 23 months prior to each DXA test identified. We conservatively excluded beneficiaries who were (i) diagnosed with any cancer, except non-melanoma skin cancer (using the Clinical Classifications [17] or (ii) diagnosed with fragility fracture in the 23 months.	-	-	CW	Morden [12]	Cure	General	Imaging
27	Abdomen CT (use of contrast)	Of all abdomen CT studies performed (with contrast, without contrast, both with and without contrast-combined study), the number of abdomen CT combined studies (with and without contrast). Exclusions for specified diagnoses.		-	unclear	QualityNet	-	Chan [19]	Cure	General	Imaging

Table A2. Low-value care measures including numerator, denominator, exclusion criteria, direction and measure source and reference specified by function according to the OECD/WHO/Eurostat Classification of Health Care Functions (n=115) (continued)

No	Measure	Measure details				Original source		Ref.			
		Numerator*	Denominator*	Exclusion	Direction	Measure	Recommendation			Function	Non-function
28	Abdomen CT use of contrast material	The number of Abdomen CT studies with and without contrast ("combined studies"). CPT 74170	The number of Abdomen CT studies performed (with contrast, without contrast or both with and without contrast). CPT 74150, 74160, 74170	-	-	QualityNet	-	Segal [20]	Cure	General	Imaging
29	CT Abdomen	All combined abdominal CT's.	Total number of abdominal CTs performed with, without, or with and without contrast.	CTs performed in patients with diagnoses of hematuria, pancreatic disorders, adrenal masses, unspecified disorders of the kidney and ureter, or malignant neoplasms of the liver, bile ducts, pancreas, kidney or liver.	-	CMS / QualityNet	-	Mathias [11]	Cure	General	Imaging
30	CT Thorax	All combined thoracic CTs	Total number of thoracic CTs performed with, without, or with and without contrast.	No CTs.	-	CMS / QualityNet	-	Mathias [11]	Cure	General	Imaging
31	Thorax CT (use of contrast)	Of all thorax CT studies performed (with contrast, without contrast, both with and without contrast-combined study), the number of thorax CT combined studies (with and without contrast).		-	unclear	QualityNet	-	Chan [19]	Cure	General	Imaging

Table A2. Low-value care measures including numerator, denominator, exclusion criteria, direction and measure source and reference specified by function according to the OECD/WHO/Eurostat Classification of Health Care Functions (n=115) (continued)

No	Measure	Measure details				Original source		Ref.	Function		Non-function
		Numerator*	Denominator*	Exclusion	Direction	Measure	Recommendation				
32	Thorax CT use of contrast material	The number of thorax CT studies with and without contrast ("combined studies"). CPT 71270	The number of thorax CT studies performed (with contrast, without contrast or both with and without contrast). CPT 71250, 71260, 71270	Exclude from the denominator if on the same claim as CPT 74140, 74160, 74170 - ICD-9 code: 593.9, 120.0, 599.70, 599.71, 599.72, 251.2, 251.0, 250.8, 270.3, 255.9, 194.xx, 277.xx, 237.xx, 155.0, 155.1, 155.2, 157.0, 157.1,157.2,157.3, 157.4, 157.8, 157.9, 189.0, 211.5, 211.6, 211.7, 223.0	-	QualityNet	-	Segal [20]	Cure	General	Imaging
33	CT of the sinuses for uncomplicated acute rhinosinusitis	Maxillofacial CT study with a diagnosis of sinusitis in the imaging claim (specified with claim codes)	Patients with sinusitis diagnosis	-	-	-	CW / Lit	Schwartz [3]	Cure	General	Imaging
34	Don't order upper-tract imaging for patients with benign prostatic hyperplasia (BPH)	Beneficiaries who received an intravenous pyelogram or an abdominal CT, MRI, or ultrasound within 60 days of the index diagnosis	Male beneficiaries diagnosed with BPH over age 65 without other indications of imaging	Cancer diagnosis at any point during study period (e.g. chronic renal failure, nephritis, calculus of kidney and ureter, kidney stones abdominal pain) within 60 days of diagnosis	-	-	CW	Colla [9]	Cure	General	Imaging

Table A2. Low-value care measures including numerator, denominator, exclusion criteria, direction and measure source and reference specified by function according to the OECD/WHO/Eurostat Classification of Health Care Functions (n=115) (continued)

No	Measure	Measure details				Original source		Ref.			
		Numerator*	Denominator*	Exclusion	Direction	Measure	Recommendation			Function	Non-function
35	Fiberoptic laryngoscopy for sinusitis diagnosis	Laryngoscopy WITH ICD-9 code indicating sinusitis on the same claim	Individuals with a diagnosis of sinusitis (acute or chronic) –inpatient or outpatient	-	-	AQC	-	Segal [20]	Cure	General	Imaging
36	Nasal endoscopy for sinusitis diagnosis	Nasal endoscopy WITH ICD-9 code indicating sinusitis on the same claim	Individual with a diagnosis of sinusitis (acute or chronic) –inpatient or outpatient	-	-	AQC	-	Segal [20]	Cure	General	Imaging
37	Simultaneous use of brain CT and sinus CT	% brain CT with a sinus CT performed on the same day at the same facility. Exclusions for specified diagnoses noted in one of the diagnoses fields of the brain CT claim.		-	lower rates is better performance	QualityNet	-	Chan [19]	Cure	General	Imaging
38	EEG monitoring in individuals presenting with syncope	EEG on the same claim as diagnosis of syncope or at any time during the hospitalization with a code for syncope	Individuals with an outpatient visit with diagnosis of syncope or hospitalization for syncope	-	-	-	NICE	Segal [20]	Cure	General	Imaging
39	Head imaging in the evaluation of syncope	CT or MR imaging of the head with a diagnosis of syncope in the imaging claim (specified with claim codes)	Patients with syncope diagnosis	-	-	-	CW / NICE	Schwartz [3]	Cure	General	Imaging
40	Screening for carotid artery disease for syncope	Carotid imaging with syncope diagnosis (specified with claim codes)	Patients with syncope diagnosis	-	-	-	CW / NICE	Schwartz [3]	Cure	General	Imaging
41	Imaging for patients at low risk for pulmonary embolism (PE)	Number of hemodynamically stable patients who receive CT pulmonary angiograms for suspected PE who have either: (a) a low clinical probability of PE (determined by structured prediction tool or implicit judgment, prior to imaging), and a negative D-dimer OR (b) a low clinical probability of PE and no D-dimer performed OR (c) no pretest probability documented		-	lower rates is better performance	NQF	-	Chan [19]	Cure	General	Imaging

Table A2. Low-value care measures including numerator, denominator, exclusion criteria, direction and measure source and reference specified by function according to the OECD/WHO/Eurostat Classification of Health Care Functions (n=115) (continued)

No	Measure	Measure details				Original source		Ref.			
		Numerator*	Denominator*	Exclusion	Direction	Measure	Recommendation			Function	Non-function
42	Cardiac imaging (preoperative risk assessment for non-cardiac low-risk surgery)	number of stress echocardiography, SPECT MPI and stress MRI studies performed at the hospital outpatients department within 30 days of an ambulatory low-risk non-cardiac surgery (e.g. endoscopic, superficial, cataract) performed at any location		-	lower rates is better performance	QualityNet	-	Chan [19]	Cure	General	Imaging
43	Don't perform preoperative cardiac tests for low-risk, non-cardiac surgeries	Beneficiaries who received a non-indicated cardiac test, including stress tests, echocardiograms, electrocardiograms, CTs, MRIs or PETs within 30 days before low-risk surgery	Beneficiaries over age 65 undergoing low-risk, non-cardiac surgery (e.g. breast surgery, transurethral resection of the prostate, corneal transplant, inguinal hernia repair, lithotripsy, arthroscopy, laparoscopic cholecystectomy)	Appropriate clinical indication on testing event claim (e.g., palpitations) or admission in the 30 days before surgery	-	PQRS	CW	Colla [9]	Cure	General	Imaging
44	Preoperative stress testing	Stress electrocardiography, echocardiography or nuclear medicine imaging specified as a preoperative assessment or occurring within 30 d before a low- or intermediate-risk non-cardiothoracic surgical procedure (specified with claim codes)	Patients undergoing selected surgeries	-	-	-	CW / Lit	Schwartz [3]	Cure	General	Imaging
45	Don't perform preoperative cardiac tests for cataract surgeries	Beneficiaries who received a non-indicated cardiac test, including stress tests, echocardiograms, electrocardiograms and advanced cardiac imaging in the 30 days before cataract surgery	Beneficiaries over age 65 undergoing cataract surgery	Appropriate clinical indication on testing event claim (e.g., palpitations) or admission in the 30 days before surgery	-	PQRS	CW	Colla [9]	Cure	General	Imaging

Table A2. Low-value care measures including numerator, denominator, exclusion criteria, direction and measure source and reference specified by function according to the OECD/WHO/Eurostat Classification of Health Care Functions (n=115) (continued)

No	Measure	Measure details				Original source		Ref.			
		Numerator*	Denominator*	Exclusion	Direction	Measure	Recommendation			Function	Non-function
46	Preoperative chest radiography	Chest radiograph specified as a preoperative assessment or occurring within 30 d before a low- or intermediate risk non-cardiothoracic surgical procedure (specified with claim codes)	Patients undergoing selected surgeries	-	-	-	CADTH / CW / Lit	Schwartz [3]	Cure	General	Imaging
47	Preoperative chest radiography in the absence of a clinical suspicion for intrathoracic pathology	71010, 71020 These codes must be in a 30 day window before the anesthesia code	All patients who had anesthesia 00100-02101 (CPT)	diagnoses 466.xx, 480.xx-488.xx, 490.xx-496.xx, 500.xx-508.xx, 510.xx-519.xx	-	-	Lit	Segal [20]	Cure	General	Imaging
48	Preoperative echocardiography	Echocardiogram specified as a preoperative assessment or obtained within 30 days before a low- or intermediate-risk non-cardiothoracic surgical procedure (specified with claim codes)	Patients undergoing selected surgeries	-	-	-	CW / Lit	Schwartz [3]	Cure	General	Imaging
49	Preoperative PFT	PFT specified as a preoperative assessment or occurring within 30 d before a low- or intermediate-risk surgical procedure (specified with claim codes)	Patients undergoing selected surgeries	-	-	-	CW	Schwartz [3]	Cure	General	Imaging
50	PTH measurement for patients with stage 1-3 CKD	PTH measurement in patients with CKD (specified with claim codes)	CKD patients	-	-	-	NICE / Lit	Schwartz [3]	Cure	General	Lab

Table A2. Low-value care measures including numerator, denominator, exclusion criteria, direction and measure source and reference specified by function according to the OECD/WHO/Eurostat Classification of Health Care Functions (n=115) (continued)

No	Measure	Measure details				Original source		Ref.	Function		Non-function
		Numerator*	Denominator*	Exclusion	Direction	Measure	Recommendation				
51	Diagnostic tests, such as immunoglobulin testing, in the evaluation of allergy	Use of CPT 82701, 82784, 82785, 82787, 86005 on the same claim as a code for diagnoses in the denominator column	477.0, 477.1, 477.2, 477.8, 477.9, 493.0, 493.02, 493.9, 493.90, 493.92, 708.0, 995.3	-	-	-	CW	Segal [20]	Cure	General	Lab
52	Serological tests for Helicobacter pylori	Any code indicating testing for H. pylori	Whole population	-	-	-	NICE	Segal [20]	Cure	General	Lab
53	Hypercoagulability testing for patients with deep vein thrombosis	Laboratory tests for hypercoagulable states within 30 d after diagnosis of lower-extremity deep vein thrombosis or pulmonary embolism (specified with claim codes)	Patients with deep vein thrombosis	-	-	-	CW / Lit	Schwartz [3]	Cure	General	Lab
54	Abx for acute asthma exacerbation	Visits by adults with acute asthma exacerbation who receive any abx	Visits by adults with acute asthma exacerbation	None	-	-	NAEPP	Kale [27]	Cure	General	Pharmaceuticals

Table A2. Low-value care measures including numerator, denominator, exclusion criteria, direction and measure source and reference specified by function according to the OECD/WHO/Eurostat Classification of Health Care Functions (n=115) (continued)

No	Measure	Measure details				Original source		Ref.			
		Numerator*	Denominator*	Exclusion	Direction	Measure	Recommendation			Function	Non-function
55	Abx for acute bronchitis	Visits by adults with bronchitis who received any abx	Visits by adults with acute bronchitis	Visits by adults with HIV, cystic fibrosis, cancer, chronic bronchitis, emphysema, bronchiectasis, extrinsic allergic alveolitis, chronic airway obstruction, tuberculosis, or pneumoconioses	-	NCQA	-	Kale [27]	Cure	General	Pharmaceuticals
56	Antibiotics for acute bronchitis	% patients aged 18-64 with acute bronchitis who were not dispensed an antibiotic prescription		-	higher rates is better performance	NGC / NQF	-	Chan [19]	Cure	General	Pharmaceuticals
57	Abx for URTI	Visits by adults with uncomplicated URTI who received any antibiotic medication	Visits by adults with uncomplicated URTI	Visits by adults with HIV, COPD or cancer	-	-	ICSI	Kale [27]	Cure	General	Pharmaceuticals

Table A2. Low-value care measures including numerator, denominator, exclusion criteria, direction and measure source and reference specified by function according to the OECD/WHO/Eurostat Classification of Health Care Functions (n=115) (continued)

No	Measure	Measure details				Original source		Ref.			
		Numerator*	Denominator*	Exclusion	Direction	Measure	Recommendation			Function	Non-function
58	Abx other than nitrofurantoin, trimethoprimsulbactam, or quinolone use for UTI	Visits by female adults with uncomplicated UTI who received abx other than nitrofurantoin, trimethoprim-sulbactam, or quinolone	Visits by female adults with uncomplicated URTI	Visits by female adults with vaginitis/cervicitis, skin infections, kidney infections, STD, history of DM, cancer, pregnancy nephrolithiasis or urologic procedures	-	-	Lit	Kale [27]	Cure	General	Pharmaceuticals
59	Sinus CT or antibiotics for uncomplicated acute rhinosinusitis	Any occurrence of sinus CT (CPT 70486, 70487, 70488) in the 3 months preceding the diagnosis of acute sinusitis	461.0, 461.1, 461.2, 461.3, 461.8, 461.9 AND NO code in the preceding 3 months for any of these	code in the preceding 3 months for 473.0, 473.1, 473.2, 473.3, 473.8, 473.	-	-	CW	Segal [20]	Cure	General	Pharmaceuticals
60	Medication use for urinary incontinence	% female patients ≥ 65 with urinary incontinence who were prescribed a medication to treat the urinary incontinence and who had a trial of behavioral therapy (e.g. bladder training, pelvic floor muscle training, prompted voiding) documented		-	higher rates is better performance	NGC	-	Chan [19]	Cure	General	Pharmaceuticals
61	Don't use opioid or butalbital treatment for migraine, except as a last resort	Beneficiaries who filled an opioid or butalbital prescription within 21 days of the office visit with migraine diagnosis	Beneficiaries over age 65 with a diagnosed migraine and no other indication for opioids	An 'E' code, inpatient admission, back pain, abdominal pain, surgery, fracture, cancer or hospice enrollment within 60 days of index visit	-	-	CW	Colla [9]	Cure	General	Pharmaceuticals

Table A2. Low-value care measures including numerator, denominator, exclusion criteria, direction and measure source and reference specified by function according to the OECD/WHO/Eurostat Classification of Health Care Functions (n=115) (continued)

No	Measure	Measure details				Original source		Ref.	Function		Non-function
		Numerator*	Denominator*	Exclusion	Direction	Measure	Recommendation				
62	Antibiotics for pharyngitis without strep confirmation	% patients aged 2-18 diagnosed with pharyngistis who were dispensed an antibiotic and received Group A streptococcustest		-	higher rates is better performance	NGC / NQF	-	Chan [19]	Cure	General	Pharmaceuticals
63	Antibiotics for URI	% patients aged 3 months to 18 with URI who were not dispensed an antibiotic prescription		-	higher rates is better performance	NGC / NQF	-	Chan [19]	Cure	General	Pharmaceuticals
64	Antihistamines or decongestants for OME	% patients aged 2 months to 12 years with OME who were not prescribed or recommended to receive either antihistamines or decongestants except with documented medical rationale.		-	higher rates is better performance	NGC / NQF	-	Chan [19]	Cure	General	Pharmaceuticals
65	Systemic antimicrobial therapy for AOE	% patients aged ≥ 2 with AOE who were not prescribed systemic antimicrobial therapy except with documented medical rationale.		-	higher rates is better performance	NGC / NQF	-	Chan [19]	Cure	General	Pharmaceuticals

Table A2. Low-value care measures including numerator, denominator, exclusion criteria, direction and measure source and reference specified by function according to the OECD/WHO/Eurostat Classification of Health Care Functions (n=115) (continued)

No	Measure	Measure details				Original source		Ref.	Function		Non-function
		Numerator*	Denominator*	Exclusion	Direction	Measure	Recommendation				
66	Systemic antimicrobial therapy for OME	% patients aged 2 months to 12 years with OME who were not prescribed systemic antimicrobials except with documented medical rationale.		-	higher rates is better performance	NGC / NQF	-	Chan [19]	Cure	General	Pharmaceuticals
67	Systemic corticosteroids for OME	% patients aged 2 months to 12 years with OME who were not prescribed systemic corticosteroid except with documented medical rationale.		-	higher rates is better performance	NGC / NQF	-	Chan [19]	Cure	General	Pharmaceuticals
68	Intravenous unfractionated heparin	% patients ≥ 18 with diagnosis of ischemic stroke who did not receive heparin. Patients undergoing carotid endarterectomy, carotid angioplasty-stenting, or had documented medical reasons for receiving heparin excluded		-	-	NGC	-	Chan [19]	Cure	General	Pharmaceuticals
69	Chronic wound care	% visits for patients ≥ 18 with a diagnosis of chronic skin ulcer without the use of a wound surface culture technique.		-	higher rates is better performance	NGC	-	Chan [19]	Cure	General	-
70	Chronic wound care	% visits for patients ≥ 18 with a diagnosis of chronic skin ulcer without a prescription or recommendation to use wet to dry dressings		-	higher rates is better performance	NGC	-	Chan [19]	Cure	General	-
71	More than 1 emergency department visit in last 30 days of life	More than 2 visits with location code or CPT code indicating ED use within 30 days before death	Individuals with death during our observation period	-	-	NQF	-	Segal [20]	Cure	General	-

Table A2. Low-value care measures including numerator, denominator, exclusion criteria, direction and measure source and reference specified by function according to the OECD/WHO/Eurostat Classification of Health Care Functions (n=115) (continued)

No	Measure	Measure details				Original source		Ref.	Function		Non-function
		Numerator*	Denominator*	Exclusion	Direction	Measure	Recommendation				
72	Caesarean section	Patients with ICD-9-CM Procedure Code for cesarean section (numerator);	nulliparous patients delivered of a live term singleton newborn in vertex presentation (denominator);	contraindications to vaginal delivery, age <8 or ≥ 65, length of stay >120 days, clinical trial enrollment (excluded)	lower rates is better performance	NQF	-	Chan [19]	Cure	General	-
73	Elective delivery	% patients with elective vaginal deliveries or elective cesarean sections among all patients delivering newborns at ≥ 37 and < 39 weeks of gestation completed. Patients with diagnosis justifying elective delivery are excluded.		-	lower rates is better performance	NQF	-	Chan [19]	Cure	General	-
74	Epidural steroid injections for back pain	% patients with back pain who received an epidural steroid injection in the absence of radicular pain AND those patients with radicular pain who received an epidural steroid injection without image guidance		-	lower rates is better performance	NQF	-	Chan [19]	Cure	General	-
75	Surgical timing for back pain	% patients with documentation of red flags who had surgery within the first 6 weeks of back pain onset.		-	lower rates is better performance	NQF	-	Chan [19]	Cure	General	-
76	Vertebroplasty or kyphoplasty for osteoporotic vertebral fractures	Vertebroplasty/kyphoplasty for vertebral fracture (specified with claim codes)	Patients with osteoporosis	-	-	-	Lit	Schwartz [3]	Cure	General	-
77	Arthroscopic surgery for knee osteoarthritis	Arthroscopic debridement/ chondroplasty of the knee (specified with claim codes)	Patients with arthritis	-	-	-	NICE / Lit	Schwartz [3]	Cure	General	-
78	Traction for low back pain	Traction with diagnosis of low back pain	Low back pain diagnosis	-	-	-	IOM	Segal [20]	Cure	General	-

Table A2. Low-value care measures including numerator, denominator, exclusion criteria, direction and measure source and reference specified by function according to the OECD/WHO/Eurostat Classification of Health Care Functions (n=115) (continued)

No	Measure	Measure details				Original source		Ref.			
		Numerator*	Denominator*	Exclusion	Direction	Measure	Recommendation			Function	Non-function
79	Laminectomy or spinal fusion	Laminectomy or spinal fusion	Everyone	those with a clear indication (radicular symptoms* and we will be liberal with this) *symptoms clearly of herniated disc—radicular pain	-	NGC	-	Segal [20]	Cure	General	-
80	Carotid endarterectomy in asymptomatic patients	Carotid endarterectomy for patients without a history of stroke or TIA and without stroke, TIA, or focal neurological symptoms noted in claim (specified with claim codes)	All patients	-	-	-	CW / Lit	Schwartz [3]	Cure	General	-
81	IVC filters to prevent pulmonary embolism	Any IVC filter placement (specified with claim codes)	All patients	-	-	-	Lit	Schwartz [3]	Cure	General	-

Table A2. Low-value care measures including numerator, denominator, exclusion criteria, direction and measure source and reference specified by function according to the OECD/WHO/Eurostat Classification of Health Care Functions (n=115) (continued)

No	Measure	Measure details				Original source		Ref.	Function		Non-function
		Numerator*	Denominator*	Exclusion	Direction	Measure	Recommendation				
82	Percutaneous coronary intervention with balloon angioplasty or stent placement for stable coronary disease	Coronary stent placement or balloon angioplasty for patients with an established diagnosis of ischemic heart disease or angina (≥6 months before the procedure); procedure not associated with an ED visit, which might be indicative of acute coronary syndrome (specified with claim codes)	IHD patients	-	-	-	Lit	Schwartz [3]	Cure	General	-
83	Renal artery angioplasty or stenting	Renal/visceral angioplasty or stent placement (specified with claim codes)	Patients with hypertension	-	-	-	Lit	Schwartz [3]	Cure	General	-
84	Routine monitoring of digoxin in patients with congestive heart failure	Any measure of digoxin with no hospitalizations or ER visits during that year.	All patients* with CHF *will include atrial fibrillation patients as well	-	-	-	NICE	Segal [20]	Cure	Specialized	Lab
85	Homocysteine testing for cardiovascular disease	Homocysteine testing (specified with claim codes)	All patients	No diagnosis for folate or B12 deficiencies in claim and no folate or B12 testing in prior claims	-	-	Lit	Schwartz [3]	Cure	General	Lab

Table A2. Low-value care measures including numerator, denominator, exclusion criteria, direction and measure source and reference specified by function according to the OECD/WHO/Eurostat Classification of Health Care Functions (n=115) (continued)

No	Measure	Measure details				Original source		Ref.	Non-function	-
		Numerator*	Denominator*	Exclusion	Direction	Measure	Recommendation		Function	General
										Cure
86	Chronic wound care	% patients ≥ 18 with a diagnosis of chronic skin ulcer undergoing debridement with documentation of wound characteristics (including size, nature of wound base tissue, and amount of drainage) prior to debridement.	-	-	higher rates is better performance	NGC	-	Chan [19]		
87	Hysterectomy for benign disease	Any hysterectomy (not specified for malignancy treatment)	All women	those with a malignancy diagnosis	-	NGC	-	Segal [20]	-	Specialized
										Cure
88	Don't recommend percutaneous feeding tubes in patients with advanced dementia	Beneficiaries with two observed dementia diagnoses residing in an institution who received a feeding tube	Institutionalized beneficiaries over age 65 with diagnosed dementia	None	-	-	CW	Colla [9]	-	Inpatient
										LTC
89	Inappropriate medications in the elderly	Visits by adults aged ≥ 65 y who received any of 33 potentially inappropriate medications	Visits by adults aged ≥65 y with reported medications	Visits by adults aged ≥65 with diabetes	-	-	Lit	Kale [27]	Pharmaceuticals	Outpatient
										LTC

Table A2. Low-value care measures including numerator, denominator, exclusion criteria, direction and measure source and reference specified by function according to the OECD/WHO/Eurostat Classification of Health Care Functions (n=115) (continued)

No	Measure	Measure details				Original source		Ref.	Function		Non-function
		Numerator*	Denominator*	Exclusion	Direction	Measure	Recommendation				
90	Don't use antipsychotics as first choice to treat behavioral and psychological symptoms of dementia	Beneficiaries who received one or more prescriptions for an antipsychotic following two observed dementia diagnoses	Beneficiaries over age 65 with diagnosed dementia	Severe mental illness during the study period	-	-	CW	Colla [9]	LTC	Outpatient	Pharmaceuticals
91	Cardiac stress imaging (testing in asymptomatic, low-risk patients)	% of all stress SPECT MPI, stress echo, CCTA, and CMR performed in asymptomatic, low CHD risk patients for initial detection and risk assessment		-	lower rates is better performance	NQF	-	Chan [19]	Preventive	-	Imaging
92	Don't order cardiac tests on low-risk, asymptomatic patients	Beneficiaries who received a non-indicated cardiac test, echocardiograms, electrocardiograms, advanced cardiac imaging	Low-risk beneficiaries ages 66-80	Indications of cardiac disease or other conditions that could indicate cardiac testing (e.g. HIV/aids, diabetes, peripheral vascular disease, pulmonary disease, cancer) or use of a prescription drug associated with the above conditions in a calendar year; enrollment in hospice; appropriate clinical indication on testing event claim.	-	-	CW	Colla [9]	Preventive	-	Imaging

Table A2. Low-value care measures including numerator, denominator, exclusion criteria, direction and measure source and reference specified by function according to the OECD/WHO/Eurostat Classification of Health Care Functions (n=115) (continued)

No	Measure	Measure details				Original source		Ref.	Function	Non-function
		Numerator*	Denominator*	Exclusion	Direction	Measure	Recommendation			
93	Endoscopy and polyp surveillance	% patients ≥ 50 receiving a surveillance colonoscopy, with a history of a colonic polyp in previous colonoscopy who had a follow-up interval of > 3 years since their last colonoscopy		-	higher rates is better performance	NGC / NQF	-	Chan [19]	Preventive	Imaging
94	Endoscopy and polyp surveillance	% patients ≥ 50 receiving a screening colonoscopy without biopsy or polypectomy who had a recommend follow-up interval of at least 10 years for repeat colonoscopy.		-	higher rates is better performance	NGC / NQF	-	Chan [19]	Preventive	Imaging
95	Colorectal cancer screening for older patients	Colorectoral cancer screening (colonoscopy, sigmoidoscopy, barium enema, or fecal occult blood testing) for patients aged over 75 y (specified with claim codes)	Patients over 75	-	-	-	USPTF / Lit	Schwartz [3]	Preventive	Imaging
96	Imaging studies in melanoma	% patients with stage 0 or 1A melanoma, without signs or symptoms, for whom no diagnostic imaging studies were ordered		-	higher rates is better performance	NGC / NQF	-	Chan [19]	Preventive	Imaging
97	Bone scan for staging low-risk patients	% patients with a diagnosis of prostate cancer at low risk of recurrence receiving interstitial prostate brachytherapy, OR external beam radiotherapy to the prostate, OR radical prostatectomy, OR cryotherapy who did not have a bone scan performed at any time since the diagnosis of prostate cancer.		-	higher rates is better performance	NGC / NQF	-	Chan [19]	Preventive	Imaging

Table A2. Low-value care measures including numerator, denominator, exclusion criteria, direction and measure source and reference specified by function according to the OECD/WHO/Eurostat Classification of Health Care Functions (n=115) (continued)

No	Measure	Measure details				Original source		Ref.	Function	Non-function
		Numerator*	Denominator*	Exclusion	Direction	Measure	Recommendation			
98	PET, CT, and radionuclide bone scan in individuals with low-risk prostate cancer	PET, CT, or radionuclide bone scan AFTER diagnosis	Men with low risk for prostate CA	-	-	-	CW	Segal [20]	Preventive	Imaging
99	Mammography follow-up rates	All patients undergoing diagnostic mammography or breast ultrasound study within 45 days of screening mammography.	All patients receiving a screening mammography study.	No patients.	-	CMS / QualityNet	-	Mathias [11]	Preventive	Imaging
100	Mammography follow-up rates	Number of patients who had a diagnostic mammography or an ultrasound of the breast following a screening mammography study within 45 days.		-	unclear	QualityNet	-	Chan [19]	Preventive	Imaging
101	Mammography screening for women aged ≥75 y	Visits by women ages ≥ 75 yr who received a mammogram	Visits by women aged ≥75 y	Visits by women aged ≥75 y with history of breast cancer, breast mass or lump	-	-	UPSTF	Kale [27]	Preventive	Imaging
102	Screening for asymptomatic carotid artery stenosis in the general adult population	CPT 93880 or 3100F, ONLY IN outpatient setting (not ER)	All people	ICD-9 codes for: 785.9, 784.2, 362.34, 435.9, 433.10, 342.90, 780.2, 781.3, 437.0	-	-	UPSTF	Segal [20]	Preventive	Imaging

Table A2. Low-value care measures including numerator, denominator, exclusion criteria, direction and measure source and reference specified by function according to the OECD/WHO/Eurostat Classification of Health Care Functions (n=115) (continued)

No	Measure	Measure details				Original source		Ref.	Function	Non-function
		Numerator*	Denominator*	Exclusion	Direction	Measure	Recommendation			
103	Screening for carotid artery disease in asymptomatic adults	Carotid imaging for patients without a history of stroke or TIA and without a diagnosis of stroke, TIA, or focal neurological symptoms in claim (specified with claim codes)	All patients	-	-	-	CW/ USPTF / Lit	Schwartz [3]	Preventive	- Imaging
104	Screening CME in adults in GME	Visits by adults who present for GME and are ordered a CBC	Visits by adults who present for GME	Visits by adults with cancer, hematologic abnormalities	-	-	UPSTF	Kale [27]	Preventive	- Imaging
105	Screening ECG in adults in GME	Visits by adults who present for GME and are ordered an ECG	Visits by adults who present for GME	Visits by adults with CAD, arrhytmia, chest pain, HTN, palpitations, dyspnea, or syncope	-	-	UPSTF	Kale [27]	Preventive	- Imaging
106	Screening UA in adults in GME	Visits by adult men and non-pregnant women who present for GME and are ordered a UA	Visits by adults who present for GME	Visits by adults with urologic disease, pregnancy, or diseases of genital organs	-	-	UPSTF	Kale [27]	Preventive	- Imaging

Table A2. Low-value care measures including numerator, denominator, exclusion criteria, direction and measure source and reference specified by function according to the OECD/WHO/Eurostat Classification of Health Care Functions (n=115) (continued)

No	Measure	Measure details				Original source		Ref.	Function	Non-function
		Numerator*	Denominator*	Exclusion	Direction	Measure	Recommendation			
107	Screening x-ray in adults in GME	Visits by adults who present for GME and are ordered a chest x-ray	Visits by adults who present for GME	None	-	-	UPSTF	Kale [27]	Preventive	- Imaging
108	PSA testing for men aged ≥75 y	PSA test for patients aged over 75 y (specified with claim codes)	Men over 75	-	-	-	USPTF	Schwartz [3]	Preventive	- Lab
109	Prostate cancer screening in men aged >75 y	Visits by men aged ≥75 yr who are ordered a PSA	Visits by men aged ≥75 y	Visits by adult men aged ≥75 y with prostate cancer	-	-	UPSTF	Kale [27]	Preventive	- Lab
110	Don't screen women older than 65 years of age for cervical cancer who have had adequate prior screening and are not otherwise at high risk for cervical cancer	Beneficiaries who received a Pap test	Female beneficiaries at low risk for cervical cancer over age 65	Gynaecological cancers, HIV/aids, diethylstilbestrol use, or a previous Pap test during the study period	-	NQF	CW	Colla [9]	Preventive	- Lab
111	Cervical cancer screening in woman aged >65 y	Visits by women ≥ 65 yr who were ordered a Pap test	Visits by women aged ≥65 y	Visits by women aged ≥65 y with cervical cancer, uterine cancer, cervical dysplasia, or vaginal bleeding	-	-	UPSTF	Kale [27]	Preventive	- Lab

Table A2. Low-value care measures including numerator, denominator, exclusion criteria, direction and measure source and reference specified by function according to the OECD/WHO/Eurostat Classification of Health Care Functions (n=115) (continued)

No	Measure	Measure details				Original source		Ref.			
		Numerator*	Denominator*	Exclusion	Direction	Measure	Recommendation			Function	Non-function
112	Cervical cancer screening for women aged ≥65 y	Screening papanicolaou test for women ages over 65 y (specified with claim codes)	Women over 65	-	-	-	CW / USPTF	Schwartz [3]	Preventive	-	Lab
113	Follow-up tumor marker studies in asymptomatic women with previously treated breast cancer	82378 (CEA), 86300 (CA 15-3) (CA 27.29)	Breast cancer is 174.0-174.9	-	-	-	Lit	Segal [20]	Preventive	-	Lab
114	Don't perform population based screening for 25-OH-Vitamin D deficiency	Beneficiaries who received a test for vitamin D deficiency	Low-risk beneficiaries over age 65	Beneficiaries with osteoporosis, fragility fracture, kidney disease, renal dialysis during the same calendar year	-	-	CW	Colla [9]	Preventive	-	
115	Cancer screening for patients with CKD receiving dialysis	Screening for cancer of the breast, cervix, colon, or prostate for patients with CKD receiving dialysis services. (specified with claim codes)	Patients with CKD	-	-	-	CW / Lit	Schwartz [3]	Preventive	-	Lab/Im

Abx: Antibiotics; NGC: National Guideline Clearing House; CMS: Centers for Medicare & Medicaid Services; NQF: National Quality Forum; Lit: Literature; AQC: Alternative Quality Contract (Blue Cross Blue Shield); NICE: National Institute for Clinical Excellence (UK); CADTH: Canadian Agency for Drugs and Technologies in Health; NAEPP: National Asthma Education and Prevention Program; NCQA: National Committee for Quality Assurance; ICSI: Institute for Clinical Systems Improvement; IOM: Institute of Medicine; USPSTF: US Preventive Services Task Force; *if denominator and numerator were not reported separately in the included paper, we reported them accordingly.

A3. Low-value care recommendations

ICHA-HC			First author	Low-value care recommendation
Alternative	-	-	Elshaug	Acupuncture for Bell's palsy
Alternative	-	-	Elshaug	Acupuncture for depression
Alternative	-	-	Elshaug	Acupuncture for induction of labor
Alternative	-	-	Elshaug	Acupuncture for irritable bowel syndrome (IBS)
Alternative	-	-	Elshaug	Acupuncture for lower urinary tract symptoms (LUTS) in men
Alternative	-	-	Elshaug	Acupuncture for peripheral joint osteoarthritis
Alternative	-	-	Elshaug	Acupuncture for the management of otitis media with effusion (OME)
Alternative	-	-	Elshaug	Acupuncture for uterine fibroids
Alternative	-	-	Elshaug	Acupuncture to treat hyperbilirubinemia
Alternative	-	-	Elshaug	Acupuncture, acupressure and hypnosis for women in labour
Alternative	-	-	Elshaug	Laser acupuncture for carpal tunnel syndrome
Cure	Dental	-	Elshaug	Emergency pulpectomy
Cure	Dental	-	Elshaug	Porcelain dental crowns
Cure	Dental	-	Elshaug	Occlusal adjustment for temporomandibular joint dysfunction
Cure	General	-	Chan	More than one ED visit in the last 30 days of life.
Cure	General	-	Chan	Routine labor induction
Cure	General	-	Chan	Referring OME early in the course of the problem.
Cure	General	-	Chan	Potentially preventable ED visits.
Cure	General	-	Chan	Routine epidural analgesia.
Cure	General	-	Chan	Routine fetal movement counting.
Cure	General	-	Chan	Routine vaginal examination to assess gestational age, predict preterm birth, or estimate tight passage during birth.
Cure	General	-	Chan	Rupturing membranes (amniotomy) after the start of spontaneous labor whether labor is progressing well or prolonged.
Cure	General	-	Elshaug	Caesarean section without medical indication
Cure	General	-	Elshaug	Vertebroplasty for painful osteoporotic vertebral factures
Cure	General	-	Elshaug	Open surgery for carotid occlusive disease
Cure	General	-	Elshaug	Vena Caval Filters for the prevention of pulmonary embolism

ICHA-HC			First author	Low-value care recommendation
Cure	General	-	Elshaug	Endovascular repair of infrarenal abdominal aortic aneurysms
Cure	General	-	Elshaug	Routine episiotomy associated with spontaneous vaginal birth
Cure	General	-	Elshaug	Routine episiotomy associated with vaginal birth following previous third- or fourth degree trauma
Cure	General	-	Elshaug	Routine dilation and curettage for missed abortion
Cure	General	-	Elshaug	Dilatation and curettage as a diagnostic tool OR therapeutic treatment
Cure	General	-	Elshaug	Amnioinfusion for the treatment of women with meconiumstained liquor
Cure	General	-	Elshaug	Chest physiotherapy as an adjunctive treatment for adults with pneumonia
Cure	General	-	Elshaug	Hospitalization for bed rest in multiple pregnancy
Cure	General	-	Elshaug	Neonatal circumcision
Cure	General	-	Elshaug	Urinary flow-rate measurement in men with lower urinary tract symptoms (LUTS).
Cure	General	-	Hicks	Do not use inferior vena cava filters routinely in patients with acute venous thromboembolism
Cure	General	-	Keyhani	Carotid endarterectomy for all indications
Cure	General	-	Korenstein	Carotid endarterectomy for carotid stenosis
Cure	General	-	Korenstein	IVC filter
Cure	General	-	Korenstein	Chiropractic for low back pain
Cure	General	-	Korenstein	Nasopharyngeal washings
Cure	General	Imaging	Chan	Imaging studies in patients with nonspecific low back pain and no red flags.
Cure	General	Imaging	Chan	Performing an imaging stress test as the initial diagnostic test in patients with known or suspected CAD who are able to exercise and have no resting ECG abnormalities that may interfere with interpretation of test results.
Cure	General	Imaging	Chan	Coronary angiography in patients with chronic stable angina with well-controlled symptoms on medical therapy or who lack specific high-risk criteria on exercise testing.
Cure	General	Imaging	Chan	Obtaining diagnostic images for minor head injuries without loss of consciousness or other risk factors.
Cure	General	Imaging	Chan	Imaging for uncomplicated head ache.
Cure	General	Imaging	Chan	Imaging studies in patients with recurrent, classic migraine headache and normal findings on neurological examination.

ICHA-HC			First author	Low-value care recommendation
Cure	General	Imaging	Chan	Repeat CT scans in patients with functional abdominal pain syndrome if no major changes in clinical findings or symptoms.
Cure	General	Imaging	Chan	Sinus imaging for patients with acute rhinosinusitis in absence of predisposing factors for atypical microbial causes.
Cure	General	Imaging	Chan	CT or MRI to evaluate simple syncope in patients with normal findings on neurologic examination.
Cure	General	Imaging	Chan	Routinely performing ECHO in the evaluation of syncope, unless the history, physical examination, and ECG do not provide a diagnosis or underlying heart disease is suspected.
Cure	General	Imaging	Chan	Pre-op chest radiography in the absence of a clinical suspicion for intrathoracic pathology.
Cure	General	Imaging	Chan	Performing radionuclide imaging as part of routine follow-up in asymptomatic patients.
Cure	General	Imaging	Chan	Routine periodic ECHO in asymptomatic patients with mild aortic stenosis for more frequently than every 3-5 years.
Cure	General	Imaging	Chan	Routinely repeat ECHO in asymptomatic patients with mild mitral regurgitation and normal left ventricular size and function.
Cure	General	Imaging	Chan	Performing ECHO in asymptomatic patients with innocent-sounding heart murmurs.
Cure	General	Imaging	Chan	Electronical fetal monitoring during labor without access to fetal scalp sampling or continuous electronic fetal monitoring.
Cure	General	Imaging	Chan	Follow-up imaging for clinically inconsequential adnexal cysts.
Cure	General	Imaging	Chan	Home uterine activity monitoring to prevent preterm birth.
Cure	General	Imaging	Chan	Routine ultrasound after 24 weeks.
Cure	General	Imaging	Chan	Routine ultrasound to estimate fetal size if large baby is suspected.
Cure	General	Imaging	Elshaug	Imaging in cases of low back pain*
Cure	General	Imaging	Elshaug	Exercise electrocardiogram (ECG) for angina
Cure	General	Imaging	Elshaug	Plan x-rays of the skull for diagnosing significant brain injury
Cure	General	Imaging	Elshaug	Routine monitoring of bone mineral density after starting bisphosphonate treatment
Cure	General	Imaging	Elshaug	Cardiac stress testing on low risk patients before major surgery
Cure	General	Imaging	Elshaug	Preoperative chest x-ray

ICHA-HC			First author	Low-value care recommendation
Cure	General	Imaging	Elshaug	Cardiotocography for antepartum fetal assessment / antenatal cardiotocography for fetal assessment
Cure	General	Imaging	Elshaug	The routine anomaly scan (at 18 weeks 0 days to 20 weeks 6 days) for Down's syndrome screening using soft markers
Cure	General	Imaging	Korenstein	Imaging for low back pain
Cure	General	Imaging	Korenstein	PT for low back pain
Cure	General	Imaging	Korenstein	Preoperative pulmonary function testing
Cure	General	Imaging	Onuoha	Preoperative baseline diagnostic cardiac testing (TTE or TEE) or cardiac stress test in asymptomatic stable patients with known cardiac disease (e.g. CAD, vulvular disease) undergoing low-risk or moderate-risk non-cardiac surgery
Cure	General	Imaging	Schuur	Do not order magnetic resonance imaging of the lumbar spine for patients with lower back pain without high-risk features.
Cure	General	Imaging	Schuur	Do not order CT of the head for patients with mild traumatic head injury who do not meet New Orleans Criteria or Canadian CT Head Rule.
Cure	General	Imaging	Schuur	Do not order CT to diagnose pulmonary embolism without first risk stratifying for pulmonary embolism (pretest probability and D-dimer tests if low probability).
Cure	General	Imaging	Wood	Patients Who Have No Cardiac History and Good Functional Status Do Not Require Preoperative Stress Testing Before Non-cardiac Thoracic Surgery
Cure	General	Imaging	Wood	Before Cardiac Surgery There Is No Need for Pulmonary Function Testing in the Absence of Respiratory Symptoms
Cure	General	Lab	Bulger	Do not perform repetitive complete blood count (CBC) and chemistry testing in the face of clinical and lab stability.
Cure	General	Lab	Chan	Routine chlamydia screening.
Cure	General	Lab	Chan	Routine hepatitis C screening.
Cure	General	Lab	Chan	Annual lipid screening for patients without lipid-lowering drug, diet therapy, or reasons for changing lipid profiles.
Cure	General	Lab	Chan	Broad spectrum testing rather than focus on likely source.
Cure	General	Lab	Chan	Repeat blood tests - creatinine within 10 days.
Cure	General	Lab	Chan	Repeat blood tests - ferritin within 6 weeks.
Cure	General	Lab	Chan	Repeat blood tests - HDL cholesterol within 6 weeks.

ICHA-HC			First author	Low-value care recommendation
Cure	General	Lab	Chan	Repeat blood tests - hemoglobin A1c within 12 weeks
Cure	General	Lab	Chan	Repeat blood tests - hemoglobin within 10 days.
Cure	General	Lab	Chan	Repeat blood tests - liver function tests (ALT/AST) within 6 weeks.
Cure	General	Lab	Chan	Repeat blood tests - sodium within 10 days.
Cure	General	Lab	Chan	Repeat blood tests - Thyroid stimulating hormone within 6 weeks.
Cure	General	Lab	Chan	Repeat blood tests - total cholesterol within 6 weeks.
Cure	General	Lab	Chan	Routine bacterial vaginosis screening.
Cure	General	Lab	Chan	Routine preterm labor screening.
Cure	General	Lab	Chan	Routine toxoplasmosis screening.
Cure	General	Lab	Chan	Serologic testing for Lyme disease in patients with chronic nonspecific symptoms and no clinical evidence of disseminated Lyme disease.
Cure	General	Lab	Chan	Serologic testing for suspected early Lyme disease.
Cure	General	Lab	Chan	Unnecessary laboratory tests, targeting panels (e.g. thyroid, SMA 20).
Cure	General	Lab	Chan	Unnecessary laboratory tests, targeting special testing (e.g. Lyme disease with regional considerations).
Cure	General	Lab	Elshaug	Measurement of parathyroid hormone (PTH) levels in people with stage 1, 2, 3A or 3B chronic kidney disease (CKD)
Cure	General	Lab	Elshaug	Chlamydia screening in routine antenatal care
Cure	General	Lab	Elshaug	Screening for hepatitis C virus in pregnant women
Cure	General	Lab	Elshaug	Biochemical tests of placental function for assessment in pregnancy
Cure	General	Lab	Elshaug	Blood biochemical testing in children with dehydration
Cure	General	Lab	Elshaug	C-reactive protein tests
Cure	General	Lab	Elshaug	Factor V Leiden, thrombophilia genetic mutations
Cure	General	Lab	Elshaug	Female hormone testing in women with heavy menstrual bleeding (HMB)
Cure	General	Lab	Elshaug	Fetal blood sample (FBS) with evidence of acute fetal compromise
Cure	General	Lab	Elshaug	Genetic testing of fragile X syndrome - population screen

ICHA-HC			First author	Low-value care recommendation
Cure	General	Lab	Elshaug	Human leukocyte antigen (HLA) DQ2/ DQ8 testing in the initial diagnosis of coeliac disease
Cure	General	Lab	Elshaug	Immunoglobulin G / A (IgG/IgA) anti-gliadin antibody (AGA) test in the diagnosis of coeliac disease.
Cure	General	Lab	Elshaug	Liver function tests - Statin therapy
Cure	General	Lab	Elshaug	Measurement of calcium levels in people with stage 1, 2, 3A or 3B chronic kidney disease (CKD)
Cure	General	Lab	Elshaug	Measurement of phosphate levels in people with stage 1, 2, 3A or 3B chronic kidney disease (CKD)
Cure	General	Lab	Elshaug	Microscopy for testing for the presence of hematuria
Cure	General	Lab	Elshaug	Nucleic acid amplification tests for diagnosis of Neisseria gonorrhea and Chlamydia trachomatis rectal infections
Cure	General	Lab	Elshaug	Preimplantation genetic screening for aneuploidy
Cure	General	Lab	Elshaug	Rectal biopsy in suspected Hirschsprungs disease
Cure	General	Lab	Elshaug	Routine blood tests in children with fever
Cure	General	Lab	Elshaug	Routine screening for preterm labour
Cure	General	Lab	Elshaug	Screening for gestational diabetes using fasting plasma glucose, random blood glucose, glucose challenge test and urinalysis for glucose.
Cure	General	Lab	Elshaug	Serum cholesterol concentrations in pregnancy
Cure	General	Lab	Elshaug	Serum ferritin tests in adults (in patients with chronic fatigue syndrome)
Cure	General	Lab	Elshaug	Serum ferritin tests in women with heavy menstrual bleeding
Cure	General	Lab	Elshaug	Testing for diarrhea in children
Cure	General	Lab	Elshaug	Tests for folate levels in patients with chronic fatigue syndrome
Cure	General	Lab	Elshaug	Tests for vitamin B12 deficiency in patients with chronic fatigue syndrome
Cure	General	Lab	Elshaug	Troponin levels in acute pulmonary embolism patients
Cure	General	Lab	Elshaug	Umbilical cord blood direct antiglobulin test (DAT) (Coombs' test) to predict significant hyperbilirubinemia
Cure	General	Lab	Elshaug	Urinary protein measurement in pregnant woman as a predictor of complications of pre-eclampsia

ICHA-HC			First author	Low-value care recommendation
Cure	General	Lab	Elshaug	Urine testing in infants and children for urinary tract infection (UTI)
Cure	General	Lab	Elshaug	Vertebral biopsy
Cure	General	Lab	Rouster-Stevens	Do not test for Lyme disease as a cause of musculoskeletal symptoms without an exposure history and appropriate examination findings.
Cure	General	Lab	Schuur	Do not order coagulation studies for patients without hemorrhage or suspected coagulopathy (e.g., with anticoagulation therapy, clinical coagulopathy).
Cure	General	Pharmaceuticals	AGS Choosing Wisely Workgroup	Don't use antimicrobials to treat bacteriuria in older adults unless specific urinary tract symptoms are present.
Cure	General	Pharmaceuticals	Bulger	Do not prescribe medications for stress ulcer prophylaxis to medical inpatients unless at high risk for GI complications.
Cure	General	Pharmaceuticals	Chan	Routinely prescribe antibiotics for acute mild to moderate sinusitis unless symptoms last for 7 or more days OR symptoms worsen after initial clinical improvement.
Cure	General	Pharmaceuticals	Chan	Using non-generic statins when initiating lipid-lowering drug therapy.
Cure	General	Pharmaceuticals	Chan	Use of brand over generic Rx for bronchitis, hyperlipidemia, hypo-functioning thyroid gland, ischemic heart disease.
Cure	General	Pharmaceuticals	Chan	Diethylstilbestrol to prevent miscarriage.
Cure	General	Pharmaceuticals	Chan	Fenfluramine plus phentermine to treat obesity.
Cure	General	Pharmaceuticals	Chan	Long-term acid suppression therapy (proton pump inhibitors or histamine-2 receptor antagonists) should be titrated to the lowest effective dose needed to achieve therapeutic goals.
Cure	General	Pharmaceuticals	Chan	Routine iron supplementation.
Cure	General	Pharmaceuticals	Chan	SSRIs in patients with migraine or tension-type headaches..
Cure	General	Pharmaceuticals	Chan	Thalidomide for sedation in pregnant women.
Cure	General	Pharmaceuticals	Chan	Triparanol (MER-29) for cholesterol reduction.
Cure	General	Pharmaceuticals	Chan	NSAIDS in individuals with HTN or heart failure or CKD of all causes, including diabetes.
Cure	General	Pharmaceuticals	Keyhani	Antibiotics for acute respiratory tract infections
Cure	General	Pharmaceuticals	Keyhani	Antibiotics for URI use of diagnostic tests
Cure	General	Pharmaceuticals	Korenstein	Antibiotics for URI, acute bronchitis

ICHA-HC			First author	Low-value care recommendation
Cure	General	Pharmaceuticals	Korenstein	Acid blockers
Cure	General	Pharmaceuticals	Korenstein	Bronchodilators for bronchiolitis obstructive diseases
Cure	General	Pharmaceuticals	Quinonez	Do not use bronchodilators in children with bronchiolitis.
Cure	General	Pharmaceuticals	Quinonez	Do not use systemic corticosteroids in children under 2 years of age with a lower respiratory tract infection.
Cure	General	Pharmaceuticals	Quinonez	Do not treat gastroesophageal reflux in infants routinely with acid suppression therapy.
Cure	General	Pharmaceuticals	Wiener	Do not routinely offer pharmacologic treatment with advanced vasoactive agents approved only for the management of pulmonary arterial hypertension to patients with PH resulting from left heart disease or hypoxemic lung diseases (group II or III PH).
Cure	General	Pharmaceuticals	Williams	Avoid nonsteroidal anti-inflammatory drugs (NSAIDS) in individuals with hypertension, heart failure, or CKD of all causes, including diabetes.
Cure	Specialized	-	Bulger	Avoid transfusions of red blood cells for arbitrary hemoglobin or hematocrit thresholds and in the absence of symptoms or active coronary disease, heart failure, or stroke.
Cure	Specialized	-	Bulger	Do not place, or leave in place, urinary catheters for incontinence or convenience or monitoring of output for non–critically ill patients (acceptable indications: critical illness, obstruction, hospice, peri-operatively for <2 days for urologic procedures; use weights instead to monitor diuresis).
Cure	Specialized	-	Chan	Chemotherapy in the last 14 days of life.
Cure	Specialized	-	Chan	Cancer-directed therapy for solid tumor patients with the following characteristics: low performance status (3 or 4), no benefit from prior evidence-based interventions, not eligible for a clinical trial, and no strong evidence supporting the clinical value of further anticancer treatment.
Cure	Specialized	-	Chan	Traction to treat low back pain.
Cure	Specialized	-	Chan	Administering ESAs to CKD patients with hemoglobin levels ≥ 10 g/dL without symptoms of anemia.
Cure	Specialized	-	Chan	Aggressive interventional procedures.
Cure	Specialized	-	Chan	Autologous bone marrow transplant with high-dose chemotherapy for advanced breast cancer.

ICHA-HC			First author	Low-value care recommendation
Cure	Specialized	-	Chan	Extracranial-intracranial bypass to reduce the risk of ischemic stroke.
Cure	Specialized	-	Chan	Gastric bubble for morbid obesity.
Cure	Specialized	-	Chan	Gastric freezing for peptic ulcer disease.
Cure	Specialized	-	Chan	Mammary artery ligation for CAD
Cure	Specialized	-	Chan	Optic nerve decompression surgery for NAION.
Cure	Specialized	-	Chan	Potentially preventable hospital admissions lasting < 24 hours.
Cure	Specialized	-	Chan	Radiation therapy for acne
Cure	Specialized	-	Chan	Recommending replacement immunoglobin therapy for recurrent infections unless impaired antibody responses to vaccines are demonstrated.
Cure	Specialized	-	Chan	Spinal manipulation for treatikng migraine or cluster headaches.
Cure	Specialized	-	Chan	Subcutaneous interferon alfa-2a to treat age-related macular degeneration.
Cure	Specialized	-	Chan	Supplemental oxygen for healthy premature babies.
Cure	Specialized	-	Chan	Unwarranted procedures, targeting knee/ hip replacement.
Cure	Specialized	-	Chan	White cell stimulating factors for primary prevention of febrile neutropenia for patients < 20% risk for this complication.
Cure	Specialized	-	Elshaug	Hysterectomy as a first-line treatment solely for heavy menstrual bleeding
Cure	Specialized	-	Elshaug	Surgical approach to hysterectomy for benign gynecological disease, abdominal hysterectomy (AH), vaginal hysterectomy (VH) and laparoscopic hysterectomy (LH)
Cure	Specialized	-	Elshaug	Radical prostatectomy
Cure	Specialized	-	Elshaug	Radical prostatectomy and external beam radiation therapy
Cure	Specialized	-	Elshaug	Upper airway surgery for obstructive sleep apnea syndrome
Cure	Specialized	-	Elshaug	Radiotherapy for patients with metastatic spinal cord disease
Cure	Specialized	-	Elshaug	Radiotherapy with the intention of preventing metastatic spinal cord compression (MSCC) in patients with asymptomatic spinal metastases.
Cure	Specialized	-	Elshaug	Spinal surgery with the intention of preventing metastatic spinal cord compression (MSCC)
Cure	Specialized	-	Elshaug	Prostatectomy for early stage prostate cancer

ICHA-HC			First author	Low-value care recommendation
Cure	Specialized	-	Elshaug	Active surveillance for men with high-risk localized prostate cancer (active surveillance includes PSA testing and prostate biopsy)
Cure	Specialized	-	Elshaug	Surgery for obstructive sleep apnea
Cure	Specialized	-	Elshaug	Radiotherapy for patients with metastatic spinal cord compression (MSCC) and planned surgery
Cure	Specialized	-	Elshaug	Posterior decompression alone in patients with metastatic spinal cord compression (MSCC).
Cure	Specialized	-	Elshaug	Adjuvant radiotherapy with surgery for endometrial cancer
Cure	Specialized	-	Elshaug	Anal fistula surgery in patients with inflammatory bowel disease
Cure	Specialized	-	Elshaug	Complementary therapies for chronic fatigue syndrome/myalgic encephalomyelitis.
Cure	Specialized	-	Elshaug	Conventional photon irradiation in treatment of chordoma
Cure	Specialized	-	Elshaug	Coronary stenting (angioplasty) for stable angina and in diabetic patients with multi-vessel disease
Cure	Specialized	-	Elshaug	Diagnosis of primary tumor site in metastatic cancer
Cure	Specialized	-	Elshaug	Dicectomy
Cure	Specialized	-	Elshaug	External fixation versus conservative treatment for distal radial fractures in adults
Cure	Specialized	-	Elshaug	Extracorporeal shock wave lithotripsy (ESWL) versus percutaneous nephrolithotomy (PCNL) or retrograde intrarenal surgery (RIRS) for kidney stones
Cure	Specialized	-	Elshaug	Extrapleural pneumonectomy for mesothelioma
Cure	Specialized	-	Elshaug	Femoral central vein catherization
Cure	Specialized	-	Elshaug	Hypothermia for traumatic head injury
Cure	Specialized	-	Elshaug	Implantable cardioverter defibrillators
Cure	Specialized	-	Elshaug	Interventions for treating acute Achilles tendon ruptures
Cure	Specialized	-	Elshaug	Intracavity lavage to reduce the risk of surgical site infection
Cure	Specialized	-	Elshaug	IVU for urothelial tumors
Cure	Specialized	-	Elshaug	Laparoscopic vs open colposuspension for urinary incontinence in women
Cure	Specialized	-	Elshaug	Medial pinning of supracondylar humeral fractures
Cure	Specialized	-	Elshaug	Needling for encapsulated trabeculectomy filtering blebs

ICHA-HC			First author	Low-value care recommendation
Cure	Specialized	-	Elshaug	Neurosurgical clipping for patients with aneurysmal subarachnoid hemorrhage
Cure	Specialized	-	Elshaug	Off-pump heart bypass
Cure	Specialized	-	Elshaug	Open total mesorectal excision for rectal cancer
Cure	Specialized	-	Elshaug	Pelvic lymphadenectomy for the management of endometrial cancer
Cure	Specialized	-	Elshaug	Postoperative radiotherapy for non-small cell lung cancer
Cure	Specialized	-	Elshaug	Prophylactic surgical litigation of patent ductus arteriosus for prevention of mortality and morbidity in extremely low birth weight infants
Cure	Specialized	-	Elshaug	Radiofrequency facet joint denervation
Cure	Specialized	-	Elshaug	Radiotherapy following mastectomy to patients with early invasive breast cancer at low risk of local recurrence
Cure	Specialized	-	Elshaug	Radiotherpay for neovascular age-related macular degeneration
Cure	Specialized	-	Elshaug	Removal of adenoids
Cure	Specialized	-	Elshaug	Rhinomanometry and acoustic rhinometry
Cure	Specialized	-	Elshaug	Rubber band ligation versus excisional haemorrhoidectomy for haemorrhoids
Cure	Specialized	-	Elshaug	Scalpel versus no-scalpel incision for vasectomy
Cure	Specialized	-	Elshaug	Sentinel lymph node biopsy (SLNB) in patients with a preoperative diagnosis of ductal carcinoma in situ (DCIS).
Cure	Specialized	-	Elshaug	Standard central venous catheters
Cure	Specialized	-	Elshaug	Stem Cell transplantation for AML
Cure	Specialized	-	Elshaug	Suprapubic urinary catheter
Cure	Specialized	-	Elshaug	Temporary defunctioning stoma in people undergoing anal sphincter repair
Cure	Specialized	-	Elshaug	Tension free repair for asymptomatic inguinal hernia
Cure	Specialized	-	Elshaug	Total fundoplication for gastroesophageal reflux disease
Cure	Specialized	-	Elshaug	Transit studies to diagnose idiopathic constipation
Cure	Specialized	-	Elshaug	Transurethral resection of the prostate for symptomatic benign prostatic obstruction
Cure	Specialized	-	Elshaug	Tube thoracostomy (TT) in thoracic surgery clinics
Cure	Specialized	-	Elshaug	Ultrasound-guided internal jugular (USIJ) versus the subclavian (SC) vein approach for central venous cannulation (CVC)

ICHA-HC			First author	Low-value care recommendation
Cure	Specialized	-	Elshaug	UVB therapy for vitiligo
Cure	Specialized	-	Elshaug	Whole brain radiotherapy for the treatment of multiple brain metastases
Cure	Specialized	-	Halpern	Don't transfuse red blood cells in hemodynamically stable, non-bleeding ICU-patients with a hemoglobin concentration greater than 7 mg/dL.
Cure	Specialized	-	Halpern	Don't use parenteral nutrition in adequately nourished critically ill patients within the first seven days of an ICU stay.
Cure	Specialized	-	Halpern	Don't deeply sedate mechanically ventilated patients without a specific indication and without daily attempts to lighten sedation.
Cure	Specialized	-	Halpern	Don't continue life support for patients at high risk for death or severely impaired functional recovery without offering patients and their families the alternative of care focused entirely on comfort.
Cure	Specialized	-	Hicks	In situations where transfusion of RBCs is necessary, transfuse the minimum number of units required to relieve symptoms of anemia or to return the patient to a safe hemoglobin range (7-8 g/dL in stable, non-cardiac in-patients)
Cure	Specialized	-	Hicks	Do not administer plasma or prothrombin complex concentrates for non-emergent reversal of vitamin K antagonists (ie, outside of the setting of major bleeding, intracranial hemorrhage, or anticipated emergent surgery)
Cure	Specialized	-	Korenstein	Hysterectomy
Cure	Specialized	-	Korenstein	Transfusion
Cure	Specialized	-	Korenstein	Chemotherapy (colon cancer)
Cure	Specialized	-	Korenstein	Coronary artery bypass graft
Cure	Specialized	-	Korenstein	Coronary revascularization for coronary artery disease
Cure	Specialized	-	Korenstein	Oxygen therapy
Cure	Specialized	-	Korenstein	PTCA
Cure	Specialized	-	Korenstein	Tympanostomy tubes
Cure	Specialized	-	Onuoha	Intraoperative administration of packed red blood cells in a young healthy patient without ongoing blood loss and hemoglobin concentration of >6 g/dL unless symptomatic or hemodynamically unstable

ICHA-HC			First author	Low-value care recommendation
Cure	Specialized	-	Onuoha	Intraoperative routine administration of colloid (dextrans, hydroxyethyl starches, albumin) for volume resuscitation without appropriate indications; clinicians should refer to current data for its use in specific populations
Cure	Specialized	-	Onuoha	Intraoperative routine use of pulmonary arterial catheter for cardiac surgery in patients with low risk of hemodynamic complications (especially with concomitant use of alternative diagnostic tools, e.g., TTE or TEE)
Cure	Specialized	-	Quinonez	Do not use continuous pulse oximetry routinely in children with acute respiratory illness unless they are on supplemental oxygen.
Cure	Specialized	-	Wiener	For patients recently discharged on supplemental home oxygen following hospitalization for an acute illness, do not renew the prescription without assessing the patient for ongoing hypoxemia.
Cure	Specialized	-	Williams	Don't administer erythropoiesis-stimulating agents (ESAs) to CKD patients with hemoglobin levels ≥ 10 g/dl without symptoms of anemia.
Cure	Specialized	-	Williams	Don't place peripherally inserted central catheters (PICCs) in stage 3–5 CKD patients without consulting nephrology.
Cure	Specialized	Imaging	Bulger	Do not order continuous telemetry monitoring outside of the ICU without using a protocol that governs continuation.
Cure	Specialized	Imaging	Chan	Unwarranted diagnostic procedures, targeting uncomplicated chest/thorax CT screening.
Cure	Specialized	Imaging	Chan	Performing DEXA screening for osteoporosis in women < 65 years or men <70 years with no risk factors.
Cure	Specialized	Imaging	Chan	Performing follow-up imaging studies for incidentally discovered pulmonary nodules ≤ 4 mm in low-risk individuals.
Cure	Specialized	Imaging	Chan	Obtaining CT scans in a patient with pneumonia that is confirmed by chest radiography in the absence of complicating clinical or radiographic features.
Cure	Specialized	Imaging	Chan	Antinuclear antibody test in patients with nonspecific symptoms or in patients with fibromyalgia.
Cure	Specialized	Imaging	Chan	CT for the evaluation of suspected appendicitis in children until after ultrasound has been considered as an option.

ICHA-HC			First author	Low-value care recommendation
Cure	Specialized	Imaging	Chan	Performing follow-up surveillance examination in < 3 years if patient had a second endoscopy that confirms the absence of dysplasia on biopsy.
Cure	Specialized	Imaging	Chan	Performing imaging studies as the initial diagnostic test in patients with low pretest of VTE.
Cure	Specialized	Imaging	Chan	Performing pre-discharge chest radiography for hospitalized patients with community-acquired pneumonia who are making satisfactory clinical recovery.
Cure	Specialized	Imaging	Chan	Repeat imaging studies (< 60 days since prior test).
Cure	Specialized	Imaging	Chan	Repeating screening ultrasonography for following a negative study.
Cure	Specialized	Imaging	Chan	Unwarranted diagnostic procedures, targeting bone or joint x-ray prior to conservative therapy, without red flags.
Cure	Specialized	Imaging	Chan	Unwarranted diagnostic procedures, targeting endoscopy.
Cure	Specialized	Imaging	Chan	Unwarranted diagnostic procedures, targeting chest x-ray, preoperative, on admission or routine monitoring.
Cure	Specialized	Imaging	Elshaug	Chest radiograph in acute respiratory infections
Cure	Specialized	Imaging	Elshaug	Routine chest x-rays on children with fever (without features of serious illness)
Cure	Specialized	Imaging	Elshaug	Halter monitoring (24 hour ECG) in young patients with palpitations and history indicating ectopic beats
Cure	Specialized	Imaging	Elshaug	Computerized tomography (CT) of the pelvis in men with low- or intermediate-risk localized prostate cancer
Cure	Specialized	Imaging	Elshaug	Chest x-ray in children with symptoms and signs suggesting pneumonia
Cure	Specialized	Imaging	Elshaug	Abdominal ultrasound to diagnose idiopathic constipation
Cure	Specialized	Imaging	Elshaug	Angiography in lower limb vascular trauma patients
Cure	Specialized	Imaging	Elshaug	Auditory brainstem responses for diagnosing CFS
Cure	Specialized	Imaging	Elshaug	Chest x-ray for diagnosis of acute coronary syndrome
Cure	Specialized	Imaging	Elshaug	CT or MRI in primary aldosteronism
Cure	Specialized	Imaging	Elshaug	CT or ultrasound to diagnose appendicitis
Cure	Specialized	Imaging	Elshaug	CT scans (head) in children with low risk of clinically important brain injuries after trauma

ICHA-HC			First author	Low-value care recommendation
Cure	Specialized	Imaging	Elshaug	Cystoscopy for men with uncomplicated lower urinary tract symptoms (LUTS).
Cure	Specialized	Imaging	Elshaug	Endoscopic retrograde cholangiopancreatiography in acute gallstone pancreatitis without cholangitis
Cure	Specialized	Imaging	Elshaug	Fluorimetry or endoscopy to assess dysphasia
Cure	Specialized	Imaging	Elshaug	Imaging of the upper urinary tract in men with uncomplicated lower urinary tract symptoms (LUTS)>
Cure	Specialized	Imaging	Elshaug	Inappropriate indication for upper endoscopy
Cure	Specialized	Imaging	Elshaug	Mammography of the ipsilateral soft tissues after mastectomy
Cure	Specialized	Imaging	Elshaug	Plain abdominal radiograph to diagnose idiopathic constipation in children and young people
Cure	Specialized	Imaging	Elshaug	Plain radiographs of the spine to make or to exclude the diagnosis of spinal metastases or metastatic spinal cord compression (MSCC).
Cure	Specialized	Imaging	Elshaug	Routine daily chest radiographs in intensive care
Cure	Specialized	Imaging	Elshaug	Routine imaging of the spine in patients with a previous diagnosis of malignancy
Cure	Specialized	Imaging	Elshaug	Routine screening for cardiac anomalies using nuchal translucency
Cure	Specialized	Imaging	Elshaug	Routine spinal magnetic resonance imaging (MRI) for all men with hormone-refractory prostate cancer and known bone metastases
Cure	Specialized	Imaging	Elshaug	Routine ultrasound in infants or children for (UTI)
Cure	Specialized	Imaging	Elshaug	Saline infusion sonography as a first-line diagnostic tool
Cure	Specialized	Imaging	Hicks	Limit surveillance CT scans in asymptomatic patients after curative-intent treatment for aggressive lymphoma
Cure	Specialized	Imaging	Keyhani	Coronary angiography for acute myocardial infarction
Cure	Specialized	Imaging	Keyhani	Coronary angiography - for all indications
Cure	Specialized	Imaging	Keyhani	Radionuclide myocardial perfusion imaging
Cure	Specialized	Imaging	Korenstein	Coronary angiography for myocardial infarction, coronary artery disease
Cure	Specialized	Imaging	Korenstein	Radiographs in acute respiratory illnesses for bronchiolitis or croup, asthma
Cure	Specialized	Imaging	Korenstein	CT scan for epilepsy
Cure	Specialized	Imaging	Korenstein	Endoscopy (upper) for bleeding (upper), peptic ulcer disease

ICHA-HC			First author	Low-value care recommendation
Cure	Specialized	Imaging	Quinonez	Do not order chest radiographs in children with asthma or bronchiolitis.
Cure	Specialized	Imaging	Rouster-Stevens	Do not routinely perform surveillance joint radiographs to monitor JIA disease activity.
Cure	Specialized	Imaging	Schuur	Do not order computed tomography (CT) of the cervical spine for patients after trauma who do not meet the National Emergency X-ray Utilization Study (NEXUS) low-risk criteria9 or the Canadian C-Spine Rule.
Cure	Specialized	Imaging	Wiener	Do not perform CT scan surveillance for evaluation of indeterminate pulmonary nodules at more frequent intervals or for a longer period of time than recommended by established guidelines.
Cure	Specialized	Imaging	Wiener	Do not perform chest CT angiography to evaluate for possible PE in patients with a low clinical probability and negative results of a highly sensitive D -dimer assay.
Cure	Specialized	Imaging	Wood	Do Not Initiate Routine Evaluation of Carotid Artery Disease Before Cardiac Surgery in the Absence of Symptoms or Other High-Risk Criteria
Cure	Specialized	Imaging	Wood	Do Not Perform a Routine Pre-discharge Echocardiogram After Cardiac Valve Replacement Surgery
Cure	Specialized	Imaging	Wood	Patients With Suspected or Biopsy Proven Stage I NSCLC Do Not Require Brain Imaging Before Definitive Care in the Absence of Neurologic Symptoms
Cure	Specialized	Lab	Chan	Performing unproven diagnostic tests (e.g. IgG or IGE) in the evaluation of allergy
Cure	Specialized	Lab	Chan	Performing follow-up complete blood counts, blood chemistry studies, tumor marker studies, chest radiography, or imaging studies other than appropriate breast imaging in asymptomatic women with previously treated breast cancer
Cure	Specialized	Lab	Chan	Measuring brain natriuretic peptide in the initial evaluation of patients with typical findings of heart failure.
Cure	Specialized	Lab	Chan	Pre-op coagulation tests when risk or predisposing factors for bleeding and history of abnormal bleeding are absent.
Cure	Specialized	Lab	Chan	Routine pre-op laboratory tests in otherwise healthy patients undergoing elective surgery.
Cure	Specialized	Lab	Chan	Routinely do diagnostic testing in patients with chronic urticaria
Cure	Specialized	Lab	Elshaug	Tissue biopsy to reassess HER2 status
Cure	Specialized	Lab	Elshaug	Tissue biopsy to reassess ER status

ICHA-HC			First author	Low-value care recommendation
Cure	Specialized	Lab	Elshaug	Assessing progesterone receptor status of tumors in patients with invasive breast cancer
Cure	Specialized	Lab	Elshaug	Inflammatory markers for prediction of recurrent stroke
Cure	Specialized	Lab	Elshaug	Measurement of alfa-fetoprotein in alpha-fetoprotein producing gastric cancers
Cure	Specialized	Lab	Elshaug	Measurement of bilirubin levels in babies who are not visibly jaundiced
Cure	Specialized	Lab	Elshaug	Mortality markers in end stage renal disease
Cure	Specialized	Lab	Elshaug	Troponin Tests for evaluation of heart attack/heart injury
Cure	Specialized	Lab	Elshaug	Uro4 HB&L system for the rapid diagnosis of lower respiratory tract infections in intensive care units
Cure	Specialized	Lab	Halpern	Don't order diagnostic tests at regular intervals (such as every day), but rather in response to specific clinical questions.
Cure	Specialized	Lab	Hicks	Do not test for thrombophilia in adult patients with venous thromboembolism occurring in the setting of major transient risk factors (surgery, trauma, or prolonged immobility)
Cure	Specialized	Lab	Korenstein	Tumor markers
Cure	Specialized	Lab	Onuoha	Preoperative baseline laboratory studies (CBC, BMP or CMP, coagulation studies) in healthy patients without significant systemic disease (ASA I or II), when blood loss (or fluid shifts) is expected to be minimal
Cure	Specialized	Lab	Rouster-Stevens	Do not order autoantibody panels unless positive ANAs and evidence of rheumatic disease
Cure	Specialized	Lab	Rouster-Stevens	Do not repeat a confirmed positive ANA in patients with established JIA or SLE.
Cure	Specialized	Lab	Rouster-Stevens	Do not perform methotrexate toxicity labs more often than every 12 weeks when patients are on stable doses.
Cure	Specialized	Lab/Im	Korenstein	Post-cancer surveillance
Cure	Specialized	Pharmaceuticals	Chan	Chelation therapy to prevent or reverse atherosclerosis.
Cure	Specialized	Pharmaceuticals	Chan	Hydralazine for CHF.
Cure	Specialized	Pharmaceuticals	Chan	Lidocaine to prevent arrhythmia and sudden death in AMI.
Cure	Specialized	Pharmaceuticals	Chan	Quinidine for suppressing recurrences of atrial fibrillation.
LTC	Day	-	Elshaug	Lower-extremity arteriovenous access for hemodialysis

ICHA-HC			First author	Low-value care recommendation
LTC	Day	-	Williams	Don't initiate chronic dialysis without ensuring a shared decision-making process between patients, their families, and their physicians.
LTC	Inpatient	-	AGS Choosing Wisely Workgroup	Don't recommend percutaneous feeding tubes in patients with advanced dementia; instead, offer oral assisted feeding.
LTC	Inpatient	-	AGS Choosing Wisely Workgroup	Avoid using physical restraints to manage behavioral symptoms of hospitalized older adults with delirium
LTC	Inpatient	Pharmaceuticals	AGS Choosing Wisely Workgroup	Don't use benzodiazepines or other sedative–hypnotics in older adults as first choice for insomnia, agitation, or delirium.
LTC	Inpatient	Pharmaceuticals	AGS Choosing Wisely Workgroup	Avoid using prescription appetite stimulates or high-calorie supplements for treatment of anorexia or cachexia in older adults; instead optimize social supports, providing feeding assistance, and clarify patient goals and expectations
LTC	Inpatient	Pharmaceuticals	AGS Choosing Wisely Workgroup	Don't prescribe a medication without conducting a drug regimen review
LTC	Inpatient	Pharmaceuticals	AGS Choosing Wisely Workgroup	Don't prescribe cholinesterase inhibitors for dementia without periodic assessment for perceived cognitive benefits and adverse gastrointestinal effects.
LTC	Outpatient	-	AGS Choosing Wisely Workgroup	Don't use antipsychotics as first choice to treat behavioral and psychological symptoms of dementia.
LTC	Outpatient	-	Chan	Monitoring patients with asthma or COPD by using full pulmonary function testing that includes lung volumes and diffusing capacity, rather than spirometry alone (or peak expiratory flow rate monitoring in asthma).
LTC	Outpatient	-	Elshaug	Spirometry during COPD exacerbation and treatment monitoring
LTC	Outpatient	Lab	Elshaug	Frequent monitoring HbA1C levels in adults with diabetes
LTC	Outpatient	Pharmaceuticals	AGS Choosing Wisely Workgroup	Avoid using medications to achieve hemoglobin A1c <7.5% in most adults age 65 and older; moderate control is generally better.
LTC	Outpatient	Pharmaceuticals	Amos	Amitryptiline in ≥65 year olds
LTC	Outpatient	Pharmaceuticals	Amos	Chlorpropamide in ≥65 year olds
LTC	Outpatient	Pharmaceuticals	Amos	Citalopram > 20 mg / day in ≥65 year olds

ICHA-HC			First author	Low-value care recommendation
LTC	Outpatient	Pharmaceuticals	Amos	Clonidine (oral) in ≥65 year olds
LTC	Outpatient	Pharmaceuticals	Amos	Disopyramide in ≥65 year olds
LTC	Outpatient	Pharmaceuticals	Amos	Escitalopram >10 mg/day in ≥65 year olds
LTC	Outpatient	Pharmaceuticals	Amos	Indomethacin in ≥65 year olds
LTC	Outpatient	Pharmaceuticals	Amos	Ketorolac max 2 days in ≥65 year olds
LTC	Outpatient	Pharmaceuticals	Amos	Methyldopa in ≥65 year olds
LTC	Outpatient	Pharmaceuticals	Amos	Nifedipine (short acting) in ≥65 year olds
LTC	Outpatient	Pharmaceuticals	Amos	NSAID's > 15 days in ≥65 year olds
LTC	Outpatient	Pharmaceuticals	Amos	Oestrogen in ≥65 year olds
LTC	Outpatient	Pharmaceuticals	Amos	Orphenadrine in ≥65 year olds
LTC	Outpatient	Pharmaceuticals	Amos	Pentazocine in ≥65 year olds
LTC	Outpatient	Pharmaceuticals	Amos	Spironolactone >25 mg/day in ≥65 year olds
LTC	Outpatient	Pharmaceuticals	Amos	Testosterone in ≥65 year olds
LTC	Outpatient	Pharmaceuticals	Chan	Inappropriate medication use, targeting polypharmacy (for multiple chronic conditions; of antipsychotics).
LTC	Outpatient	Pharmaceuticals	Chan	Overuse or early use of third-line treatment, for example Avandia for diabetes.
Preventive	-	-	Chan	Screening for COPD with spirometry in individuals without respiratory symptoms
Preventive	-	-	Elshaug	Spirometry for COPD screening
Preventive	-	Imaging	AGS Choosing Wisely Workgroup	Don't recommend screening for breast or colectoral cancer, nor prostate cancer (with the prostate-specific antigen test) without considering life expectancy and the risks of testing, over-diagnosis and overtreatment
Preventive	-	Imaging	Chan	Coronary heart disease screening using ECG, exercise treadmill test (ETT), electron beam CT in low-risk adults
Preventive	-	Imaging	Chan	Obtaining ECGs to screen for cardiac disease in patients at low to average risk for CAD.
Preventive	-	Imaging	Chan	Obtaining exercise ECG for screening in low-risk asymptomatic adults.
Preventive	-	Imaging	Chan	Ordering annual ECG or any other cardiac screening for asymptomatic, low-risk patients.
Preventive	-	Imaging	Chan	Screening for colectoral cancer in adults older than 75 years or in adults with a life expectancy of less than 10 years.
Preventive	-	Imaging	Chan	Repeating colonoscopy within 5 years in: (a) asymptomatic patients with low-risk adenomas, (b) patients with one or two small (< 1 cm) adenomatous polyps, without high-grade dysplasia, completely removed via high-quality colonoscopy.

ICHA-HC			First author	Low-value care recommendation
Preventive	-	Imaging	Chan	Carotid artery stenosis screening in general adult population
Preventive	-	Imaging	Chan	Repeat cardiac studies within a 3-month period.
Preventive	-	Imaging	Chan	Using MRI rather than mammography as the breast cancer screening test of choice for average-risk women.
Preventive	-	Imaging	Elshaug	Fecal occult blood screening for colorectal cancer
Preventive	-	Imaging	Keyhani	Surveillance endoscopy
Preventive	-	Imaging	Korenstein	Cardiac stress test
Preventive	-	Imaging	Korenstein	Fecal occult blood screening for colorectal cancer
Preventive	-	Imaging	Korenstein	Echocardiogram
Preventive	-	Imaging	Korenstein	Periodic health examination: electrocardiogram
Preventive	-	Imaging	Korenstein	Colonoscopy for colon cancer screening and follow-up
Preventive	-	Imaging	Korenstein	Periodic health examination: chest radiography
Preventive	-	Imaging	Wiener	Do not perform CT scan screening for lung cancer among patients at low risk for lung cancer.
Preventive	-	Lab	Chan	Prostate cancer screening in males ≥ 75 years or with a life expectancy of < 10 years.
Preventive	-	Lab	Chan	Pap tests in females <21 years or post-hysterectomy for benign disease; cervical cancer screening in average in average to low-risk females ≥ 65 years or post-hysterectomy for benign disease.
Preventive	-	Lab	Chan	Using CA-125 antigen levels to screen woman for ovarian cancer in the absence of increased risk.
Preventive	-	Lab	Chan	Blood chemistry panels or urinalyses for screening in asymptomatic, healthy adults.
Preventive	-	Lab	Chan	BRCA mutation testing for breast and ovarian cancer in low-risk females.
Preventive	-	Lab	Chan	Screening low-risk individuals for hepatitis B virus infection.
Preventive	-	Lab	Elshaug	Prostate specific antigen (PSA) testing
Preventive	-	Lab	Elshaug	The routine measurement of vitamin D levels in stage 1, 2, 3A or 3B chronic kidney disease (CKD) is not recommended
Preventive	-	Lab	Elshaug	Chlamydia screening in under 25 year olds
Preventive	-	Lab	Korenstein	Prostate-specific antigen :prostate cancer screening
Preventive	-	Lab	Korenstein	Papanicolaou test: cervical cancer screening

ICHA-HC			First author	Low-value care recommendation
Preventive	-	Lab	Korenstein	Periodic health examination: urinalysis
Preventive	-	Lab/Im	Chan	Routine cancer screening for dialysis patients with limited life expectancies without signs or symptoms.
Preventive	-	Lab/Im	Williams	Don't perform routine cancer screening for dialysis patients with limited life expectancies without signs or symptoms.
Rehabilitative	-	-	Elshaug	CBT for schizophrenia, bipolar disorder and major depression
Rehabilitative	-	-	Elshaug	Electroconvulsive therapy (ECT) for people with moderate depression
Rehabilitative	-	-	Elshaug	Social skills training (as a specific intervention) to people with schizophrenia
Rehabilitative	-	-	Elshaug	Structural neuroimaging techniques (either magnetic resonance imaging (MRI) or computed axial tomography (CT) scanning) for the management of first-episode psychosis

Lab.: Laboratory; Im.: Imaging


3

CHAPTER 3

Unraveling the drivers of regional variation in healthcare spending by analyzing prevalent chronic diseases

Published as
De Vries EF, Heijink R, Struijs JN, Baan CA
Unraveling the drivers of regional variation in healthcare spending by analyzing prevalent chronic diseases
BMC Health Serv Res. 2018, 18(1):323

ABSTRACT

Background: To indicate inefficiencies in health systems, previous studies examined regional variation in healthcare spending by analyzing the entire population. As a result, population heterogeneity is taken into account to a limited extent only. Furthermore, it clouds a detailed interpretation which could be used to inform regional budget allocation decisions to improve quality of care of one chronic disease over another. Therefore, we aimed to gain insight into the drivers of regional variation in healthcare spending by studying prevalent chronic diseases.

Methods: We used 2012 secondary health survey data linked with claims data, healthcare supply data and demographics at the individual level for 18 Dutch regions. We studied patients with diabetes (n=10,767) and depression (n=3,735), in addition to the general population (n=44,694). For all samples, we estimated the cross-sectional relationship between spending, supply and demand variables and region effects using linear mixed models.

Results: Regions with above (below) average spending for the general population mostly showed above (below) average spending for diabetes and depression as well. Less than 1% of the a-priori total variation in spending was attributed to the regions. For all samples, we found that individual-level demand variables explained 62-63% of the total variance. Self-reported health status was the most prominent predictor (28%) of healthcare spending. Supply variables also explained, although a small part, of regional variation in spending in the general population and depression. Demand variables explained nearly 100% of regional variation in spending for depression and 88% for diabetes, leaving 12% of the regional variation left unexplained indicating differences between regions due to inefficiencies.

Conclusions: The extent to which regional variation in healthcare spending can be considered as inefficiency may differ between regions and disease-groups. Therefore, analyzing chronic diseases, in addition to the traditional approach where the general population is studied, provides more insight into the causes of regional variation in healthcare spending, and identifies potential areas for efficiency improvement and budget allocation decisions.

BACKGROUND

The sustainability of healthcare systems is a common challenge for Western countries. One of the main responses is to improve population health and quality of care, while slowing down the expenditure growth by population health management (PHM). PHM aims to integrate services across healthcare, prevention, social care, welfare and public health for a pre-specified population within the region [1]. Commonly, these initiatives are set-up by a network of health insurers, health care providers and other health organizations that together develop interventions. These interventions include so-called citizen-centered interventions, which targets citizens at risk and are tailored for pre-defined subpopulations such as citizens who smoke, diabetes patients or multi-morbid patients. In a number of countries, stakeholders within specific regions have adopted the PHM approach. Examples are Gesundes Kinzigtal in Germany [2], Accountable Health Communities in the USA [3] and Dutch PHM initiatives [4, 5]. In order to develop policy to improve efficiency and to make budget allocation decisions, these initiatives require insight into the performance of (the organization of) their health services. One way to inform these decisions is to study the drivers of variation in healthcare spending between regions, as regional variation in healthcare spending, not caused by differences in medical need, is said to *indicate* inefficiency [6, 7].

Research has shown that regional variation in healthcare spending is determined by the interplay of demand and supply factors under the influence of system factors [8, 9]. Demand factors labeled as medical need, such as demographics and health status, are generally considered justifiable causes of variation in healthcare spending. Empirical research has shown that such factors explain a large part of the regional variation in healthcare spending (e.g. [8, 10, 11]). Regional variation as a result of supply factors (e.g. competition, capacity or physician beliefs) is considered to be generally undesirable. In empirical studies, those factors have been found to be of varying impact, depending on the study context, or system factors (e.g. [11, 12]). System factors such as (insurance) regulation, price setting or payment methods influence the dynamics of demand and supply.

Previous studies (e.g. [6, 10, 13]) generally examined regional variation in healthcare spending studying sources of variation that affect spending for the entire population in a region, such as local price levels or general economic circumstances. However, there is substantial population heterogeneity that clouds a more detailed interpretation, as other causes of variation might differ between disease groups across regions. Population heterogeneity between regions exists, for example in terms of the prevalence of (chronic)

diseases and multi-morbidities; a region may have a relatively high prevalence for certain disease groups and lower prevalence for others. Furthermore, the extent to which regional differences are caused by disease severity or other patient characteristics or the level of treatment standardization may vary between disease groups [14, 15]. These more detailed explanations remain undetected when studying the general population only. Also, because of lack of knowledge and data, it is complicated to control for all relevant sources of variation in healthcare spending. In order to inform, for instance, budget allocation decisions for quality improvements or care standardization efforts of one chronic disease over another, a more detailed approach may be needed.

Therefore, this study aimed to gain further insight into the sources of regional variation in healthcare spending by zooming in on prevalent chronic diseases (disease-approach). We applied this to the context of 18 Dutch PHM regions. We expected variation patterns to differ between disease-groups within the population. This is in part intrinsically to the disease and in part in line with the degree to which consensus is reached on how to treat a certain disease [8, 15]. We selected two types of prevalent disease groups: patients with diabetes and patients with depression. We chose patients with diabetes as treatments for diabetes are known to be highly standardized due to provider approved general treatment decisions in the Dutch Diabetes Federation Health Care Standard (DFHCS) and as a consequence of the bundled payment model that was introduced in 2010 [16]. We chose patients with depression as treatments for depression are expected to be less standardized, considering the variety of treatment options and lack of healthcare standards at a national level. Assuming that treatment options vary in costs, we expected the group of patients with depression to reflect more variation in healthcare spending as compared to the group of patients with diabetes. We showed the results of the disease-approach in addition to the results of the traditional approach (where the general population is studied) for each step taken in the analysis of regional variation in spending, which is to 1) describe the unadjusted regional variation in healthcare spending, and 2) explore the extent to which demand and supply factors explain regional variation in healthcare spending.

METHODS

Data sources

We used data from the Dutch Public Health Monitor (DPHM) for the year 2012 [17]. We merged these data at the individual level with demographic variables from Statistics Netherlands and nationwide claims data which we obtained from Vektis [18].

The claims data included all health care use in 2012 that was covered by the basic health insurance. The records were linked using a pseudonymized personal identification code. Furthermore, at the regional level, we added health care supply data from the Netherlands Institute for Health Services Research (NIVEL), the National Institute of Public Health and the Environment (RIVM) and the Health Care Inspectorate (IGZ). We used 2012 as reference year for all data sources. For an overview of data and variables used, see Supplementary File 1.

Study population

To define the regions, we used the geographical demarcation of 18 Dutch PHM initiatives [4]. We analyzed the variation in healthcare spending across these regions, focusing on three (sub)populations: 1) the general population, 2) patients with diabetes, and 3) patients with depression. For a detailed description of the sample selection, see Supplementary File 2.

The first sample (general population) was based on participation in the DPHM survey (n=138,000), since we aimed to include health status information from that survey. We excluded individuals with missings on self-reported health status (n=82,455) and missings on the health spending variable (n=10,298). In addition, to correct for errors and costs in the data that are not linked to Diagnosis Treatment Groups (DTGs) in the previous year, we removed 1% outliers in terms of spending (above €29,468) (n=1,284) and negative values for health spending (n=1). As a result, 44,694 individuals from the general population were included in the sample. The second study sample consisted of diabetes patients. We selected them by their participation in a bundled payment program for diabetes care, the Pharmacy-based Cost Group (PCG) or the DTGs referring to diabetes type II (n=10,767), which is known to include nearly all diabetes type II patients [19]. The PCGs are – among other variables - used in the Dutch risk equalization system to identify risk-profiles that predict healthcare spending in the following year [20]. PCGs aim to identify persons with chronic diseases based on claims data of the previous year for medication for which it is known that they are used for that particular disease. The third study sample was created by selecting patients with depression (n=3,735) in a similar way as the diabetes patients. We used PCGs and DTGs referring to depression.

Econometric specification

To allow for the complex structure of the data, i.e. individuals nested within regions and the skewed distribution of healthcare costs as dependent variable, we performed multilevel analyses using linear mixed models (LMM) with a random intercept for region effects. We applied LMM to the log-transformed dependent variable. We

considered other methods and distributions: see Supplementary File 3 for details on the model selection. We specified the following equation (1):

$$\log(y_{ij}) = \beta_0 + \sum \beta_{ij} X_{ij} + \sum \beta_{0j} X_{0j} + \nu_{0j} + \varepsilon_{ij} \quad (1)$$

where y_{ij} is the healthcare spending of individual i in region j, β_{ij} are the fixed effects of individual level characteristics; X_{ij} is the vector of variables at the individual level; β_{0j} are the fixed effects of the region level characteristics; X_{0j} is the vector of variables at the region level; ν_{0j} is the random intercept at the region level with $\nu_{0j} \sim N(0, {\sigma_\nu}^2)$, and ε_{ij} is the residual error at the individual level with $\varepsilon_{ij} \sim N(0, {\sigma_\varepsilon}^2)$.

The outcome variable was the natural log of total curative healthcare spending in 2012 at the individual level (see Supplementary File 4). It was calculated by summing spending on general practitioner (GP)-care, hospital and specialist care, pharmaceutical care, physical therapy, mental health care and other types of care (i.e. patient transport, maternity care, medical aids and care abroad). It comprised all spending within the basic health insurance package [18].

Similar to previous studies, individual characteristics (age, gender, SES and health status) were included in the model. These variables reflect justified causes of variation in healthcare spending, which is also referred to as medical need (e.g. [6]). The self-reported health status variable was derived from the DPHM survey consisting of three levels from bad to very good [17]. Self-reported health status is generally considered to be a good predictor for future health as it encompasses more than standard objective measure of health (e.g. blood pressure, presence of disease) and reflects preclinical diseases or worsening of diseases [21]. In addition, using the self-reported health status in addition to using claims data derived health status alone, resolves the bias resulting from claims data derived health status, which is not independent from the supply-side and therefore may also reflect supplier-induced utilization. We used 2008 and 2012 Diagnosis-based Cost Groups (DCGs), which are cost-profiles of the diseases based on diagnosis information [22] and PCGs from the Dutch risk equalization model [20], each referring to data from previous years. Using claims data from previous years may tackle the endogeneity issues that arise when using spending and health data from the same year [23]. The DCG variable ranged from 0 to 13 or 15 (13 clusters for 2008 and 15 for 2012) where a higher DCG number is equal to being in a higher cost cluster. We transformed the DCG into dummy variables, reflecting being in either a specific DCG or not. The PCG variable ranged from 0 to 20 or 25 (20 clusters for 2008 and 25 for 2012), where a higher number means being in a higher cost cluster. As it was

possible for an individual to be in more PCGs at the same time (i.e. reflecting multiple diseases), we constructed an aggregate score from the PCG by summing the PCGs for each individual.

Supply variables, in this study mainly about the access to healthcare, reflecting generally unjustified variation, were included at the individual and at the regional level. At the individual level, we added distance to provider in meters from Statistics Netherlands. At the regional level, we used number of providers per 1000 population, which we constructed using the four-digit postal codes of providers in the Netherlands using information from NIVEL, RIVM and IGZ.

The random intercept at the regional level in the full model is considered to *indicate* variation in efficiency between regions, as it reflects regional variation which is not due to the demand and supply variables and the random error at the individual level in the model.

Statistical analyses

First, in order to gain insight in the a-priori variation of the three samples, we showed the deviation from the mean in percentages per region and the overall coefficient of variation (CoV; ratio of the standard deviation and the mean) at the regional level. The a-priori regional variation includes all justified and unjustified variation in healthcare spending between regions.

Second, we studied the extent to which demand and supply factors explained variation in healthcare spending at the individual and regional level. For all samples separately, to identify the contribution of the variables to the model, we fitted LMM models by gradually adding groups of variables to a null model without covariates. This strategy is common in analyzing regional variation in healthcare spending and is previously used by e.g. Gopffarth et al [11] and Newhouse et al [6]. We started with demographics (age and gender), followed by health status variables (first self-reported health status and then the DCGs and PCGs). Consecutively, we added the supply variables one by one. Finally, we selected the best-fit model based on subsequent nested fits for each sample using the likelihood ratio test (LRT), Akaike Information Criteria (AIC) and Bayesian Information Criterion (BIC). We chose the more expanded model when the p-value for the LRT < 0.01, and checked whether the AIC and BIC corresponded accordingly. After fitting the model, we calculated Pearson's correlation coefficients for the observed and predicted means of natural log of the curative healthcare spending to estimate how the

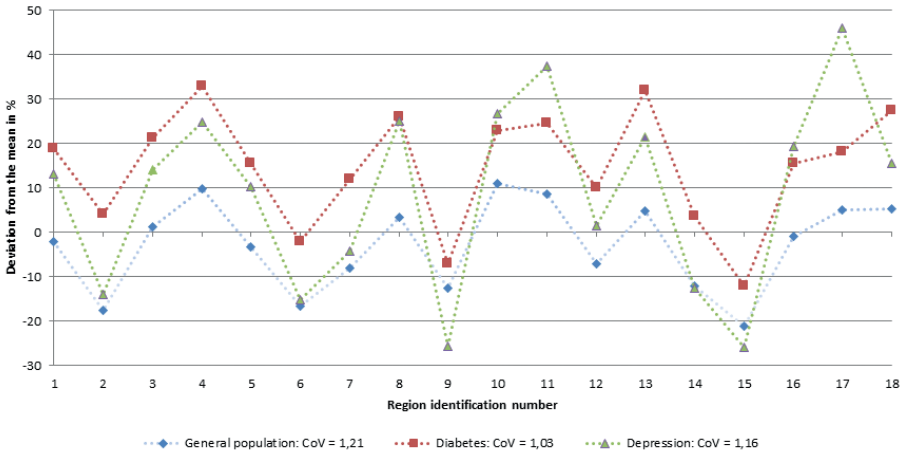
predicted means changed by adding the covariates. In addition, we plotted the variance at each level for all samples.

Data management and analyses were performed using SPSS Version 20 and Stata/MP 14.0.

RESULTS

Unadjusted regional variation in healthcare spending

Figure 1 displays the unadjusted regional variation shown as deviation from the mean in percentages per sample per region. It includes both justified and unjustified variation in healthcare spending. For the diabetes sample, region 13 and 4 ranked highest in spending per capita (pc). The two lowest spending regions were 15 and 9. The depression sample had highest spending pc in regions 17 and 11; lowest in region 15 and 9 (see Supplementary File 4 for more details and for the exact levels of spending per region). Regions with above (below) average spending for the general population mostly showed above (below) average spending for diabetes and depression as well. The disease-groups showed less variability compared to the general population (CoV of 1.21). The diabetes sample showed less variation than the depression sample, which is reflected in the CoV of 1.03 for diabetes and 1.16 for depression.



CoV: Coefficient of Variation (ratio of standard deviation and mean); data label: region identification number

Figure 1: Unadjusted regional variation in healthcare spending in the general population and the disease-approach

Variance (un)explained by demand and supply factors

Table 1 shows the model estimates for the null and the best-fitted model for all samples. The results are shown on the log-transformed scale. For all samples, the best-fitted model included demand variables as demographic variables and health status variables (both self-reported and claims data derived). Supplementary File 5 illustrates that self-reported health status explains a substantial part of healthcare spending; the total variance decreased approximately 28% on the log-scale. Supply variables were found to add small but significantly on the log-scale, to the models for depression and the general population (not for diabetes). The best-fitted depression model included distance in meters to a physical therapist. For the general population, distance in meters to GP, hospital and pharmacy were additionally included. The Pearson's correlations (Table 1) between the overall mean in the null models and predicted mean in the full models confirmed the influence of the covariates by showing coefficients of 0.12 (diabetes), 0.14 (general population) and 0.16 (depression), which was largely due to the influence of the demand variables.

Figure 2 demonstrates the proportion of the variance that is explained after fitting the models. A-priori, more than 99% of the total variance was attributed to the individual level (also see Table 1), leaving less than 1% attributed to the regional level. At the individual level, 62-63% of the variance on the log-scale was explained by demand variables. At the regional level, the covariates explained relatively more variation for the general population and the depression sample (96% and 100% respectively) than for the diabetes sample (88%).

DISCUSSION

The aim of this study was to gain more in-depth insight into regional variation in healthcare spending using prevalent chronic diseases. We used samples of patients with diabetes and depression, as we expected more homogeneity due to treatment standardization in the first compared to the latter. To our knowledge, we were the first to apply such an approach. We found indications that levels and sources of variation in healthcare spending seem to differ between disease-groups. The results showed that unadjusted regional variation in healthcare spending was smaller for the sample of patients with diabetes, than for patients with depression and the general population. Regions with above (below) average spending for the general population mostly showed above (below) average spending for diabetes and depression as well. A-priori, more than 99% of the total variance was concentrated at the individual level, leaving less than 1%

Table 1: LMM model estimates for the traditional approach and disease-based approach

			General Dutch population (n=44,694)		Diabetes (n=10,767)		Depression (n=3,735)	
			Model 0	Model 10	Model 0	Model 5	Model 0	Model 9
Variable			Beta (se)	Beta (se)	Beta (se)	Beta (se)	Beta (se)	Beta (se)
intercept (patient level)			7,64 (0,03)***	6,21 (0,03)***	8,04 (0,02)***	6,95 (0,05)***	7,72 (0,05)***	6,20 (0,08)***
age				0,01 (0,00)***		0,00 (0,00)***		0,01 (0,00)***
gender				0,03 (0,01)***		0,04 (0,01)***		0,03 (0,03)
self-reported health status	fair			0,44 (0,01)***		0,31 (0,01)***		0,60 (0,04)***
	poor			0,77 (0,02)***		0,59 (0,02)***		0,94 (0,05)***
claims data derived	*DCG 2012*	1		0,90 (0,03)***		0,68 (0,03)***		0,80 (0,14)***
		2		0,79 (0,02)***		0,54 (0,04)***		0,75 (0,08)***
		3		0,81 (0,02)***		0,60 (0,04)***		0,78 (0,10)***
		4		1,02 (0,02)***		0,81 (0,03)***		1,05 (0,07)***
		5		1,09 (0,02)***		0,86 (0,03)***		1,02 (0,09)***
		6		1,04 (0,02)***		0,92 (0,03)***		0,97 (0,08)***
		7		1,13 (0,03)***		0,99 (0,04)***		1,03 (0,13)***
		8		1,30 (0,06)***		1,06 (0,11)***		1,37 (0,21)***
		9		1,25 (0,04)***		1,08 (0,06)***		0,86 (0,18)***
		10		1,38 (0,04)***		1,18 (0,07)***		1,12 (0,17)***
		11		1,45 (0,12)***		1,35 (0,17)***		1,07 (0,48)**
		12		1,14 (0,06)***		1,04 (0,09)***		1,10 (0,29)***
		13		1,83 (0,09)***		1,68 (0,14)***		1,55 (0,32)***
		14		1,96 (0,31)***		1,49 (0,69)**		(omitted)
		15		1,99 (0,50)***		(omitted)		2,34 (0,96)**
demand	PCG 2012			0,250 (0,01)***		0,19 (0,01)***		0,22 (0,02)***

Table 1: LMM model estimates for the traditional approach and disease-based approach (continued)

			General Dutch population (n=44,694)		Diabetes (n=10,767)		Depression (n=3,735)	
			Model 0	Model 10	Model 0	Model 5	Model 0	Model 9
	Variable		Beta (se)	Beta (se)	Beta (se)	Beta (se)	Beta (se)	Beta (se)
supply	*distance to care provider in meters*							
		GP		0,00 (0,00)				
		pharmacy		0,00 (0,00)				
		hospital		0,00 (0,00)				
		physical therapist		0,00 (0,00)***				0,00 (0,00)***
re	*intercept – variance at the regional level*		0,02 (0,01)	0,00 (0,00)	0,01 (0,00)	0,00 (0,00)	0,04 (0,02)	0,00 (0,00)
	random error – variance at the individual level		1,24 (0,01)	0,75 (0,01)	0,79 (0,01)	0,48 (0,01)	1,49 (0,03)	0,92 (0,02)
Postestimation statistics	*ICC*		0,02 (0,01)	0,00 (0,00)	0,01 (0,00)	0,00 (0,00)	0,02 (0,01)	0,00 (0,00)
	AIC		136507	114297	28054	22696	12120	10351
	BIC		136533	114532	28075	22856	12139	10494
	log likelihood		-68250	-57122	-14024	-11326	-6057	-5153
	PC mean and predicted mean			0,14***		0,12***		0,16***
	PC empirical Bayes mean			0,90***		0,94***		0,70***
	total variance on the log-scale		1,54	0,57	0,62	0,23	2,22	0,85
	total variance individual level		1,54	0,57	0,62	0,23	2,22	0,85
	total variance regional level		0,00	0,00	0,00	0,00	0,00	0,00

DCG: Diagnosis Cost Group; PCG: Pharmacy-based Cost Group; GP: General Practicioner; AIC: Aikaike Information Criterium; BIC: Bayesian Information Criterion; PC: Pearson's Correlation; **: p-value < 0.05; ***: p-value < 0.01; re: random effects

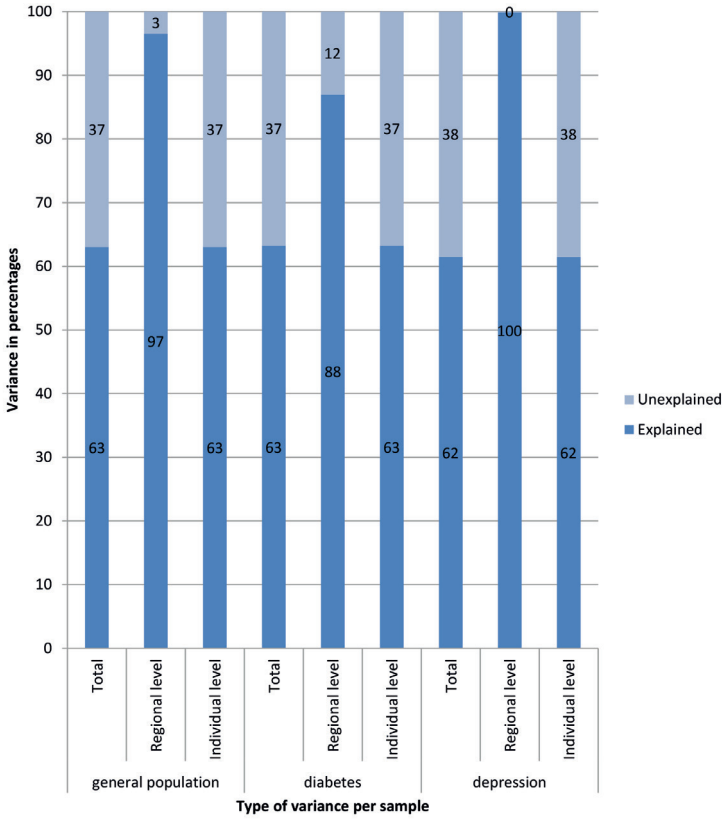


Figure 2: Variance in healthcare spending explained by covariates in the general population and the disease-approach

at the regional level. We found that demand variables explained 62-63% of the total variance in healthcare spending. Self-reported health status was the most prominent variable in the model, explaining 28% of the variation. Supply variables added nearly 0%, but significantly to the model in the general population sample (distance to GP, hospital, pharmacy and physical therapy) and the depression sample (distance to physical therapy). For the general population and the depression sample, the covariates in the model explained 96-100% of the variation at the regional level as opposed to 88% for the diabetes sample.

The finding that variation in healthcare spending is largely explained by demand factors and that the variation at the regional level is limited, is in accordance with work from e.g. Lavergne, Barer [24] and Gopffarth, Kopetsch [11]. Both studies were performed

using a traditional population-based approach and performed their studies in similar health systems as the Netherlands. Based on their results, both authors concluded that regional differences in healthcare spending do not clearly reflect inefficiencies. Therefore, Lavergne, Barer [24] suggests that policy reforms should be rather targeted at system wide efficiency improvements, than at high-spending regions. Although the results of our study similarly showed a small amount of variation that was left at the regional level, we think that unexplained variance that might indicate inefficiency cannot be ruled out based on two findings. First, at the regional level, the diabetes sample showed relatively more unexplained variance than the other samples. This indicates that variance at the regional level was caused by factors outside of the model, as for example differences in efficiency in organizations that were influenced by the disease standardization. In contrast, the other study samples showed less variability. For the depression sample, all of the variation was explained at the regional level. This might be caused by variability in disease severity or treatment differences between individuals within regions, rather than across regions. Second, at the individual level, supply variables were small, but significant in the general population and depression sample and not in the diabetes sample. Even though we found small regional effects, we believe this insight is useful, as more efforts at the regional level are expected in the near future, and therefore more variability at the regional level may be encountered in analyses of disease-groups.

Our study has several limitations. First and most important, by using the self-reported health status variable which we derived from the Dutch Public Health Monitor survey, we were limited to a subsample of the nationwide patient-level claims data. Consequently, we encountered loss of precision in our analyses as the region sample sizes were less than 30 for a maximum of three regions in the disease-groups. Additionally, the external validity of this study depends directly on that of the Dutch Public Health Monitor, which shows selection bias that was corrected for by using a weighing factor (unpublished work of Carolien vd Brink submitted to TSG). Nevertheless, we believe that for the purpose of this study, namely to show a method on how to gain insight into differences between disease-groups within and across regions, this had no substantial impact on our conclusions. Second, controlling for supply and demand factors in the analyses should be improved. For example, we were not able to capture variables that inform cultural differences between physician treatments or the level of standardization of treatments per region. Third, due to computational issues in STATA we could not retransform the log-transformed healthcare spending variable and calculate variance measures on the linear scale (see Supplementary File 3). Interpretation of the results is therefore more difficult and less precise. Consequently, the extent to which variance is explained might differ when measured on the linear scale. However, and in extension

to previous work [11, 25], we used a more complex hierarchical structure to account for the nested structure of individuals into regions and therefore avoided the disadvantages (i.e. ecological fallacy) of an aggregate design. As this study has a descriptive character and we did not aim to infer causally, we are confident our conclusions hold up in a qualitative sense.

Despite these limitations, this study contributes to the literature that aims to investigate the role of demand and supply factors in explaining geographical variation in health care spending. First, this study benefits from the homogenous Dutch healthcare system. In contrast to, for example, the US, financing structures and schemes are defined at the country level in the Netherlands. Therefore, they do not influence regional variation in healthcare spending. In addition, as opposed to studies from the U.S., Dutch claims data cover roughly 95% of the population, as private health insurance is mandatory. Second, we were able to reduce bias that results from using claims data derived health status, which was encountered in e.g. Lavergne, Barer [24] and Wennberg, Sharp [26], by including a self-reported health status variable from a large health survey. Finally, by using a disease-approach, we were able to create a more homogenous study population in advance, which enriches the understanding of causes of regional variation in spending between groups within the population. To our knowledge, we are the first to apply such an approach. It provides a novel process of modeling regional variation in healthcare spending that may be followed in similar future studies. The notion that regional variation in healthcare spending might be a composition of different variation patterns of several disease-groups is important when developing regional healthcare policies. We suggest further research to investigate whether regions are consistent over a larger set of disease-groups and specific cost data to better interpret the results research has shown this far. Moreover, in the coming years data will become available to analyze the effects of the PHM regions. Additionally, more research is needed to fill in the gaps of regional variation that remains unexplained. A suggestion is to operationalize cultural differences to include in the analysis (previously mentioned by Kopetsch and Schmitz [25]). At the individual level, cultural differences might influence healthcare utilization on the demand side and at the organizational level or physician level cultural differences may influence the supply side. In addition, a longitudinal analysis of regional variation in healthcare spending using disease-groups might be interesting to start unraveling the causal relationship instead of a more descriptive analysis as used in this study.

Conclusion

The extent to which regional variation in healthcare spending can be considered as inefficiency may differ between regions and disease-groups. Therefore, an approach

analyzing chronic diseases, in addition to the traditional approach, where the general population is studied, provides more detailed insight into the causes of regional variation in healthcare spending and identifies potential areas for efficiency improvement and budget allocation decisions.

ACKNOWLEDGMENTS

We thank Hanneke W. Drewes and Albert Wong from the Dutch National Institute of Public and the Environment for their helpful comments on a previous version of the manuscript.

REFERENCES

1. Steenkamer, B.M., et al., *Defining Population Health Management: A Scoping Review of the Literature*. Popul Health Manag, 2017;20(1):74-85.
2. Hildebrandt, H., Schulte, T., Stunder, B., *Triple Aim in Kinzigtal, Germany: improving population health, integrating health care and reducing costs of care-lessons for the UK?* J Integr Care, 2012;20(4):205-222.
3. Alley, D.E., Asomugha, C.N., Conway, P.H., Sanghavi, D.M., Accountable Health Communities - Addressing Social Needs through Medicare and Medicaid. N Engl J Med, 2016;374:8-11.
4. Lemmens, L.C., Drewes, H.W., Lette, M., Baan C,A., *Een populatiegerichte aanpak voor verbinding van preventie, zorg en welzijn: de beweging in beeld*. Ned Tijdschr Geneeskd, 2017;161:D849.
5. Drewes, H.W., Struijs, J.N., Baan C.A,. *How the Netherlands Is Integrating Health and Community Services*. NEJM Catalyst, 2016.
6. Institute of Medicine (IOM). *Variation in Health Care Spending: Target Decision Making, Not Geography*. Washington DC: The National Academies Press; 2013.
7. Fisher, E.S., Bynum, J.P., Skinner, J.S.. *Slowing the growth of health care costs--lessons from regional variation*. N Engl J Med, 2009; 360(9):849-52.
8. Skinner, J., *Chapter Two - Causes and Consequences of Regional Variations in Health Care*. In: Pauly, M.V., McGuire, T.G., Barros, P.P., editors. *Handbook of Health Economics*. Oxford: Elsevier; 2011. p. 45-93.
9. Manning, W.G., Norton, E.C., Wilk, A.S., *Explaining geographic variation in health care spending, use and quality, and associated methodological challenges*. In: *Geographic Variation in Health Care Spending and Promotion of High-Value Care – Interim Report*. Institute of Medicine. 2012.
10. Zuckerman, S., Waidmann, T., Berenson, R., Hadley, J., *Clarifying sources of geographic differences in Medicare spending*. N Engl J Med, 2010; 363(1):54-62.
11. Gopffarth, D., Kopetsch, T., Schmitz, H.. *Determinants of Regional Variation in Health Expenditures in Germany*. Health Econ, 2015;25(7):801-815.
12. Cutler, D., Skinner, J., Stern, A.D., Wennberg, D., *Physician Beliefs and Patient Preferences: A New Look at Regional Variation in Health Care Spending*. In: *The National Bureau of Economic Research (NBER) Working Paper Series*. NBER, Cambridge, 2013.
13. Newhouse, J.P., Garber, A.M., *Geographic variation in Medicare services*. N Engl J Med, 2013;368(16):1465-8.
14. Häkkinen, U., et al. *Health care performance comparison using a disease-based approach: The EuroHOPE project*. Health Policy, 2013;112(1–2):100-109.
15. Heijink, R., Engelfriet, P., Rehnberg, C., Kittelsen, S.A., Häkkinen, U., *A Window on Geographic Variation in Health Care: Insights from EuroHOPE*. Health Econ, 2015;24 Suppl 2:164-77.
16. Struijs, J.N., Baan, C.A., *Integrating care through bundled payments--lessons from The Netherlands*. N Engl J Med, 2011;364(11):990-1.
17. Brink, C.L. van den., *Gezondheidsmonitor GGD'en, CBS en RIVM*. In: *Volksgezondheid Toekomst Verkenning, Nationaal Kompas Volksgezondheid*. Bilthoven, RIVM, 2013.
18. de Boo, A., *Vektis 'Informatiecentrum voor de zorg'*. TSG Tijdschr Gezondheidswet. 2011;89(7):358-359.
19. Elissen, A., Bilo, H., Struijs, J., Van Galen, M., De Vries, R., Luijk, R., et al., *Development of a national diabetes register based on claims data to support population health management in the Netherlands* (under review).
20. van de Ven, W.P., van Vliet, R.C., Lamers, L.M., *Health-adjusted premium subsidies in the Netherlands*. Health Aff (Millwood). 2004;23(3):45-55.
21. Benyamini, Y., *Why does self-rated health predict mortality? An update on current knowledge and a research agenda for psychologists*. Psychol Health, 2011; 26(11):1407-13.

22. Zorginstituut Nederland (ZiN). *Wat is risicoverevening?* Retrieved from: https://www.zorginstituutnederland.nl/financiering/risicoverevening-zvw/wat-is-risicoverevening Accessed 1 Nov 2017.

23. Portrait, F.R., van der Galien, O., van den Berg, B., *Measuring Healthcare Providers' Performances Within Managed Competition using Multidimensional Quality and Cost Indicators*. Health Econ, 2015;25(4):408-423.

24. Lavergne, M.R., Barer, M., Law, M.R., Wong, S.T., Peterson, S., McGrail, K., *Examining regional variation in health care spending in British Columbia, Canada*. Health Policy, 2016;120(7):739-48.

25. Kopetsch, T., Schmitz, H., *Regional variation in the utilisation of ambulatory services in Germany*. Health Econ, 2014;23(12):1481-92.

26. Wennberg, D.E., Sharp, S.M., Bevan, G., Skinner, J.S., Freedman, J.O., Gottlieb, D.J., et al. *A population health approach to reducing observational intensity bias in health risk adjustment: cross sectional analysis of insurance claims*. BMJ, 2014;348:g2392.

SUPPLEMENTARY FILES

S1. Data and sources

	Variable	Operationalization	Type	Data source
Supply	Provider concentration	Number of providers per 1000 population	ratio	Netherlands Institute for Health Services Research (NIVEL), the National Institute of Public Health and the Environment (RIVM) and the Health Care Inspectorate (IGZ).
		GP		
		Pharmacy		
		Hospital		
		Physiotherapist		
Demographics	Age	Age in years	ratio	Statistics Netherlands
	Gender	Percentage female		Statistics Netherlands

	Variable	Operationalization	Type	Data source
Health status	Self-reported health status	How do you perceive your health in general?	categorical (good - fair - poor)	Dutch Health Monitor survey 2012
	Claims data health status	Number of Pharmacy Cost Groups in 2008	ratio	Vektis 2008/2012
		Number of Pharmacy Cost Groups in 2012	ratio	
		Type of Diagnosis-based Cost Group in 2008	categorical (none, 1-13)	
		Type of Diagnosis-based Cost Group in 2012	categorical (none, 1-15)	
		Number of chronic illnesses	categorical (none, 1-3)	Claims data Vektis 2012
Supply	Accessibility	Distance to nearest provider in meters	ratio	Statistics Netherlands
		GP		
		Pharmacy		
		Hospital		
		Physiotherapist		
Outcome variable	Healthcare spending	Natural log of euro's spend on healthcare 2012	ratio	Claims data Vektis 2012
		Total curative healthcare spending		
		GP		
		Pharmacy		
		Specialized care		
		Physiotherapist		
		Mental health care		

GP: General Practitioner

S2. Sample selection

Sample:	**General population**		**Diabetes**		**Depression**	
Reason	*included*	*excluded*	*included*	*excluded*	*included*	*excluded*
Participants of the Dutch Health Monitor survey	387195					
Link available with claims data	362905	24290				
Who are inhabitants of one of the 18 PHM sites	138732	224173				
Without missings in explanatory variables	56277	82455				
Without missings in dependent variable	45979	10298				
Without 1% outliers in dependent variable (at the top)	44695	1284				
Without negative values for the dependent variable	44694	1				
Study sample total population	**44694**		44694		44694	
Study sample diabetes			**10767**	33927		
Study sample depression					**3735**	40959

S3. Statistical analyses: model selection

The selection of a method for the analysis of healthcare spending data is not straightforward, as difficulties arise due the distribution of healthcare spending, which is known to be non-negative, skewed and heavy-tailed. However, in general, generalized linear models (GLM) are used as it deals with the skewed distribution of healthcare spending [1, 2]. GLM requires a mean function and a variance function to be specified beforehand based on the probability distribution of the dependent variable. The mean function relates the mean to some function of covariates and the variance function relates the mean to the variance on the raw scale. For spending data, options for distributions typically include (Generalized) Gamma, Poisson, inverse Gaussian and Gaussian distributions. Link functions include identity, inverse square or power. A major advantage of GLM over Ordinary Least Squares (OLS) using a log-transformed response is that the expected value of the response on the original response scale can be retrieved straightforwardly, without the need for a specific retransformation.

As the gamma family and log link is often used in previous literature for handling data on healthcare spending (e.g. [3, 4]), we used this distribution and link as a starting point. For the gamma distribution the variance is a square function of the mean, which can be checked by performing the modified Park test [2]. We performed the modified Park test by fitting the GLMM for the log of the squared residuals using the log of the predicted y. The coefficient of the log of the predicted y indicates the functional form of the underlying distribution: beta ≈ 0 indicates Gaussian distribution; beta ≈ 1 Poisson; beta ≈ 2 gamma; and beta ≈ 3 indicates inverse Gaussian or Wald distribution. However, due to computational problems, STATA was not able to fit the GLMM with gamma family and log link for the large number of observations and variables in addition to the nested structure. However, in much smaller samples the model with gamma family and log link did converge. Therefore, we consecutively tried fitting Poisson, inverse Gaussian and Gaussian distributions. Due to aforementioned computational problems, STATA was only able to fit the model using the Gaussian distribution. These sorts of computational problems are seen before in e.g. Tsiachristas and Rutten-van Mölken [5] and Mohnen, Molema [4]. In order to refrain from losing too much statistical power we chose a different model instead; he only link option in combination with Gaussian distribution is identity (i.e. Linear Mixed Models (LMM))

A disadvantage of using the Gaussian distribution is that the results are sensitive to extreme values, can produce out-of-range values as negative predicted spending, and is likely to be inefficient for small to medium sample sizes since the underlying distribution in reality is not normal [6]. To accommodate a Gaussian distribution, we decided to

log-transform the outcome variable. Modeling the log-transformed variable instead of the raw cost variable may solve the problem of skewness and allows for comparison on the log-transformed scale. Although the results are more precise and robust by the transformation, a major drawback is that the interpretation of the results are not straightforward. The comparison of the means of the log-transformed scale cannot be translated directly to comparison of means on the raw scale by inverting the log, as the log of the expected value does not equal the expected value of the log (i.e. $\ln(E[y|x]) \neq E[\ln(y|x)]$). Appropriate solutions for this problem (e.g. Duan's non-parametric smearing factor [7]) may be applied to transform the contribution of the covariates. However, variance estimates remain on the log-scale. As the results of this study focuses on the variances and therefore still had to be displayed on the log-scale, we decided not to retransform the covariates.

References

1. Manning, W.G., Basu, A., and Mullahy, J., *Generalized modeling approaches to risk adjustment of skewed outcomes data.* J Health Econ, 2005. 24(3): p. 465-88.
2. Manning, W.G. and Mullahy, J., *Estimating log models: to transform or not to transform?* J Health Econ, 2001. 20(4): p. 461-94.
3. Iversen, T., et al., *Comparative Analysis of Treatment Costs in EUROHOPE.* Health Econ, 2015. 24: p. 5-22.
4. Mohnen, S.M., et al., *Cost Variation in Diabetes Care across Dutch Care Groups?* Health Serv Res, 2016.
5. Tsiachristas, A. and Rutten-van Mölken, M., *Exploring the variability of patient costs in disease management programs: A hierarchical modelling approach.* Appl Econ, 2014. 46(9): p. 940-951.
6. Mihaylova, B., et al., *Review of Statistical Methods for Analysing Healthcare Resources and Costs.* Health Econ, 2011. 20(8): p. 897-916.
7. Duan, N., *Smearing Estimate: A Nonparametric Retransformation Method.* J Am Stat Assoc, 1983. 78(383): p. 605-610.

S4. Descriptive statistics of healthcare spending per region

General population sample

region	n		total		gp		hospital		pharmacy		mental healthcare	
	reality	sample	mean	sd	mean	sd	mean	sd	mean	sd	mean	sd
1	170247	2898	3886	4319	215	145	1955	3395	878	1042	151	966
2	648250	1967	3277	4269	202	140	1892	3436	628	799	93	995
3	104883	786	4024	5132	189	122	2129	3849	796	1246	87	981
4	269786	4032	4366	4994	203	127	2300	3984	886	1055	107	924
5	182888	660	3845	5110	216	166	2217	3989	771	1298	105	771
6	418896	977	3315	4494	175	123	1797	3405	750	1689	119	867
7	274957	2003	3655	4874	195	131	2219	4076	684	990	110	903
8	518308	6975	4104	4905	194	144	2292	3948	786	1081	104	974
9	113025	238	3473	4803	174	91	1873	3586	830	1551	160	1170
10	183053	2256	4413	5007	221	140	2307	3756	905	1085	170	1459
11	87667	779	4316	5077	197	125	2551	4172	896	1331	130	1078
12	687486	7470	3688	4553	185	118	1979	3655	790	1049	67	837
13	126698	2601	4163	5106	210	144	2433	4165	806	1116	76	762
14	60680	299	3491	4513	198	123	1979	3817	707	1159	64	400
15	40529	88	3131	4117	159	95	1696	2905	750	1591	14	70
16	293083	4231	3936	4823	201	127	2323	3968	753	1077	89	706
17	13692	348	4172	4750	225	137	2251	3803	891	918	45	360
18	1034788	6086	4180	4976	201	122	2448	4087	860	1164	103	896
total	5228916	44694	3973	4816	199	132	2215	3870	806	1104	101	923

Diabetes sample

region	**n**		**total**		**gp**		**hospital**		**pharmacy**		**mental healthcare**	
	reality	sample	mean	sd	mean	sd	mean	sd	mean	sd	mean	sd
1		687	4723	4298	240	152	2074	3351	1181	1002	125	800
2		410	4135	4574	220	145	2022	3527	964	1214	16	217
3		187	4814	4873	184	106	2485	3665	974	902	16	114
4		1104	5277	5348	217	130	2515	4208	1128	927	104	1163
5		95	4591	4616	234	128	2143	3004	1086	1283	165	1145
6		185	3884	4238	187	131	1804	3180	870	984	129	836
7		393	4448	4765	209	154	2292	3883	908	942	147	1468
8		1632	5010	5208	209	151	2598	4172	1020	956	76	726
9		45	3691	3022	166	71	1357	1767	1333	1955	20	135
10		550	4886	4859	235	135	2262	3620	1272	1445	63	723
11		231	4949	5275	212	146	2541	4091	1095	1089	189	1564
12		1907	4369	4558	197	126	2044	3555	1052	1197	35	523
13		589	5244	5276	237	174	2725	4271	1108	1175	88	1242
14		73	4114	4824	227	139	2039	4230	802	716	134	689
15		15	3493	3467	177	147	1417	2957	789	639	0	0
16		940	4586	4680	209	138	2343	3702	1002	1212	84	767
17		75	4695	4345	242	159	2150	3083	1052	770	7	41
18		1649	5067	5274	215	131	2802	4350	1103	1019	83	626
total		10767	4786	4938	214	140	2389	3914	1068	1097	79	831

Depression sample

region	n		total		gp		hospital		pharmacy		mental healthcare	
	reality	sample	mean	sd	mean	sd	mean	sd	mean	sd	mean	sd
1		300	4491	4590	219	142	1936	3298	896	906	657	1859
2		184	3415	4812	199	126	1705	3576	629	718	456	2203
3		75	4530	5520	179	106	2484	4581	711	774	356	1767
4		330	4960	5261	216	156	2229	3660	987	1013	628	2092
5		61	4383	5546	245	189	2308	4222	802	858	596	1836
6		97	3366	4110	208	134	1599	3003	690	1015	458	1605
7		138	3804	5147	216	182	1685	3209	802	983	654	2591
8		502	4971	5633	209	159	2404	4007	864	919	752	2815
9		37	2950	4543	146	52	1544	3563	541	847	407	1301
10		203	5034	6419	222	143	1897	3730	989	1082	1005	3734
11		69	5459	5514	239	177	2524	3492	1259	1896	630	2102
12		647	4032	4857	195	120	1779	3139	892	1176	451	2210
13		212	4823	5349	232	148	2548	4432	902	959	330	1021
14		24	3474	4530	181	67	1251	1687	1314	3191	198	555
15		12	2941	4538	161	99	1758	3651	433	479	101	172
16		402	4746	5515	220	150	2553	4416	814	1307	550	1680
17		15	5799	3699	306	208	2715	3347	1147	821	856	1503
18		427	4592	5031	212	134	2249	3747	958	1319	623	1932
total		3735	4493	5233	211	144	2125	3736	881	1126	588	2216

GP: general practicioner; sd: standard deviation

S5. Complete LMM model estimates for the general population

			Model 0	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	Model 10
Variable			**B (se)**	**B (se)**	**B (se)**	**B (se)**	**B (se)**	**B (se)**	**B (se)**	**B (se)**	**B (se)**	**B (se)**	**B (se)**
fixed effects	intercept (patient level)		7,64 (0,03)***	7 (0,04)***	6,37 (0,03)***	6,30 (0,04)***	6,18 (0,03)***	6,18 (0,03)***	6,20 (0,03)***	6,19 (0,03)***	6,20 (0,03)***	6,21 (0,03)***	6,21 (0,03)***
demand	pension age			0,02 (0,00)***	0,01 (0,00)***	0,01 (0,00)***	0,01 (0,00)***	0,01 (0,00)***	0,01 (0,00)***	0,01 (0,00)***	0,01 (0,00)***	0,01 (0,00)***	0,01 (0,00)***
	gender			-0,02 (0,01)*	-0,05 (0,01)***	0,02 (0,01)	0,06 (0,01)***	0,04 (0,01)***	0,03 (0,01)***	0,03 (0,01)***	0,03 (0,01)***	0,03 (0,01)***	0,03 (0,01)***
self-reported health status													
	fair				0,71 (0,01)***			0,44 (0,01)***	0,44 (0,01)***	0,44 (0,01)***	0,44 (0,01)***	0,43 (0,01)***	0,44 (0,01)***
	poor				1,30 (0,02)***			0,77 (0,02)***	0,77 (0,02)***	0,77 (0,02)***	0,77 (0,02)***	0,77 (0,02)***	0,77 (0,02)***
claims data derived health status													
	DCG*	1				0,45 (0,03)***	1,00 (0,03)***	0,90 (0,03)***	0,90 (0,03)***	0,90 (0,03)***	0,90 (0,03)***	0,90 (0,03)***	0,90 (0,03)***
		2				0,28 (0,04)***	0,89 (0,02)***	0,79 (0,02)***	0,79 (0,02)***	0,79 (0,02)***	0,79 (0,02)***	0,79 (0,02)***	0,80 (0,02)***
		3				0,34 (0,05)***	0,90 (0,02)***	0,81 (0,02)***	0,81 (0,02)***	0,81 (0,02)***	0,81 (0,02)***	0,81 (0,02)***	0,81 (0,02)***
		4				0,56 (0,05)***	1,11 (0,02)***	1,02 (0,02)***	1,02 (0,02)***	1,02 (0,02)***	1,02 (0,02)***	1,02 (0,02)***	1,02 (0,02)***
		5				0,42 (0,05)***	1,21 (0,02)***	1,09 (0,02)***	1,09 (0,02)***	1,09 (0,02)***	1,09 (0,02)***	1,09 (0,02)***	1,09 (0,02)***
		6				0,58 (0,06)***	1,19 (0,02)***	1,04 (0,02)***	1,04 (0,02)***	1,04 (0,02)***	1,04 (0,02)***	1,04 (0,02)***	1,04 (0,02)***

Variable		Model 0 B (se)	Model 1 B (se)	Model 2 B (se)	Model 3 B (se)	Model 4 B (se)	Model 5 B (se)	Model 6 B (se)	Model 7 B (se)	Model 8 B (se)	Model 9 B (se)	Model 10 B (se)
	7				0,95 (0,08)***	1,28 (0,03)***	1,13 (0,03)***	1,13 (0,03)***	1,13 (0,03)***	1,13 (0,03)***	1,13 (0,03)***	1,13 (0,03)***
	8				1,05 (0,09)***	1,50 (0,06)***	1,30 (0,06)***	1,30 (0,06)***	1,30 (0,06)***	1,30 (0,06)***	1,30 (0,06)***	1,30 (0,06)***
	9				0,95 (0,09)***	1,41 (0,04)***	1,25 (0,04)***	1,25 (0,04)***	1,25 (0,04)***	1,25 (0,04)***	1,25 (0,04)***	1,25 (0,04)***
	10				0,72 (0,12)***	1,54 (0,04)***	1,38 (0,04)***	1,38 (0,04)***	1,38 (0,04)***	1,38 (0,04)***	1,38 (0,04)***	1,38 (0,04)***
	11				0,99 (0,19)***	1,70 (0,12)***	1,45 (0,12)***	1,45 (0,12)***	1,45 (0,12)***	1,45 (0,12)***	1,45 (0,12)***	1,45 (0,12)***
	12				0,89 (0,08)***	1,33 (0,06)***	1,14 (0,06)***	1,14 (0,06)***	1,14 (0,06)***	1,14 (0,06)***	1,14 (0,06)***	1,14 (0,06)***
	13				0,94 (0,14)***	2,06 (0,10)***	1,83 (0,09)***	1,83 (0,09)***	1,83 (0,09)***	1,83 (0,09)***	1,82 (0,09)***	1,83 (0,09)***
	14					2,05 (0,32)***	1,96 (0,31)***	1,96 (0,31)***	1,96 (0,31)***	1,96 (0,31)***	1,96 (0,31)***	1,96 (0,30)***
	15					2,28 (0,52)***	2,00 (0,50)***	2,00 (0,50)***	2,00 (0,50)***	1,99 (0,50)***	2,00 (0,50)***	1,99 (0,50)***
PCG 2008					0,33 (0,01)***							
PCG 2012						0,32 (0,01)***	0,25 (0,01)***	0,25 (0,01)***	0,25 (0,01)***	0,25 (0,01)***	0,25 (0,01)***	0,25 (0,01)***

Variable	Model 0 B (se)	Model 1 B (se)	Model 2 B (se)	Model 3 B (se)	Model 4 B (se)	Model 5 B (se)	Model 6 B (se)	Model 7 B (se)	Model 8 B (se)	Model 9 B (se)	Model 10 B (se)
supply											
distance to care provider in meters											
GP							0,00 (0,00)***				0,00 (0,00)***
pharmacy								-0,00 (0,00)***			0,00 (0,00)***
hospital									0,00 (0,00)***		0,00 (0,00)***
physical therapist										0,00 (0,00)***	0,00 (0,00)***
Random effects											
intercept – variance at the region level	0,02 (0,01)	0,01 (0,00)	0,00 (0,00)	0,01 (0,00)	0,01 (0,00)	0,00 (0,00)	0,00 (0,00)	0,00 (0,00)	0,00 (0,00)	0,00 (0,00)	0,00 (0,00)
random error – variance at the individual level	1,24 (0,01)	1,20 (0,01)	1,02 (0,01)	1,10 (0,01)	0,81 (0,01)	0,75 (0,01)	0,75 (0,01)	0,75 (0,01)	0,75 (0,01)	0,75 (0,01)	0,75 (0,01)
Post estimation statistics											
ICC	0,02 (0,01)	0,01 (0,00)	0,00 (0,00)	0,01 (0,00)	0,01 (0,00)	0,01 (0,00)	0,01 (0,00)	0,01 (0,00)	0,01 (0,00)	0,01 (0,00)	0,00 (0,00)
AIC	136507	135161	127804	131208	117692	114311	114305	114310	114308	114293	114297
BIC	136533	135205	127865	131373	117875	114512	114514	114519	114517	114502	114532
log likelihood	-68250	-67576	-63895	-65585	-58825	-57133	-57129	-57131	-57130	-57123	-57122
likelihood ratio test		1349***	7361***	3981***	17501***	20886***	8***	3*	4**	20***	22***
PC mean and predicted mean		0,47***	0,45***	0,26***	0,15***	0,14***	0,14***	0,14***	0,14***	0,14***	0,14***
Total variance on the log scale	1,54	1,45	1,04	1,21	0,66	0,57	0,57	0,57	0,57	0,57	0,57

B: beta; se: standard error; AIC: Aikaike Information Criterion; BIC: Bayesian Information Criterion; PC: Pearson's Correlation; GP: general practitioner; PCG: Pharmacy-based Cost Group; DCG: Diagnosis Cost Group; *: p-value < 0.10; **: p-value < 0.05; ***: p-value < 0.01; #:DCG in model 3 is for the year 2008. As from model 4 the year

3


4

CHAPTER 4

Measuring value in maternity care. A first attempt for Dutch Maternity Care Networks

Based on
De Vries EF, Over EAB, Wong A, Baan CA, Heijink R, Struijs JN
Measuring value in maternity care.
A first attempt with Dutch Maternity Care Networks.
(Submitted)

ABSTRACT

Objective: To explore whether Dutch Maternity Care Networks (MCNs) show differences in value using observational data.

Data sources/study setting: We linked data from three nationwide administrative registers; data on maternal and neonatal health, quality of maternity care and case-mix variables, claims data and municipality registry data for low-risk pregnancies in the Netherlands in 2011 through 2016 (n=297,960).

Study design: The variation of six value-indicators, defined as the rate of adverse health outcomes (outputs) relative to the rate of services (inputs), was measured across 91 MCNs. To explore the extent to which the value-indicators are able to reflect the actual value, additional analyses were performed. We assessed the association between inputs and outputs, the consistency across inputs and across outputs and the persistence of inputs and outputs over time. All estimates were modeled through multilevel modeling with adjustments for mother, father and child demographics, pregnancy related clinical factors, previous health spending and time.

Principal findings: We observed substantial variation in all six value-indicators, comprising combinations of inputs (caesarean sections, epidural analgesia and labor induction) and outputs (Apgar score lower than 7 after 5 minutes, hemorrhage post-partum and serious perineal damage), across MCNs after adjusting for case-mix. Additional analyses showed that MCNs with more (less) low-value services did not show more (less) adverse outcomes, except for 'caesarean section' and 'epidural' with 'low Apgar score', which were weakly negatively correlated. Furthermore, positive correlations emerged for: 1) low Apgar score after caesarean section, after epidural and after labor induction, 2) hemorrhage postpartum after labor induction and after caesarean section (consistency). Persistence over time was found for all inputs (caesarean section, epidural, labor induction).

Conclusions: We found substantial variation across MCNs for the six value-indicators. The additional analyses indicated that some of the outputs may have measured a similar concept (consistency), and that the inputs (i.e. low-value care indicators) may have captured a part of value (persistence). Yet, despite access to extensive set of case-mix variables, population heterogeneity clouded our results. Future research should further deal with the limitations of observational data and refine outcome measures that matter to the population of pregnant women in order to truly capture value at the MCN level and contribute to value-based maternity care.

INTRODUCTION

As a reaction to sustainability issues of health systems in Western countries, regional networks of payers and professionals [1, 2] increasingly seek to improve *value*. Higher-value care means that the ratio between health outcomes and spending increases [3]. An example of such networks are the Dutch maternity care networks (MCNs), generally consisting of obstetricians (organized in medical specialist partnerships) in hospitals, several independent midwife organizations and community maternity caregivers [4]. Insight into the performance, with respect to value, within and across MCNs, can assist these networks in monitoring progress towards value, in developing policies and interventions, and in making budget allocation decisions [5].

A common method to analyze the performance of regional networks is to study the variation of outcomes across regions. The reasoning is that variation, which is not caused by medical need, is unwarranted and, therefore, points at inefficiencies [6, 7]. To date, most studies have examined the variation in health spending (e.g. [8-10]), utilization (e.g. [6, 11-13]) or health outcomes (e.g. [14]), separately. However, despite the increasing attention for the concept of value, studies that aim to measure performance of organizations with respect to value (i.e. comprising both outcomes and inputs), are scarce [15-22]. A few studies assessed the value of care, implicitly, by specifying *outputs* and *inputs* relative to each other, or explicitly by using a production function framework to develop a value-indicator [16]. Those studies assessed the value of care across hospitals [16, 20, 22], mental care providers [17], diabetes care groups [18] or physical therapists [19]. We found one study assessing the value of care across regional provider networks for high-cost seniors [21]. As far as we know, an analysis of the performance of maternity care networks (MCNs) with respect to the value of care has not been performed before.

Moreover, previous studies that, explicitly or implicitly, analyzed the variation in value used *health spending* as input, which result in endogeneity problems [20]. Endogeneity occurs because health spending, as is registered by insurers and alike, includes spending on the initial treatment as well as on the complications that may be caused by the treatment itself. Moreover, healthcare spending may also reflect the health status of the patients next to the intensity with which patients are treated by health care providers, even when an attempt to adjust for case-mix has been made. This affects not only value analyses, but also studies analyzing the variation in healthcare spending separately. Although most studies using health spending acknowledge endogeneity bias in the limitations section of their papers (e.g. [5, 18, 22]), only a few studies actively attempt to deal with this problem.

This study aims to explore value in Dutch maternity care, by (1) selecting several maternal care services, of which the utilization reflects the intensity with which patients are treated (*inputs*), (2) selecting several neonatal or maternal health outcomes that may be affected by those services (*outputs*), and (3) assessing to what extent Dutch MCNs show variations in the relationship between the services and outcomes. In the remainder of this article, we define 'value-indicators' as specific combinations of input and output as derived from Dutch observational data, that potentially reflect how more input will affect the output. In this study, we propose six value-indicators across MCNs. By selecting the utilization of *low-value* services, we can largely avoid problems with endogeneity that would have come with considering health care spending as an input. Low-value services are services that provide little to no benefits, or even cause harm. Previous studies (e.g. [23, 24]) have demonstrated the benefits of using low-value services over aggregate measures such as spending [12]. Not only do they allow for pinpointing exact sources of overuse, but they also provide a sense of how much health care resources can be saved or redirected, by comparing the most efficient and least efficient regions.

We further examine the extent to which our proposed value-indicators reflect actual value, by a series of additional analyses regarding the underlying inputs and outputs of the value-indicators. First, we assessed the association between the inputs and the outputs per value-indicator to explore their relation. When this relationship is highly variable and/or noisy, the proposed value-indicator and the observational data may perhaps not be suitable. Second, we measure the consistency across inputs and outputs in order to assess whether they measure the same concept. If the consistency is low, it indicates that underlying value is multifaceted and not easily measured by just one indicator. Third, we study the persistence of the inputs and outputs over time. The presence of persistency over time would suggest that we might capture (a part of) value across MCNs, as we do not expect 'true value' at the MCN level to change drastically over one year in the majority of the cases.

METHODS

Data sources

We linked data from three nationwide administrative registers; Perined (maternal and neonatal health, quality of maternity care and case-mix variables), Vektis (claims data) and Statistics Netherlands (case-mix variables) for births in 2011 through 2016.

Perined is a perinatal registration center that collects data from independent midwifes, obstetricians, obstetric active general practitioners, and neonatologists. Perined collects information on pregnant women (e.g. age, parity, medical history) and information on the pregnancy, labor and quality of care (e.g. gestational age, way of delivery, pain relief, maternal and infant health outcomes). Perined covers information on 99.7% of all pregnancies with births >22 weeks gestational age in the Netherlands [25]. We used data from pregnancies that led to a live birth or stillbirth of >22 weeks gestational age in 2011 through 2016.

Vektis collects individual claims data from all Dutch health insurers under the Health Insurance Act. The dataset of Vektis covers all claims data of more than 99% of the Dutch citizens [26]. We used detailed claims data of mothers for whom any claims on maternity care were made during 2011-2016. In addition, we used claims data of children with birth dates in 2011-2016.

Statistics Netherlands gave access to several case-mix variables: month and year of birth of the mother, father and child, gender of the child, ethnicity, household income and highest attained educational level.

Data linkage and sample

Statistics Netherlands provided a highly secured remote access platform on which we were able to safely link the data at the individual level using a pseudonimized linkage number to adhere to privacy legislation [27]. Each dataset contained unique encoded identifiers at the level of the mother, father or child. In order to link the data at the pregnancy level, data on both mothers and children needed to be assigned a pregnancy identification number. The Appendix elaborates on the specifics of the data linkage process.

We included pregnancies with date of birth in 2011 through 2016. Only low-risk pregnancies, defined as nulliparous term singleton vertex (NTSV[1]) with live births or stillbirths were included to limit bias due to population heterogeneity [28]. We excluded pregnancies with a Neonatal Intensive Care Unit (NICU)-admission, because these pregnancies were solely attributed to the nine MCNs with such a unit. In addition, we excluded pregnancies that could not be attributed to a MCN, and pregnancies that were

1 Pregnancies of the first born child with gestational age between 37 weeks and 42 weeks being a single birth (no twins) in head presentation.

attributed to a MCN for which the total number of pregnancies was too low (threshold <2000 pregnancies attributed to one MCN in 2011-2016).

Measuring the value of maternity care

As we aimed to measure value validly, we chose inputs and outputs based on the following conditions: 1) the input has to reflect one health service that occurs before the output, 2) the input has to be a low-value service with literature evidence showing wide variations across care providers, 3) there must be a possibility for the input to have a clinical relationship with the output (i.e. the utilized service could possibly prevent the adverse outcome, if used appropriately, or could possibly induce the adverse outcome if not used appropriately), which is supported with literature evidence and obstetrics gynecologist expert opinion. Only indicators for which it is not possible for the input to have a clinical relationship with the output are removed. 4) the output has to be a primary neonatal or maternal outcome in integrated maternity care.

Table 1 shows an overview of the value-indicators that we use in this study, specifying the numerator (output) and denominator (input) and the rationale of an association between input and output. Below, we briefly discuss the rationale of the inputs and outputs separately.

Inputs: low-value services in low-risk pregnancies

Several studies showed that wide variations of rates of caesarean sections (c-sections) exist across practitioners and hospitals, which cannot be explained by differences in medical need (e.g. [28, 34, 42-44]). Other low-value services in low-risk pregnancies for which wide variations exist are (routine) epidural analgesia (epidural) [28, 45] and non-medically induced induction of labor (induction) [44-47].

Outputs: neonatal and maternal health outcomes

In the absence of Patient Reported Outcome Measures (PROMS) in our dataset, we selected neonatal and maternal health outcomes as deemed relevant in integrated care by the Dutch Working Group on Indicators for Integrated Care [48] that were present in our observational dataset. The outcomes that we selected are generally referred to as the Adverse Outcome Index-5 (AOI-5) [49, 50]. Three of these five indicators reflect neonatal health outcomes; perinatal mortality between 22 weeks of gestational age and seven days after birth, NICU-admission and Apgar score <7 after 5 minutes (low Apgar score). The other two indicators reflect maternal health outcomes: hemorrhage post-partum (hemorrhage) ≥1000 ml and 3rd and 4th degree perineal tear for vaginal births (perineal damage).

Table 1: Operationalization of the value-indicators*

Value-indicator name	Numerator (output)	Denominator (input)	Rationale of association between input and output
Low Apgar score after C-section	Deliveries with a newborn with an Apgar score <7 after 5 minutes	Low-risk pregnancies (NTSV-group) with C-section	Timely and justified (e.g. in case of fetal complications) C-sections can improve neonatal outcomes (e.g. mortality, Apgar score) [29, 30]. In the absence of fetal complications, vaginal deliveries are considered to be at least as save as C-sections [29, 31] and, therefore, preferred over C-sections.
Hemorrhage after C-section	Deliveries with hemorrhage postpartum (HPP=> 1000 ml)	Low-risk pregnancies (NTSV-group) with C-section	Hemorrhage postpartum may occur after C-section [32, 33] [34].
Perineal damage after C-section			The association between C-section and perineal damage is not possible in this context as the perineal damage is specifically registered in vaginal births . Therefore, we excluded this indicator from further analysis
Low Apgar score after epidural	Deliveries with in a newborn with an Apgar score <7 after 5 minutes	Low-risk pregnancies (NTSV-group) with epidural analgesia	Epidural analgesia is associated with longer duration of delivery which, in turn, is associated with higher rates of C-section which may have an impact on neonatal health outcomes (see above). Yet, there are also studies showing that epidural analgesia has no impact on neonatal outcomes [35] [36].
Hemorrhage after epidural			To date, there is no empirical evidence that epidural analgesia has an association with hemorrhage postpartum [37]. Therefore, we excluded this indicator from further analysis.
Perineal damage after epidural			To date, there is no empirical evidence that epidural analgesia has an association with perineal damage [38]. Therefore, we excluded this indicator from further analysis.

Table 1: Operationalization of the value-indicators*

Value-indicator name	Numerator (output)	Denominator (input)	Rationale of association between input and output
Low Apgar score after labor induction	Deliveries with in a newborn with an Apgar score <7 after 5 minutes	Low-risk pregnancies (NTSV-group) with labor induction	Labor induction is associated with higher rates of C-section [39]. While in case of fetal complications, C-sections may have a positive association with neonatal health outcomes, other studies show no improvement of neonatal health outcomes after induction [40].
Hemorrhage after labor induction	Deliveries with hemorrhage postpartum (HPP=> 1000 ml)	Low-risk pregnancies (NTSV-group) with labor induction	Labor induction is associated with higher rates of hemorrhage postpartum [34, 40].
Perineal damage after labor induction	Percentage of deliveries with serious perineal damage (stage 3 or 4)	Low-risk pregnancies (NTSV-group) with labor induction	There are indications that labor induction is associated with increased serious perineal damage [41].

HPP: Hemorrhage PostPartum; NTSV: Nulliparous Term Singleton, Vertex presentation group; NICU: Neonatal Intensive Care Unit; C-section: Caesarean Section. Value indicators using perinatal mortality or NICU-admission were excluded because of too little observations and that deliveries with NICU-admission are attributed to MCNs with a NICU respectively. *Value-indicators using perinatal mortality or NICU-admission were not discussed in this section, because they were excluded (perinatal mortality had too little observations and deliveries with NICU-admission are attributed only to MCNs with such a unit available).

Case-mix variables

Based on previous research (e.g. [5, 18, 28, 42, 44, 51]), and depending on the specific input or output, we adjusted for a combination of the following case-mix variables. We included age of the mother at the time of birth, highest attained educational level, ethnic background, degree of urbanization based on the 4-digit postal code of the mother, household income, gestational age, birth weight, gestational hypertension and diabetes, history of miscarriages, maternal health spending in the second year before pregnancy, paternal age >45 years [52] and gender of the child.

Statistical analyses

First, we summarized unadjusted descriptive statistics for the inputs, outputs and case-mix variables at the MCN level (mean, minimum and maximum per pregnancy per MCN). Second, we fitted generalized mixed effects logistic models to estimate the variation at the MCN level, that is adjusted for case-mix and random errors at the pregnancy level. Because of nonrandom attributing of pregnancies to MCNs, the unadjusted variation would reflect differences in case-mix rather than MCN performance. Allowing for random effects instead of fixed effects is more efficient and limits the consequences of overfitting as smaller MCNs will shrink to zero (i.e., the average). The random effects at the MCN level are considered to reflect performances across MCNs, as it shows the variation between MCNs that is not due to the case-mix variables and the random error at the pregnancy level [53]. We specified the following equation for the probability that an input or an output occurs:

$$\Pr\left(y_{ij} = 1 \,\middle|\, x_i, v_j\right) = \frac{\exp\left(\sum_m \beta_m x_{mi} + v_j\right)}{1+\exp\left(\sum_m \beta_m x_{mi} + v_j\right)} \quad (1),$$

where y_{ij} is the binary-valued input or output of pregnancy *i* in MCN *j*, β_m is the m-th fixed effect of pregnancy level characteristics; x_{mij} is the m-th case-mix variables at the pregnancy level including fixed effects for *year* of birth; v_j is the parameter of our interest, the random intercept at the MCN level with $v_j \sim N(0, {\sigma_v}^2)$, and ε_{ij} is the residual error at the pregnancy level, which is distributed as logistic with mean 0 and variance $\pi^2/3$ and is independent from v_j. All inputs and outputs were fitted separately using backward elimination. That means that we first fitted all variables, one at the time. Variables for which the p-value was <0.3, were selected for the most extensive model. Then, the least significant variable (for alpha <0.05) was dropped until all variables contributed significantly to the model. The outputs were fitted on the study sample for which the input applied to (e.g. low Apgar score only after a c-section was performed, low Apgar score only after epidural was used, low Apgar score only after labor induction, and so on). Third, we described the variance in value across MCNs by plotting the inputs

against the outputs at the logit link scale (random MCN effects only). To increase interpretability we showed the adjusted rates (mean, minimum and maximum) on the response scale (which are essentially percentages derived through taking the inverse logit of the linear predictor function, as described between brackets in (1)). We further assessed whether the MCNs exhibit distinct profiles of value by calculating whether the random effects differed from zero for the inputs and the outputs and/or both, through studying the Empirical Bayes estimates of the random effects and their variances. Data management and analyses were performed using Stata software version 15.1.

Additional analyses

To explore whether we truly grasped the concept of value, we performed three additional analyses. First, to gain insight into the extent to whether high (or low) performing MCNs on a specific input also performed high (or low) on a specific output, the graphs were inspected and Spearmans' correlation coefficient (rho) was calculated between each input and output. Second, to assess whether the value-indicators might measure the same concept, we checked the consistency across each possible combination of inputs, and across each possible combination of outputs, using Spearmans' correlation coefficient (rho). Third, we assessed the persistence of the inputs and outputs over time. The presence of persistency over time would suggest that we may have captured (a part of) the concept of value across MCNs, as we do not expect 'true value' at the MCN level to change drastically over one year. In order to do so, we fitted new models. We repeated the procedure for the estimation of the random effects, but this time, we estimated the models for each year separately. Per indicator, the association between sequent years was calculated using Spearmans' correlation coefficient (rho) and assessed to what extent they persisted over the years.

All observed Spearman correlation coefficients (rho) were interpreted as negligible (0.00 – 0.10), weak (0.10-0.39), moderate (0.40-0.69), strong (0.70-0.89) or very strong (0.90-1.00) [54]. 95% confidence intervals were bootstrapped with 1000 replications. Statistically significant correlations that were moderate, strong or very strong were interpreted to point towards consistency or persistence.

RESULTS

Study characteristics

Our original dataset contained 834,234 pregnancies over 2011 through 2016. We excluded pregnancies with a NICU-admission (n=2,871), pregnancies that were not

low-risk (n= 533,403), pregnancies that could not be attributed to a MCN (n=5,936), or were attributed to a MCN for which the numbers of pregnancies were too low (threshold 2,000 pregnancies per MCN) (n=4,222).

Table 2 shows unadjusted descriptive statistics (mean/%, minimum and maximum value) at the MCN-level for outputs, inputs and case-mix variables. From 2011 through 2016, 91 distinct MCNs served a total of 297,960 low-risk pregnancies. That comes down to an average of 4,054 pregnancies per MCN during the whole study period, where the smallest MCN had 764 pregnancies and the largest MCN had 7,861 pregnancies. Unadjusted rates of low-value services varied widely, with c-sections ranging from 6% through 21%, epidural ranging from 5% through 58% and induction ranging from 20% through 69%, depending on the MCN. Rates for neonatal and maternal health outcomes also varied across MCNs. For example, rates of low Apgar score varied between 0.2% and 2.3%. Subsequent to the utilization of low-value services, the rates of low Apgar score ranged from 0.0% through 6.8% (after c-sections), from 0.4% through 3.6% (after epidural) and from 0.3% through 3.3% (after induction). Case-mix characteristics of the mother, father, pregnancy, and child also varied across MCNs.

Variation of value-indicators across MCNs

After we adjusted for case-mix, variation across MCNs persisted for all inputs and outputs (Table 3). Most variation was observed for the inputs indicators. For example, labor induction showed rates of 21.0% through 67.7% depending on the MCN, and 66 out of 91 MCNs showed statistically significant differences from the average. Variation also persisted for the outcomes. For example, 'low Apgar score after c-section' ranged from 1.6% through 5.2%, and 5 out of the 91 MCNs differed statistically significant from the average, for both the input (c-section) and the output (low Apgar score). The output with the largest number of MCNs (20 out of 91) that differed statistically significant from the average is 'hemorrhage after labor induction' and varied between 4.9% and 11.7%, depending on the MCN.

Table 2: Descriptive statistics at the Maternity Care Network Level for low-risk pregnancies

	Mean	Minimum	Maximum
Number of MCNs: 91			
Number of pregnancies: 297,960	4054	764	7861
Inputs – Low-value services			
C-section, %	13.53	5.97	20.69
Primary	1.23	0.26	2.99
Elective	1.62	0.00	4.11
Secondary	12.30	5.44	18.81
Epidural, %	31.44	5.28	57.98
Induction, %	54.32	20.31	68.77
Outputs: Neonatal and maternal health outcomes			
Maternal mortality, %	0.00	0.00	0.09
Perinatal mortality, %	0.02	0.00	0.26
AOI-5, %	11.64	6.51	16.64
Low Apgar score, %	1.16	0.20	2.26
After c-section*	2.57	0.00	6.79
After epidural*	1.74	0.35	3.64
After induction*	1.39	0.29	3.26
NICU-admission, %	0.00	0.00	0.00
Hemorrhage, %	7.35	3.26	10.99
After c-section*	6.89	1.13	21.02
After induction*	8.17	3.19	12.63

Table 2: Descriptive statistics at the Maternity Care Network Level for low-risk pregnancies

	Mean	Minimum	Maximum
Perineal Damage, %	3.67	0.98	6.27
After induction*	3.19	0.57	5.42
Case-mix (mother) characteristics			
Age mother, years	29.13	27.36	32.07
Educational level (SOI 2006 levels), %			
1 - nursery	1.34	0.43	5.37
2 – primary	1.99	0.67	4.62
3 – lower secondary	9.37	4.18	14.58
4 – upper secondary	43.41	18.69	61.18
5 – vocational or bachelors' degree	28.64	19.81	34.46
6 – masters' degree	13.35	2.82	39.61
7 – phd / postdoctoral	1.90	0.25	7.50
Household income (deciles), %			
1	5.50	2.62	15.03
2	7.76	3.88	15.59
3	6.54	3.34	10.32
4	9.38	4.70	13.62
5	12.19	6.59	19.10
6	14.24	6.45	19.07
7	13.97	8.40	17.53

Table 2: Descriptive statistics at the Maternity Care Network Level for low-risk pregnancies

	Mean	Minimum	Maximum
8	12.26	7.39	16.48
9	10.08	4.17	19.60
10	8.07	1.56	31.33
Etnicity, %			
Western	85.08	46.07	97.25
Urbanity (urban to rural), %			
1	22.37	0.00	92.99
2	24.96	0.13	64.04
3	18.69	0.40	39.93
4	18.50	0.15	62.88
5	15.48	0.04	52.59
Previous miscarriage, %	20.11	13.92	35.84
Gravidity >1, %	43.77	32.35	70.68
Previous (-2 years) Maternal Health Spending, €	1340.31	1040.27	2048.63
Previous (-2 years) Maternal Health Spending, %			
P0 – p33	39.91	29.85	47.45
P33-p66	37.14	33.62	41.62
P6-p100	22.95	18.26	34.13
Case-mix (father) characteristics			
Paternal age >45 years, %	16.08	11.96	34.91
Case-mix (pregnancy) characteristics			

Table 2: Descriptive statistics at the Maternity Care Network Level for low-risk pregnancies

	Mean	Minimum	Maximum
Gestational age, days	278.82	276.40	280.57
Gestational age in weeks, %			
37	8.07	5.29	14.63
38	15.37	11.03	20.07
39	25.00	22.27	29.01
40	30.68	25.84	34.08
41	20.89	15.50	26.64
Birth weight, grams	3420.90	3313.93	3486.84
Birth weight in grams, %			
P0-P33	32.89	27.31	42.84
P33-P66	36.78	33.54	38.93
P66-P100	30.32	23.07	35,58
Gestational hypertension, %	11.89	6.83	21.13
Gestational diabetes, %	2.85	0.89	8.38
Case-mix (child) characteristics			
Male gender, %	51.27	49.02	53.88

MCNs: Maternity Care Networks; C-section: Caesarean section; Epidural: Epidural Analgesia; Induction: Labor Induction; AOI-5: Averse Outcome Index 5 (incidence of (at least) one out of five adverse outcomes)); Low Apgar Score: Apgar Score <7 after 5 min; NICU: Neonatal Intensive Care Unit; Hemorrhage: Hemorrhage postpartum; Perineal Damage: Serious Perineal Damage (stage 3 or 4); *number of pregnancies in the C-section subsample (mean: 543; min: 100; max: 1008), number of pregnancies in the Epidural subsample (mean: 1450; min: 126; max: 3420), number of pregnancies in the Induction subsample (mean: 2262; min: 340; max: 4499).

Table 3: Variation in Value Across Maternity Care Networks

Value-indicators		**Adjusted# rate per 100 pregnancies**			**No. of MCNs that differ from the average**		
Input	**Output**	**Mean**	**Min**	**Max**	**Input**	**Output**	**Input and output**
C-Section		13.10	6.24	19.80	61		
	Low Apgar Score	2.60	1.61	5.22		5	5
	Hemorrhage	6.71	2.23	17.62		22	14
Epidural		31.06	4.93	58.25	75		
	Low Apgar Score	1.74	1.14	3.04		7	5
Labor Induction		54.20	21.00	67.68	66		
	Low Apgar Score	1.37	0.70	2.64		13	9
	Hemorrhage	8.15	4.93	11.65		26	20
	Perineal Damage	3.17	1.44	4.94		24	16

*Number of Maternity Care Networks that statistically differ from the mean (i.e. random effect of zero) for alpha < 0.05. #adjusted for case-mix mother, father, child and pregnancy. C-Section: Caesarean Section; Epidural: Epidural Analgesia

Figure 1 shows the random effects at the log scale for the value-indicator 'hemorrhage after labor induction' at the MCN level. Each point in the quadrants represents one MCN and confidence intervals are shown around the points. MCNs at the bottom left show higher value as they perform relatively less low-value services and have less adverse outcomes (relative to other MCNs). MCNs at the top right show lower value as they perform relatively more low-value services and have more adverse outcomes (relative to other MCNs). Figure 1a shows the MCNs (n=66) that differed from the average regarding labor inductions (input) (i.e. no overlapping confidence interval with Y-axis), figure 1b shows the MCNs (n=26) that differed with respect to hemorrhages (output) (i.e. no overlapping confidence interval with the X-axis), and figure 1c depicts the MCNs (n=20) that differed on both labor inductions (input) and hemorrhages (output), thus value, for alpha <0.05 (no overlapping of both Y- and X-axis). Eight MCNs were found to be of higher value, depicted in the lower left part of the quadrant, meaning that they performed relatively less inductions and had relatively less hemorrhages, corrected for case-mix. Five MCNs were found of lower value, meaning that they performed relatively more inductions and had relatively more hemorrhages, depicted at the higher right part of the quadrant. The Appendix contains the figures for all value-indicators. The Appendix also contains the regression results that yielded the MCN effects.

Additional analyses

Correlations between inputs and outputs: relation between input and output?

In order to gain insight into the relation between the inputs and outputs, we assessed Spearmans' correlation coefficient between each input and output. Five value-indicators did not show statistically significant correlations between input and output performances (see Appendix for the full table). This means that MCNs performing relatively more low-value services did not show relatively more adverse outcomes and vice versa. Instead, most MCNs performing relatively more low-value services had showed no differences in outcomes. The exceptions were correlations between the relative performance of MCNs regarding 'c-sections' and 'low Apgar score, and 'epidural' and 'low Apgar score', which were weakly negatively correlated (rho, -0.37; 95 percent CI -0.53; -0,20 and rho, -0.20*; 95 percent CI -0.39; -0.01 respectively). This indicates that MCNs performing relatively more c-sections or epidurals had relatively less low Apgar scores. Yet, visual inspection of the graph (see Appendix) shows that there are no MCNs that had statistically significant less than average low Apgar scores (MCNs had either an average score or higher than average score).

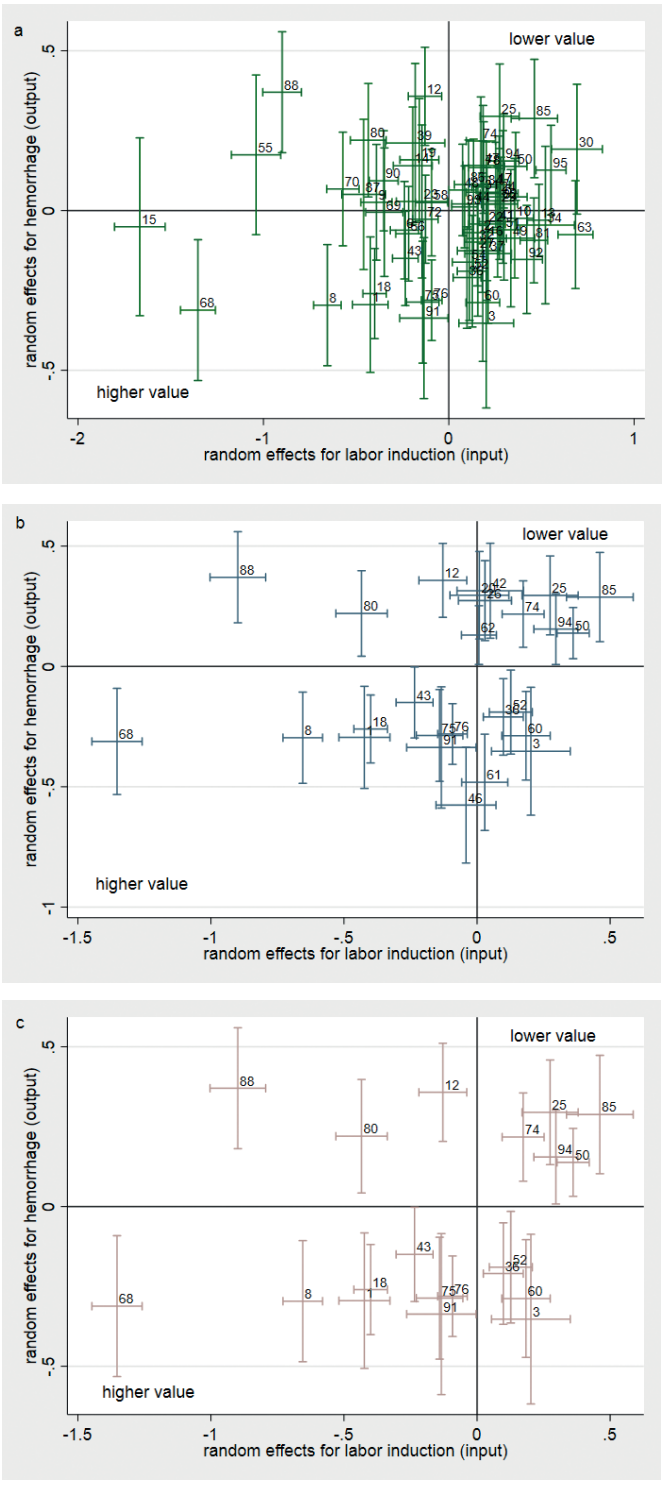


Figure 1: Maternity Care Network (MCN) performances on the value-indicator 'hemorrhage after labor induction': a) MCNs that differed from the average on labor inductions (input), b) MCNs that differed from the average on hemorrhages (output), c) MCNs that differed on both labor induction (input) and hemorrhages (output)

Consistency across inputs and outputs: do they measure the same concept?

In order to gain insight into whether the value-indicators might grasp the same concept, we assessed the consistency across inputs and outputs. The consistency of the relative performance of the MCNs across adjusted inputs and adjusted outputs was positively correlated for 6 out of 18 pairs (Table 4). Correlations were positive and statistically significant for 2 out of 3 input pairs, and for 4 out of 15 pairs of output/input. Statistically significant correlation coefficients ranged from weak (in red) to strong (in green) (rho: 0.24 with 95 percent CI, 0.05-0.42 to rho: 0.77 with 95 percent CI, 0.68-0.87). Consistency was found across all 'low Apgar score' outputs, irrespective of the inputs (labor induction, caesarean section, epidural), which indicates that MCNs showing high (or low) scores on low Apgar score after caesarean section, also showed low Apgar scores after labor induction and after epidural analgesia. This mostly indicated that the same persons were measured (low Apgar score after induction ánd a c-section, for instance). Across different outcomes, for example, across low Apgar score and hemorrhage we found low correlation coefficients.

Persistence over time: do they measure possibly the concept of value?

The persistence over two consecutive years indicates whether the inputs and outputs could possibly reflect the concept of value as we expected true value not to change much over one year. For inputs, the correlations were positive and statistically significant for all subsequent years (except for 'labor induction' between 2015 and 2016). The strongest correlation coefficients for the subsequent years were found for 'epidural' (ranging from rho: 0.84 with 95 percent CI, 0.71-0.97 to rho: 0.93 with 95 percent CI, 0.89-0.98). All inputs were considered to persist over time as they had four or more (out of five) statistically significant correlations of rho >0.40 (orange or green) in subsequent years. For outputs, 18 out of 30 subsequent years showed positive and statistically significant correlations. Statistically significant correlation coefficients ranged from weak to moderate (rho: 0,23 with 95 percent CI, [0.01,0.44] ('low Apgar score after epidural' between 2011 and 2012) to rho: 0,52 with 95 percent CI, [0.35,0.69] ('hemorrhage after c-section' between 2011 and 2012)). The output that showed the most persistence over time was 'hemorrhage after labor induction', which had two statistically significant correlations of >0.40 in subsequent years.

Table 4: Consistency between Inputs and Outputs of Value-Indicators at the MCN level (N=91)

Inputs	**Caesarean section**		**Epidural analgesia**		**Labor induction**	
Outputs		**Low Apgar score**	**Hemorrhage**	**Low Apgar score**	**Low Apgar score**	**Hemorrhage**
Caesarean section	-					
Low Apgar score		-				
Hemorrhage		0.03 [-0.18, 0.24]	-			
Epidural analgesia	0.11 [-0.09, 0.32]			-		
Low Apgar score		0.58*** [0.43, 0.73]	0.07 [-0.13, 0.28]		-	
Labor induction	0.24* [0.05, 0.42]			0.26* [0.03, 0,48]		-
Low Apgar score		0.66*** [0.53, 0.79]	0.19 [-0.01, 0.39]		0.77*** [0.68, 0.87]	-
Hemorrhage		-0.04 [-0.24, 0.17]	0.47*** [0.29, 0.65]		-0.04 [-0.25, 0.17]	-0.03 [-0.24, 0.18]	-
Perineal Damage		0.17 [-0.02 – 0.35]	0.02 [-0.19, 0.23]		0.03 [-0.16, 0.23]	0.00 [-0.21, 0.21]	0.12 [-0.08, 0.32]

Estimates are Spearman correlation coefficients (rho) and bootstrapped 95% confidence intervals for 1000 replications in brackets. N = 91. 95% confidence intervals in brackets. * p < 0.05, ** p < 0.01, *** p <0.001; Coefficients were interpreted as negligible (0.00-0.10) in red, weak (0.10-0.39) in red, moderate (0.40-0.69) in orange, strong (0.70-0.89) in green or very strong (0.90-1.00) in green.

Table 5: Persistence of Inputs and Outputs of Value-Indicators Over Time at the MCN level (N=91)

Inputs	Outputs	2011-2012	2012-2013	2013-2014	2014-2015	2015-2016	# sign	# sign and rho > 0.40
Caesarean section		0.71***	0.69***	0.64***	0.66***	0.73***	5	5
		[0.58,0.85]	[0.55,0.82]	[0.48,0.80]	[0.52,0.80]	[0.62,0.85]		
	Low Apgar score	0.09	0.13	0.23	0.29**	0.18	1	0
		[-0.12,0.31]	[-0.10,0.35]	[-0.01,0.46]	[0.07,0.51]	[-0.07,0.43]		
	Hemorrhage	0.52***	0.37***	0.12	0.19	0.34**	3	1
		[0.35,0.69]	[0.16,0.59]	[-0.13,0.38]	[-0.05,0.43]	[0.12,0.55]		
Epidural Analgesia		0.84***	0.93***	0.90***	0.93***	0.88***	5	5
		[0.71,0.97]	[0.89,0.98]	[0.84,0.96]	[0.89,0.98]	[0.80,0.95]		
	Low Apgar score	0.23*	0.09	0.31**	0.17	0.24*	3	0
		[0.01,0.44]	[-0.15,0.33]	[0.10,0.52]	[-0.06,0.40]	[0.03,0.46]		
Labor Induction		0.74***	0.57***	0.56***	0.76***	0.13	4	4
		[0.63,0.86]	[0.39,0.76]	[0.38,0.74]	[0.63,0.88]	[-0.11,0.37]		
	Low Apgar score	0.13	0.18	0.39***	0.27*	0.27*	3	0
		[-0.08,0.34]	[-0.07,0.43]	[0.19,0.59]	[0.06,0.48]	[0.05,0.49]		
	Hemorrhage	0.49***	0.44***	0.34**	0.33**	0.09	4	2
		[0.30,0.68]	[0.25,0.64]	[0.12,0.56]	[0.11,0.55]	[-0.14,0.32]		
	Perineal Damage	0.36**	0.34**	0.28*	0.48***	0.08	4	1
		[0.14,0.59]	[0.13,0.56]	[0.07,0.49]	[0.31,0.66]	[-0.15,0.31]		

Estimates are Spearman correlation coefficients (rho) and bootstrapped 95% confidence intervals for 1000 replications in brackets. N = 77. 95% confidence intervals in brackets. * p < 0.05, ** p < 0.01, *** p <0.001; Coefficients were interpreted as negligible (0.00-0.10) in red, weak (0.10-0.39) in red, moderate (0.40-0.69) in

DISCUSSION

This study aimed to explore whether it is possible to demonstrate difference in value across Dutch MCNs by studying six proposed value-indicators that could be operationalized in Dutch nationwide linked observational data from 2011 through 2016. We observed substantial variation in all six value-indicators across MCNs after adjusting for characteristics of mothers, fathers, pregnancies and children in low-risk pregnancies. Additional analyses showed no correlations between inputs and outputs, except between 'low Apgar score' and 'caesarean section' and between 'low Apgar score' and 'epidural analgesia', which were weakly negatively correlated. Additionally, we found that MCNs showing relatively high (or low) 'low Apgar score after caesarean section', also showed relatively high (or low) 'low Apgar scores after labor induction' and after epidural analgesia (consistency). We also found consistency between 'hemorrhage after induction' and 'hemorrhage after c-section'. Persistence over time was found for all inputs, i.e. the low-value care indicators (labor indication, caesarean section, epidural), for at least four out of five sequential years.

Our findings suggest that the inputs (which are known as low-value care indicators) may, in part, reflect the concept of value, because they showed variation across MCNs and persistence over time. The performance of MCNs regarding 'low Apgar score' and 'hemorrhage post-partum' seems invariant to the rate of low-value services in low-risk pregnancies (i.e. whatever treatment rate, MCNs seem to have fixed health outcome rates). This points at overuse of those services, because it seems that a reduction of the utilization of caesarean sections, epidural analgesia and labor induction is not associated with a deterioration in maternal and neonatal health outcomes at the MCN level. Therefore, policies aiming at reducing these services would improve value.

Although the low-value care indicators that we have used in this study may have captured a part of value, we considered the low-value care indicators not to be consistent. Our results showed that the use of labor inductions was weakly positively correlated with the use of caesarean sections and the use of epidural analgesia. That means that if an MCN has relatively high rates for one low-value care indicator, it may have relatively low rates for another. This indicates that value likely is multi-faceted, and not easily measured by just one indicator. Our results somewhat contrast previous research of Schwartz and colleagues, who found (moderate) consistency between different clinical disciplines within organizations [13, 24]. Possibly, some aspects of obstetrics are less influenced by practice patterns than others [13], or areas of expertise vary across MCNs.

Despite our efforts to bypass endogeneity issues by using low-value utilization indicators instead of health spending and using many case-mix variables, we still detected population heterogeneity that may have clouded our results. We found that high variability of value-indicators, both inputs and outputs, in addition to wide confidence intervals, which suggests we were not able to control sufficiently for all factors that play a role in predicting the inputs and outputs. Variances between MCNs may be driven by unobserved or unavailable characteristics of the population or care professionals that we were not able to capture, such as preferences of pregnant women, training of midwifes and obstetricians and level and history of collaboration within an MCN [13]. Nevertheless, whatever the cause of the variation may be, we did observe persistent and wide variations for the low-value care indicators across MCNs, which is consistent with previous literature [28, 34] and suggests that there is room for improvement at the MCN level [13].

The most important limitation to our study was that we had no access to patient reported outcome measures (PROMs) or patient reported experience measures (PREMs). These indicators are important for capturing the concept of value as they potentially reflect *what matters to women,* which is essential as it is part of the definition of value [3]. By lack of patient reported indicators, we used observational outcomes, which did not show persistence over time. In addition, we excluded pregnancies with a NICU-admission, because all these pregnancies were attributed to the nine MCNs that had such a unit, and would have resulted unfair high rates of negative outcomes for those MCNs. The drawback is that we excluded the pregnancies with outcomes of interest, which may have had an important impact on the variation of the outcomes across MCNs.

Future research should investigate possibilities to include PROMs and PREMs on a large scale in observational data, without putting an administrative burden on (pregnant) women and providers, for example by using mobile phone applications. Subsequently, the marginal effects of low-value services on PROMs and PREMs should be investigated, in addition to exploring other (types of) outcomes that directly reflect the value of the services delivered in terms of experiences of women. Furthermore, future research should focus on dealing with limitations of observational data, for example that it still is difficult to deal with endogeneity problems and population heterogeneity, and be transparent about it.

Conclusions

Our findings suggested differences between MCNs with respect to the selected value-indicators. In addition, our results indicated that some outputs may have measured a

similar concept (consistency), and that the inputs (i.e. low-value care indicators) may have captured a part of value (persistence). Yet, despite access to many case-mix variables from multiple datasets, we detected population heterogeneity which might have clouded our results. Future research should further deal with the limitations of observational data and refine outcome measures that matter to the population of pregnant women in order to truly capture value at the MCN level and contribute to value-based maternity care.

ACKNOWLEDGEMENTS

We thank Statistics Netherlands, Vektis and Perined for kindly providing access to their data.

REFERENCES

1. Alley, D.E., et al., *Accountable Health Communities--Addressing Social Needs through Medicare and Medicaid.* N Engl J Med, 2016. 374(1): p. 8-11.
2. Drewes, H.W., Struijs, J.N., Baan, C.A., *How the Netherlands Is Integrating Health and Community Services.* NEJM Catalyst, 2016.
3. Porter, M.E., *A strategy for health care reform--toward a value-based system.* N Engl J Med, 2009. 361(2): p. 109-12.
4. Struijs, J.N., De Vries, E.F., Van Dorst, H.D.C.A., Over, E.A.B., Baan, C.A., *Insight in outcomes, utilization and medical spending of maternity care and first experiences with bundled payments* (in Dutch). 2018, Bilthoven: The Dutch National Institute for Public Health and the Environment (RIVM).
5. de Vries, E.F., et al., *Unraveling the drivers of regional variation in healthcare spending by analyzing prevalent chronic diseases.* BMC HSR, 2018. 18(1): p. 323.
6. Newhouse, J.P., et al., *Variation in Health Care Spending: Target Decision Making, Not Geography*. 2013, National Academies Press (US): Washington (DC).
7. Fisher, E.S., Bynum, J.P., and Skinner, J.S., *Slowing the growth of health care costs--lessons from regional variation.* N Engl J Med, 2009. 360(9): p. 849-52.
8. Skinner, J., *Chapter Two - Causes and Consequences of Regional Variations in Health Care1*, in *Handbook of Health Economics*, T.G.M. Mark V. Pauly and P.B. Pedro, Editors. 2011, Elsevier. p. 45-93.
9. Manning WG, N.E., Wilk AS, *Explaining geographic variation in health care spending, use and quality, and associated methodological challenges*, 2012, IOM: Washington, DC.
10. Zuckerman, S., et al., *Clarifying sources of geographic differences in Medicare spending.* N Engl J Med, 2010. 363(1): p. 54-62.
11. Wennberg, D.E., et al., *A population health approach to reducing observational intensity bias in health risk adjustment: cross sectional analysis of insurance claims.* BMJ, 2014. 348: p. g2392.
12. Schwartz, A.L., et al., *Analysis of physician variation in provision of low-value services.* JAMA Intern Med, 2018.
13. Schwartz, A.L., et al., *Low-Value Service Use in Provider Organizations.* Health Serv Res, 2018. 53(1): p. 87-119.
14. Rosenberg, B.L., et al., *Quantifying Geographic Variation in Health Care Outcomes in the United States before and after Risk-Adjustment.* PLoS One, 2016. 11(12): p. e0166762.
15. Skinner, J. and Chandra, A., *The Past and Future of the Affordable Care Act.* JAMA, 2016. 316(5): p. 497-9.
16. Romley, J., et al., *Geographic variation in the delivery of high-value inpatient care.* PLoS One, 2019. 14(3): p. e0213647.
17. Moran, V., Jacobs, R., and Mason, A., *Variations in Performance of Mental Health Providers in the English NHS: An Analysis of the Relationship Between Readmission Rates and Length-of-Stay.* Adm Policy Ment Health, 2017. 44(2): p. 188-200.
18. Portrait, F.R., O. van der Galien, and B. Van den Berg, *Measuring Healthcare Providers' Performances Within Managed Competition using Multidimensional Quality and Cost Indicators.* Health Econ, 2015.
19. Steenhuis, S., et al., *Good, better, best? A comprehensive comparison of healthcare providers' performance: An application to physiotherapy practices in primary care.* Health Policy, 2017.
20. Hakkinen, U., et al., *Quality, cost, and their trade-off in treating AMI and stroke patients in European hospitals.* Health Policy, 2014. 117(1): p. 15-27.
21. Muratov, S., et al., *Regional variation in healthcare spending and mortality among senior high-cost healthcare users in Ontario, Canada: a retrospective matched cohort study.* BMC Geriatr, 2018. 18(1): p. 262.
22. Gutacker, N., et al., *Truly inefficient or providing better quality of care? Analysing the relationship between risk-adjusted hospital costs and patients' health outcomes.* Health Econ, 2013. 22(8): p. 931-47.

23. Colla, C.H., et al., *Choosing wisely: prevalence and correlates of low-value health care services in the United States.* J Gen Intern Med, 2015. 30(2): p. 221-8.

24. Schwartz, A.L., et al., *Measuring low-value care in Medicare.* JAMA Intern Med, 2014. 174(7): p. 1067-76.

25. PeriNed, *Perinatale Zorg in Nederland 2015*, PeriNed, Editor. 2016: Utrecht.

26. De Boo, A., *Vektis 'Center of information for health services' (in Dutch).* Tijdschrift voor gezondheidswetenschappen, 2011. **89**(7): p. 358-359.

27. General Data Protection Regulation (in Dutch). 2019.

28. Seijmonsbergen-Schermers, A.E., et al., *Regional variations in childbirth interventions in the Netherlands: a nationwide explorative study.* BMC Pregnancy Childbirth, 2018. 18(1): p. 192.

29. Euro-Peristat Project, *European Perinatal Health Report. Core indicators of the health and care of pregnant women and babies in Europe in 2015.* 2018.

30. Molina, G., et al., *Relationship Between Cesarean Delivery Rate and Maternal and Neonatal Mortality.* Jama, 2015. 314(21): p. 2263-70.

31. Ye, J., et al., *Association between rates of caesarean section and maternal and neonatal mortality in the 21st century: a worldwide population-based ecological study with longitudinal data.* BJOG, 2016. 123(5): p. 745-53.

32. Patel, N. and M. Radeos, *Severe Delayed Postpartum Hemorrhage after Cesarean Section.* J Emerg Med, 2018. 55(3): p. 408-410.

33. Magann, E.F., et al., *Postpartum hemorrhage after cesarean delivery: an analysis of risk factors.* South Med J, 2005. 98(7): p. 681-5.

34. Glantz, J.C., *Obstetric Variation, Intervention, and Outcomes: Doing More but Accomplishing Less.* 2012. 39(4): p. 286-290.

35. Srebnik, N., et al., *The impact of epidural analgesia on the mode of delivery in nulliparous women that attain the second stage of labor.* J Matern Fetal Neonatal Med, 2019: p. 1-8.

36. Anim-Somuah, M., R.M. Smyth, and L. Jones, *Epidural versus non-epidural or no analgesia in labour.* Cochrane Database Syst Rev, 2011(12): p. Cd000331.

37. Yin, H. and R. Hu, *A cohort study of the impact of epidural analgesia on maternal and neonatal outcomes.* J Obstet Gynaecol Res, 2019. 45(8): p. 1435-1441.

38. Loewenberg-Weisband, Y., et al., *Epidural analgesia and severe perineal tears: a literature review and large cohort study.* J Matern Fetal Neonatal Med, 2014. 27(18): p. 1864-9.

39. Glantz, J.C., *Elective induction vs. spontaneous labor associations and outcomes.* J Reprod Med, 2005. 50(4): p. 235-40.

40. Vardo, J.H., L.L. Thornburg, and J.C. Glantz, *Maternal and neonatal morbidity among nulliparous women undergoing elective induction of labor.* J Reprod Med, 2011. 56(1-2): p. 25-30.

41. Pergialiotis, V., et al., *Risk factors for severe perineal lacerations during childbirth.* Int J Gynaecol Obstet, 2014. 125(1): p. 6-14.

42. Brownlee, S., et al., *Evidence for overuse of medical services around the world.* Lancet, 2017. 390(10090): p. 156-168.

43. Little, S.E., et al., *The relationship between variations in cesarean delivery and regional health care use in the United States.* Am J Obstet Gynecol, 2016. 214(6): p. 735.e1-8.

44. Elferink-Stinkens, P.M., et al., *Large differences in obstetrical intervention rates among Dutch hospitals, even after adjustment for population differences.* Eur J Obstet Gynecol Reprod Biol, 1996. 68(1-2): p. 97-103.

45. Chan, K.S., et al., *The state of overuse measurement: a critical review.* Med Care Res Rev, 2013. 70(5): p. 473-96.

46. Zhang, L., et al., *Preventive induction of labor for non-urgent indications at term and maternal and neonatal outcomes.* Reprod Health, 2016. 13: p. 46.

47. De Vries, E.F., et al., *Are low-value care measures up to the task? A systematic review of the literature.* BMC HSR, 2016. 16(1): p. 1-12.

48. Werkgroep Indicatoren Integrale Geboortezorg, *Indicatorenset Integrale Geboortezorg.* 2017, Zorginstituut Nederland: Diemen.

49. Project team MILLQ (Mother and Infant Labour Quality), *Adverse Outcome Index – Joint responsibility [De adverse Outcome Index: Samen verantwoordelijk, samen transparent – verantwoordingsdocument].* 2015: The Netherlands.

50. Mann, S., et al., *Assessing quality obstetrical care: development of standardized measures.* Jt Comm J Qual Patient Saf, 2006. 32(9): p. 497-505.

51. ICHOM, *Pregnancy and Childbirth Data Collection Reference Guide*. 2016, ICHOM.

52. Khandwala, Y.S., et al., *Association of paternal age with perinatal outcomes between 2007 and 2016 in the United States: population based cohort study.* 2018. 363: p. k4372.

53. Jacobs R, S.P., Street A., *Some alternative approaches to measuring performance*, in *Measuring Efficiency in Health Care - Analytic Techniques and Health Policy*. 2009, Cambridge University Press

54. Schober, P., Boer, C., and Schwarte, L.A., *Correlation Coefficients: Appropriate Use and Interpretation*. Anesth Analg, 2018. 126(5): p. 1763-1768.

SUPPLEMENTARY FILES

S1. Data linkage process

The linkage of the datasets and the analyses of this study were performed in the highly secured remote access platform from Statistics Netherlands (SN). Vektis and Perined uploaded their data to the secured platform from SN. To adhere to the privacy legislation, SN pseudonomized the social security numbers in the Vektis data using pseudonomized identification numbers (PINs). The linkage percentage for the Vektis-data was 99.9% for each year. Subsequently, SN linked the Perined data, which do not include social security numbers, to the municipality registration data in order to obtain PINs for mother and child. The linkage key comprised out of: the date of birth of the mother, date of birth of the child, gender of the child, single births or twins, live- or stillbirth and 4-digit zip code. The linkage percentage for the Perined-data varied between 92 and 98% depending on the year. After the linkage and pseudonomization by SN, the data was accessible for the authors.

The datasets were linked in three subsequent steps. First, we combined nationwide individual level all-payer claims data (containing total spending per year) for 2009 through 2016. Then, we linked datasets containing municipality registration data (i.e. year of birth, month of birth, gender, ethnicity), highest attained educational level and standardized household income using the unique PINs. Finally, we used Perined data, which showed one pregnancy per record, to link the father, mother and child data separately to.

In order to include the detailed claims data from Vektis on maternity spending, we additionally performed the following steps and applied the following assumptions. First, we created pregnancy-id's that could end in both a birth or a miscarriage. Therefore, we first created a list of all unique mothers (in the Vektis-data) and used the SN data to link parents to their children. We identified multiple gestations based on mothers' PIN and month of birth using the community registration set from SN. We assumed that mothers for whom no child was found in the community registration set, have had a miscarriage. A number was assigned to each pregnancy based on chronological order. In this manner, we made a distinction between births and miscarriages. Second, we linked claims data to a singular pregnancy based on the date the claim was made and the date of birth of the child based on the community registration set. A claim was linked to a pregnancy when the date fell between 9 months prior to the date of birth and two months after the date of birth. The remaining claims were labeled as miscarriage claims. We assumed

that claims in concurrent months belonged to the same miscarriage. Third, we linked the mother and child combinations to the Perined data using the pregnancy-id.

The created linked dataset could now be augmented with several SN datasets, such as datasets containing municipality registrations (age, ethnicity, gender), highest attained educational level and level of urbanization.

S2. Endogeneity of health spending

As discussed in the introduction section, healthcare spending may not be ideal for analyses of the variation of value as endogeneity problems may occur. We wanted to confirm that this endogeneity issue also applies to our dataset by inspecting the relationship between maternity spending and maternal/infant health outcomes at the individual level. In order to do so, we created a subsample (n=268,893) of our data for which detailed claims data on maternity care were available (2015-2016). To define the spending on maternity care, we categorized a single claim into a prenatal, natal and postnatal phase and summed those categories. In order to inspect the relationship between maternal and infant health outcomes, we modelled each of the health outcomes as a function of maternity spending (in 10 bins), case-mix and MCNs, where case-mix and MCNs were kept constant. We fitted logistic models using the covariates from the main analysis (age mother, educational level, household income, ethnicity, urbanity, gestational age, birth weight, gestational hypertension, gestational diabetes, healthcare spending in the previous year, gender of the child and age of the father). The marginal contributions of maternity spending on health outcomes with their confidence intervals (95%) were plotted (see Figure A1). There seems to be a positive association (i.e. the higher the adverse health outcome, the higher the maternity spending) between the adverse outcomes components separately and maternity care spending. This suggests that the spending level seems to be associated with the type of treatment (which is necessary in order to deal with the health status the mother and child are presented with), rather than quality of care, even when adjusted for case-mix.

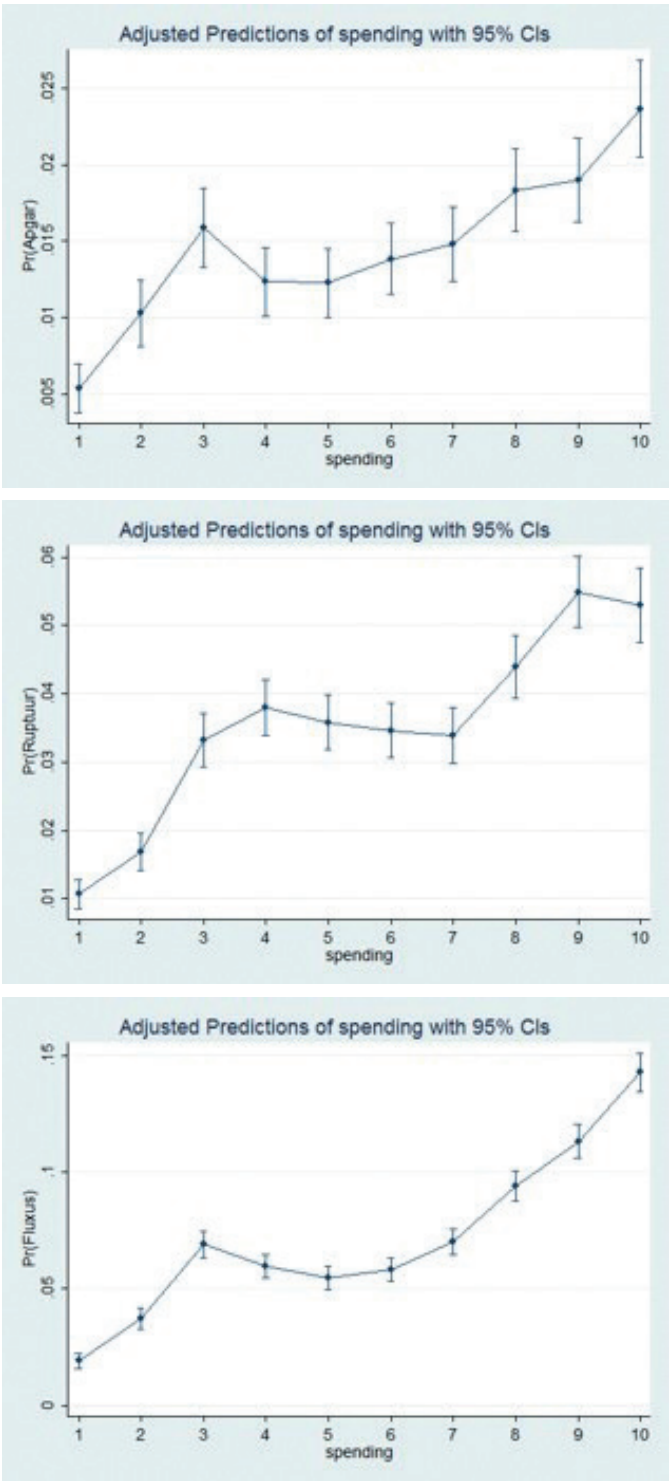


Figure A1: adjusted predictions of maternity spending on adverse neonatal and maternal outcomes (a: low Apgar score; b) hemorrhage; c) perineal damage)

S3. Graphics of the six value-indicators

The figures show the random effects at the log scale for the five additional value-indicators at the MCN level. In each plot, each point represents one MCN and confidence intervals are shown around the points. MCNs at the bottom left part of the quadrant show higher value as they perform relatively less low-value services and have less adverse outcomes (relative to other MCNs). MCNs depicted at the top right part of the quadrant show lower value as they perform relatively more low-value services and have more adverse outcomes (relative to other MCNs). For each value-indicator, the figures *a* show the MCNs that differed from the average regarding inputs. The figures *b* show the MCNs with respect to the outputs. And the figures *c* depict the MCNs that differed from the average on both inputs and outputs (i.e. value), for alpha < 0.05.

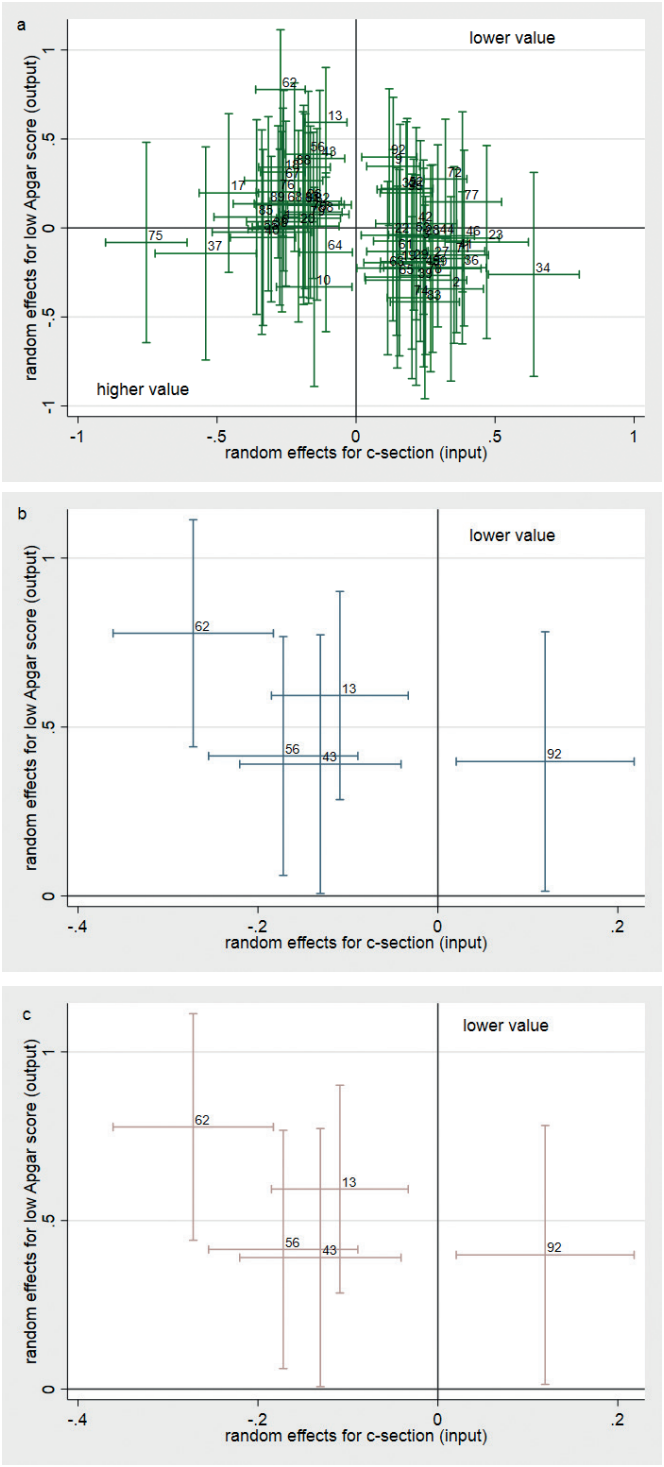


Figure A2.1. Maternity Care Network performances for 'Low Apgar score after Caesarean section'

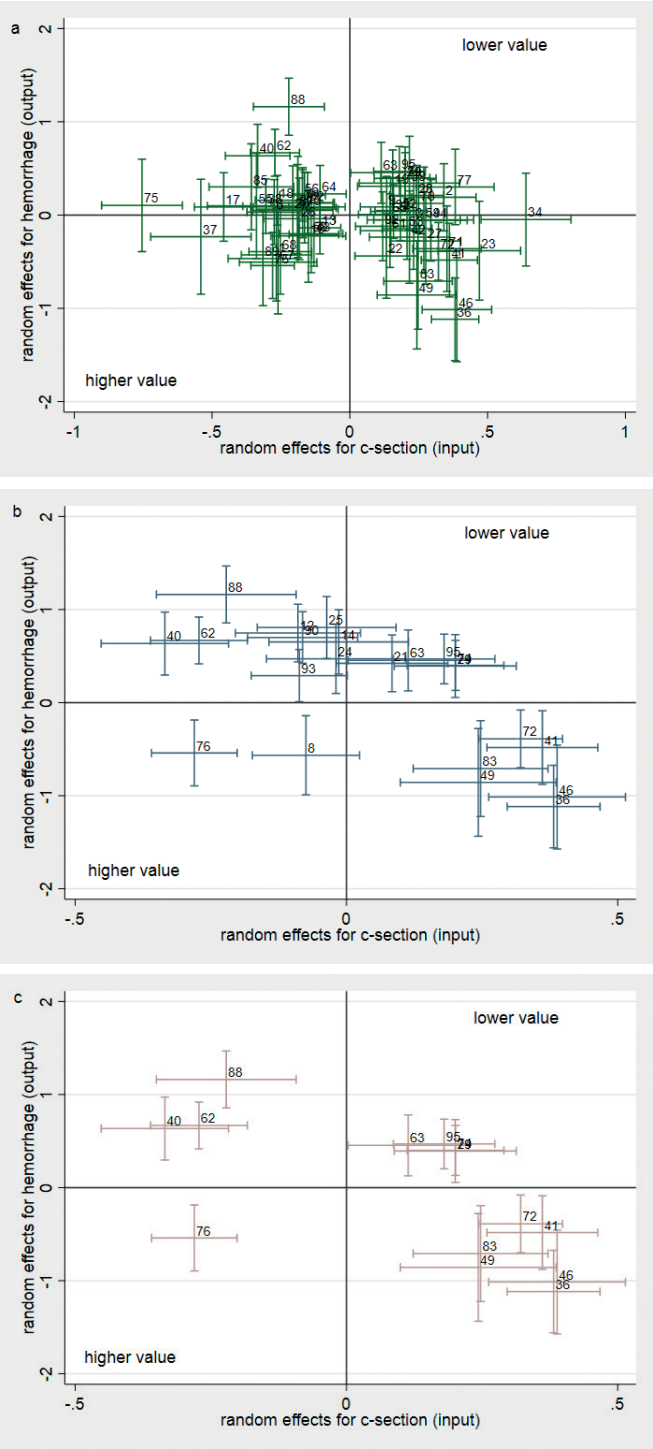


Figure A2.2. Maternity Care Network performances for 'Hemorrhage after Caesarean section'

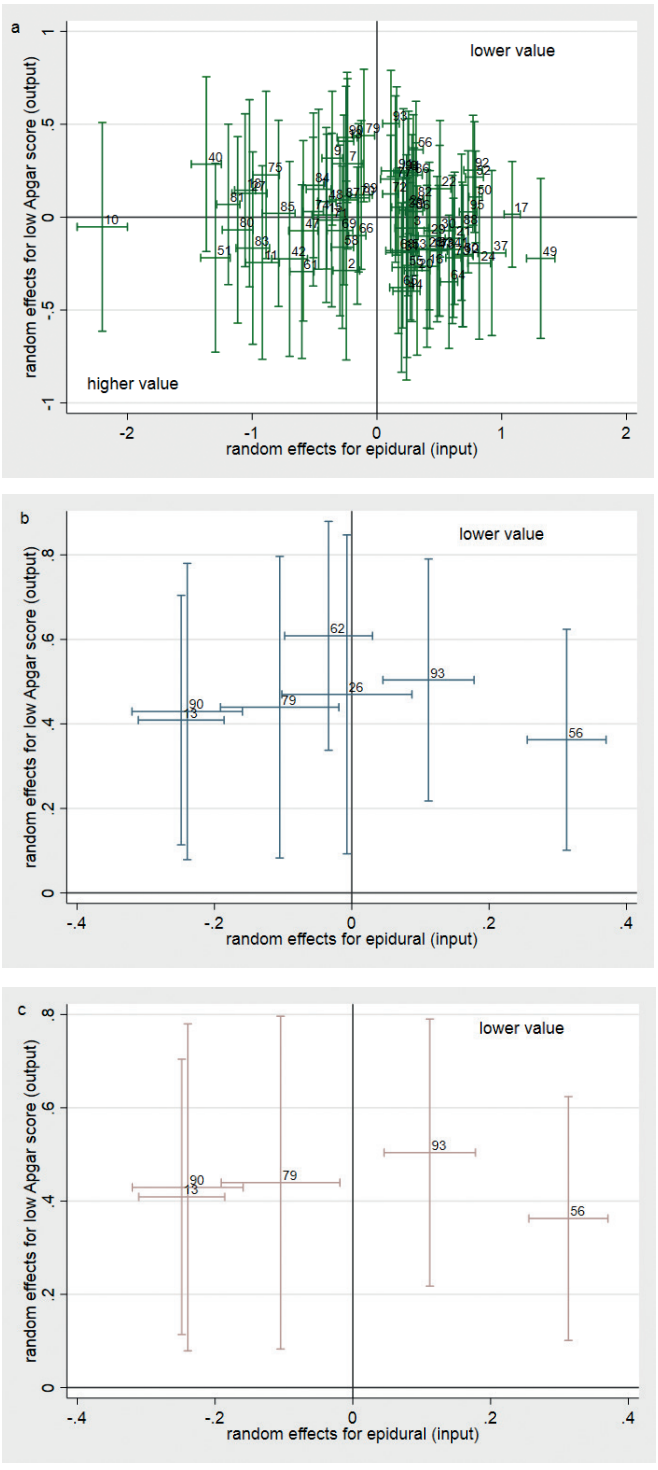


Figure A2.3. Maternity Care Network performances for 'Low Apgar Score after Epidural Analgesia'

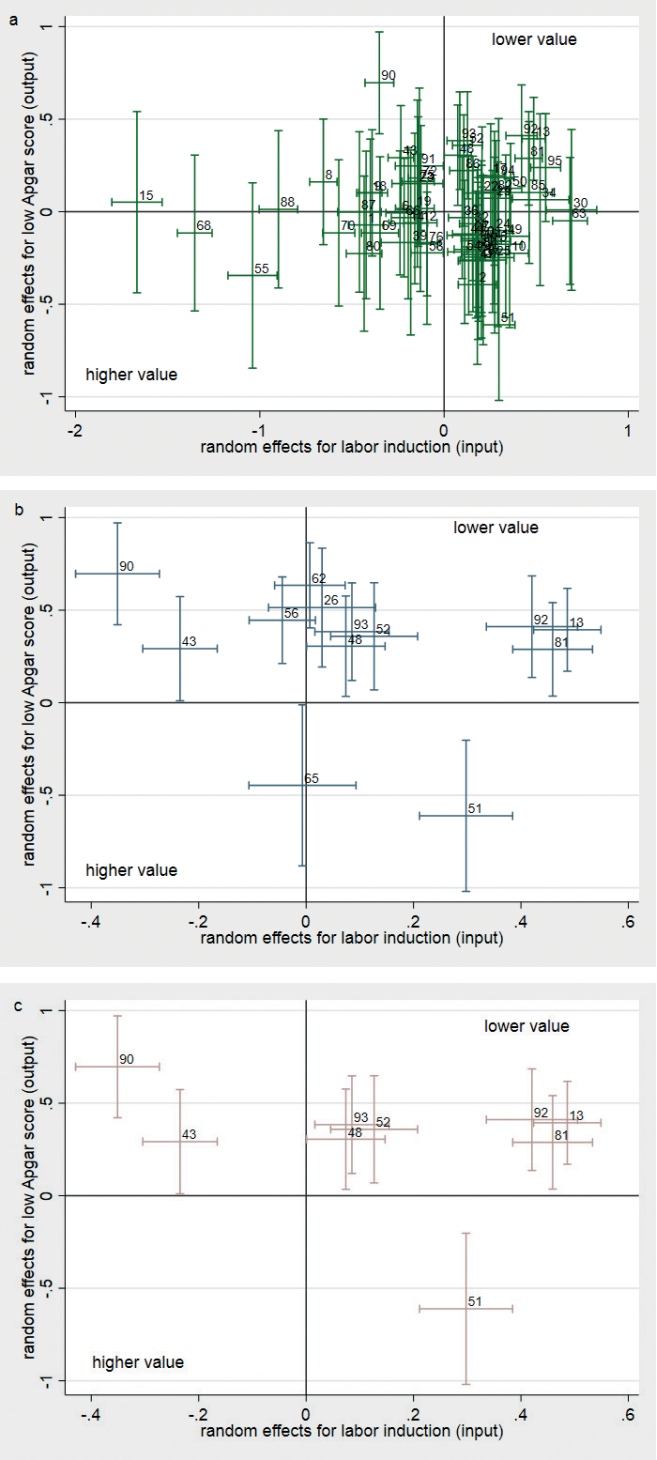


Figure A2.4. Maternity Care Network performances for 'Low Apgar Score after Labor Induction'

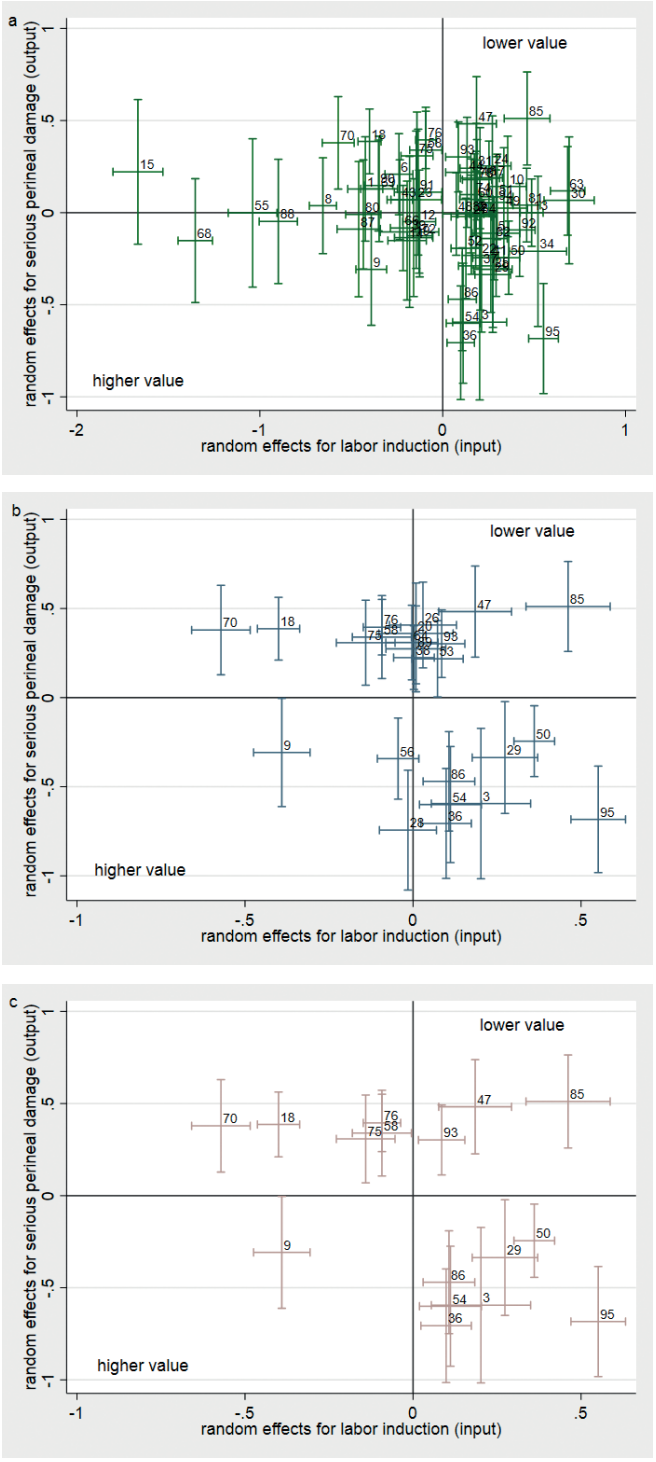


Figure A2.5. Maternity Care Network performances for 'Perineal Damage after Labor Induction

S4. Regression results for the inputs

	Caesarean Section		Epidural Analgesia		Labor Induction	
main						
Age mother	0.08***	(0.00)	-0.00	(0.00)	0.03***	(0.00)
Educational level mother (category, from low to high)						
2	0.13*	(0.06)			0.24***	(0.06)
3	0.12*	(0.05)			0.20***	(0.05)
4	-0.04	(0.05)			-0.01	(0.05)
5	-0.32***	(0.05)			-0.22***	(0.05)
6	-0.46***	(0.05)			-0.33***	(0.05)
7	-0.44***	(0.06)			-0.30***	(0.06)
Household income (category, from low to high)						
2			0.03	(0.03)	0.03	(0.03)
3			-0.05*	(0.03)	-0.03	(0.03)
4			-0.10***	(0.02)	-0.08**	(0.03)
5			-0.14***	(0.02)	-0.11***	(0.03)
6			-0.18***	(0.02)	-0.13***	(0.03)
7			-0.24***	(0.02)	-0.14***	(0.03)
8			-0.27***	(0.02)	-0.13***	(0.03)
9			-0.28***	(0.03)	-0.11***	(0.03)
10			-0.19***	(0.03)	-0.10***	(0.03)
Ethnicity mother						
Western	-0.37***	(0.02)	-0.27***	(0.01)	-0.25***	(0.01)

	Caesarean Section		Epidural Analgesia		Labor Induction	
Urbanity mother (category, from low to high)						
2	0.03	(0.02)	0.03*	(0.01)		
3	0.03	(0.02)	-0.00	(0.02)		
4	-0.01	(0.02)	-0.02	(0.02)		
5	0.02	(0.03)	-0.05*	(0.02)		
Gestational age (weeks)						
38	0.01	(0.03)	0.01	(0.02)	-0.10***	(0.02)
39	-0.15***	(0.03)	-0.10***	(0.02)	-0.37***	(0.02)
40	-0.10***	(0.03)	-0.01	(0.02)	-0.29***	(0.02)
41	0.33***	(0.03)	0.32***	(0.02)	0.33***	(0.02)
Birth weight (categories)						
Middle	0.07***	(0.02)	0.06***	(0.01)	0.02	(0.01)
Heavy	0.54***	(0.02)	0.20***	(0.01)	0.26***	(0.01)
Gestational hypertension						
Yes	0.52***	(0.02)			1.68***	(0.02)
Gestational diabetes						
Yes	0.62***	(0.03)	0.55***	(0.03)	1.28***	(0.04)
Gravidity						
>1	-0.53***	(0.01)	-0.82***	(0.01)	-1.29***	(0.01)
Maternal health spending level						
Middle	0.16***	(0.01)	0.32***	(0.01)	0.17***	(0.01)

	Caesarean Section		Epidural Analgesia		Labor Induction	
High	0.33***	(0.02)	0.49***	(0.01)	0.36***	(0.01)
Gender baby						
Male	0.20***	(0.01)			0.03**	(0.01)
Age father						
>45	0.07***	(0.02)	0.11***	(0.01)	0.05***	(0.01)
Year of birth						
2012	-0.04	(0.02)	0.01	(0.02)	-0.05**	(0.02)
2013	-0.03	(0.02)	0.11***	(0.02)	-0.00	(0.02)
2014	-0.05*	(0.02)	0.20***	(0.02)	-0.07***	(0.02)
2015	-0.07**	(0.02)	0.37***	(0.02)	-0.09***	(0.02)
2016	-0.12***	(0.02)	0.33***	(0.02)	-0.66***	(0.02)
_cons	-4.17***	(0.07)	-0.65***	(0.08)	0.01	(0.07)
var(_cons[MCN])	0.06***	(0.01)	0.35***	(0.05)	0.14***	(0.02)
N	261114		244193		220068	
AIC	187837.79		277218.28		257433.03	
BIC	188172.92		277561.67		257814.19	

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

S5. Regression results for the outputs

Input	**Caesarean Section**				**Epidural Analgesia**		**Labor Induction**					
Output	**Low Apgar Score**		**Hemorrhage**		**Low Apgar Score**		**Low Apgar Score**		**Hemorrhage**		**Perineal Damage**	
main												
Age mother			0.02**	(0.00)					0.04***	(0.00)		
Educational level mother (category, from low to high)												
2			0.19	(0.23)					0.24*	(0.12)		
3			0.02	(0.19)					0.11	(0.10)		
4			0.11	(0.18)					0.23*	(0.10)		
5			0.21	(0.18)					0.26**	(0.10)		
6			0.18	(0.19)					0.21*	(0.10)		
7			0.11	(0.24)					0.20	(0.12)		
Household income (category, from low to high)												
2	0.18	(0.17)			0.13	(0.14)	0.13	(0.12)			0.01	(0.10)
3	-0.36	(0.21)			-0.01	(0.16)	-0.18	(0.14)			0.23*	(0.10)
4	-0.01	(0.18)			0.08	(0.14)	0.07	(0.12)			0.27**	(0.09)
5	-0.09	(0.17)			0.14	(0.14)	0.07	(0.12)			0.30***	(0.09)
6	-0.05	(0.17)			0.12	(0.14)	0.06	(0.12)			0.34***	(0.09)
7	0.03	(0.17)			0.22	(0.14)	0.13	(0.11)			0.28**	(0.09)
8	0.01	(0.17)			0.16	(0.14)	0.07	(0.12)			0.31***	(0.09)
9	0.12	(0.17)			0.19	(0.14)	0.08	(0.12)			0.19*	(0.09)
10	-0.19	(0.18)			-0.11	(0.15)	-0.10	(0.13)			0.30***	(0.09)

Input	Caesarean Section				Epidural Analgesia		Labor Induction					
Output	**Low Apgar Score**		**Hemorrhage**		**Low Apgar Score**		**Low Apgar Score**		**Hemorrhage**		**Perineal Damage**	
Ethnicity mother												
Western	-0.05	(0.09)	0.03	(0.06)	-0.00	(0.07)	-0.05	(0.06)	0.30***	(0.03)	-0.15***	(0.05)
Urbanity mother (category, from low to high)												
2			-0.04	(0.07)					0.05	(0.03)		
3			0.07	(0.07)					0.07*	(0.03)		
4			-0.17*	(0.08)					0.05	(0.04)		
5			0.03	(0.09)					0.10*	(0.04)		
Gestational age (weeks)												
38	-0.02	(0.15)	-0.09	(0.10)	-0.18	(0.12)	-0.06	(0.10)	-0.09*	(0.04)	0.08	(0.08)
39	-0.03	(0.14)	-0.28**	(0.10)	-0.17	(0.12)	-0.05	(0.09)	-0.11**	(0.04)	0.20**	(0.07)
40	0.17	(0.14)	-0.39***	(0.10)	-0.02	(0.11)	0.06	(0.09)	-0.07	(0.04)	0.17*	(0.07)
41	0.37**	(0.14)	-0.44***	(0.10)	0.36**	(0.12)	0.41***	(0.09)	-0.01	(0.04)	-0.01	(0.08)
Birth weight (categories)												
Middle	-0.27**	(0.08)	0.49***	(0.07)	-0.10	(0.07)	-0.15*	(0.06)	0.34***	(0.03)	0.48***	(0.05)
Heavy	-0.72***	(0.09)	1.01***	(0.07)	-0.28***	(0.08)	-0.32***	(0.06)	0.77***	(0.03)	0.81***	(0.05)
Gestational hypertension												
Yes	0.12	(0.09)	0.01	(0.06)	0.10	(0.08)	0.13*	(0.06)	0.22***	(0.03)	-0.07	(0.04)
Gestational diabetes												
Yes	0.05	(0.16)			0.33**	(0.13)	0.29**	(0.11)			-0.17*	(0.08)
Gravidity												

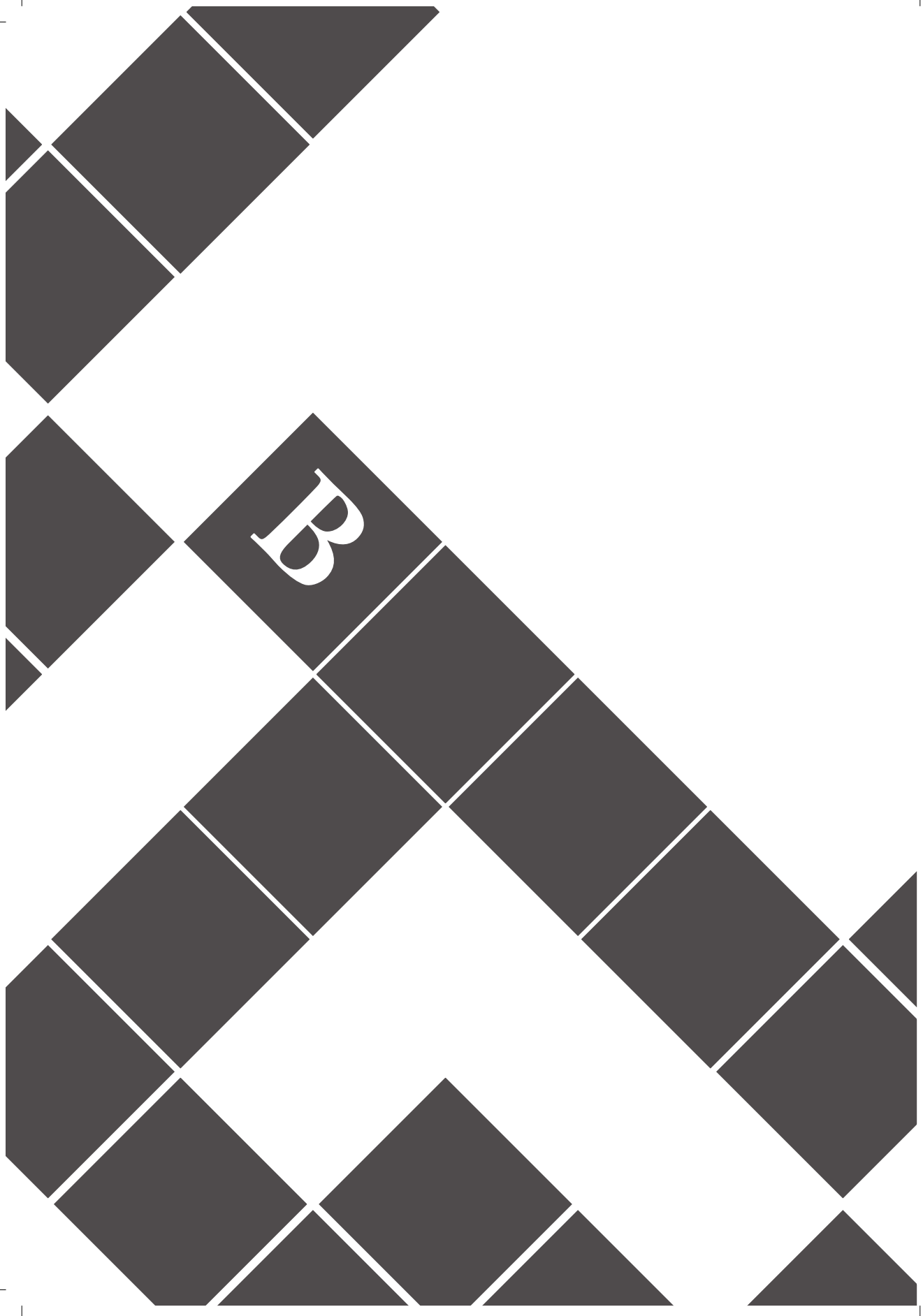
Input	**Caesarean Section**				**Epidural Analgesia**		**Labor Induction**					
Output	**Low Apgar Score**		**Hemorrhage**		**Low Apgar Score**		**Low Apgar Score**		**Hemorrhage**		**Perineal Damage**	
>1	0.01	(0.07)	-0.00	(0.05)	0.06	(0.06)	0.11*	(0.05)	0.01	(0.02)	-0.14***	(0.04)
Maternal health spending level												
Middle	-0.02	(0.08)	0.02	(0.05)	0.09	(0.06)	0.11*	(0.05)	0.01	(0.02)	-0.11**	(0.04)
High	0.01	(0.09)	0.11	(0.06)	0.17*	(0.07)	0.20***	(0.06)	0.01	(0.03)	-0.15***	(0.04)
Gender baby												
Male	0.28***	(0.07)	-0.24***	(0.04)	0.40***	(0.06)	0.39***	(0.05)	-0.24***	(0.02)	0.02	(0.03)
Age father												
>45	0.12	(0.08)	-0.02	(0.06)	0.13	(0.07)	0.12	(0.06)	-0.07**	(0.03)	-0.09	(0.04)
Year of birth												
2012	0.18	(0.12)	-0.10	(0.08)	0.02	(0.10)	0.07	(0.08)	0.03	(0.03)	-0.06	(0.05)
2013	0.25*	(0.12)	-0.12	(0.08)	0.21*	(0.10)	0.13	(0.08)	0.01	(0.03)	-0.09	(0.05)
2014	0.19	(0.12)	-0.10	(0.08)	0.15	(0.10)	0.17*	(0.08)	-0.01	(0.03)	-0.15**	(0.05)
2015	0.21	(0.12)	-0.08	(0.08)	0.12	(0.10)	0.24**	(0.08)	0.01	(0.03)	-0.10	(0.05)
2016	0.40***	(0.12)	0.02	(0.08)	0.25*	(0.10)	0.15	(0.09)	0.25***	(0.04)	0.22***	(0.05)
_cons	-3.81***	(0.22)	-3.46***	(0.26)	-4.59***	(0.18)	-4.56***	(0.14)	-4.34***	(0.13)	-3.99***	(0.12)
var(_cons[MCN])	0.12**	(0.04)	0.19***	(0.04)	0.09***	(0.03)	0.11***	(0.02)	0.04***	(0.01)	0.09***	(0.02)
N	32551		33561		76655		134512		140726		134512	
AIC	8201.45		16179.46		13840.77		20143.65		77402.67		37557.84	
BIC	8453.16		16440.51		14118.18		20437.93		77708.16		37852.12	

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

S6. Correlations between inputs and outputs

	Output		
Input	**Low Apgar Score**	**Hemorrhage**	**Perineal Damage**
Caesarean section	-0.37*** [-0.53; -0.20]	-0.16 [-0.36; 0.05]	-
Epidural Analgesia	-0.20* [-0.39; -0.01]		
Labor Induction	-0.07 [-0.27; 0.13]	0,01 [-0.20; 0.21]	-0.12 [-0.31; 0.07]

Estimates are Spearman correlation coefficient (rho) and bootstrapped 95% confidence interval for 1000 replications in brackets. N = 91. * p < 0.05, ** p < 0.01, *** p <0.001


B

PART B

Alternative payment models in population health management


5

CHAPTER 5

Alternative payment models in maternity care and their effects on health and spending: a scoping review

Based on
De Vries EF, De Bruin-Kooistra, M, Baan CA, Struijs JN
Alternative payment models in maternity care and their effects on health and spending: a scoping review
(Submitted)

ABSTRACT

Although effects of alternative payment models (APM) on health and spending are unclear, they are increasingly implemented across countries. This study provides an overview of APMs in maternity care, describing key elements and empirical evidence. A scoping review was performed in PubMed, Embase and Scopus for articles published from January 2007 through November 2018. Key elements (status, type of APM and risk mitigation strategies) were collected and evidence regarding health and spending was examined. We identified seventeen initiatives that employed pay-for-performance (n=2), shared savings (n=7) and bundled payments (n=8). Key terms (e.g. shared savings and bundled payments) were used interchangeably. APMs that shifted more financial accountability towards providers included more strategies that mitigated risks. Four studies evaluated effects of APMs; two found associated improved outcomes and two associated reduced spending. Clear definitions of key elements and understanding of how they operate into different settings is required to interpret future evidence and shape payment reform.

INTRODUCTION

Policies in both US and Europe aim for the reduction of avoidable infant mortality, pre-term rates and maternal mortality. Previous reports show that improvements can be made by optimizing the delivery of health services [1, 2]. For instance, low-value services, such as non-medically indicated caesarean sections, are increasingly performed, while high-value services, such as screening for gestational diabetes or educating women on what to expect during and after birth, are underutilized [2, 3]. To achieve optimally organized care, more coordination of care delivery is needed.

The promise of alternative payment models (APMs) is that they incentivize care coordination, and stimulate the use of high-value care and discourage the use of low-value care by increased provider accountability. In the literature, theoretical effects of APMs have been discussed [4-8] in comparison to effects of fee-for-service (FFS) models, which are commonly used in daily practice. The financial risks in the FFS models are borne largely by the payers. Since health care providers thereby run no financial risk in terms of the volume and the value of care they deliver, a FFS system inadvertently encourages providers to deliver larger volumes of care and low-value care [4-7, 9]. APMs aim to remove these incentives by shifting the accountability, for health outcomes and health spending, towards providers [4]. This serves as an incentive to avoid unnecessary care as well as encourages other cost-conscious behaviors such as downward substitution of care, task reallocation and more efficient coordination between practitioners within care. The scope of the APM determines the allocation of financial risk between provider and payer. Ideally, the performance risk (i.e. risks that are related to the providers own share in providing high quality and efficient care [7]) is allocated with the provider and the insurers risk (i.e. risks that stem from patients and their respective needs [7]) with the payer. Performance risk increases the incentives to create value, whereas insurance risk increases a providers' level of financial risk without the provider being able to control it. Therefore, the optimal allocation of risks is where, for providers, insurance risk is minimized and the performance risk is maximized [8]. To optimize the financial risk for providers, APMs may be refined by adding components such as bonuses or penalties when certain quality targets are met (e.g. in shared savings), or adding risk adjustments or stop loss provisions (i.e. risk mitigation strategies).

New contribution

Based on the promises of better quality of care, APMs are increasingly implemented in maternity care, for example in Arkansas [10-12] and in the Netherlands [13]. All adopted a different strategy for which there is currently no strong empirical evidence.

Yet, further implementation of APMs is likely to continue for maternity care [14, 15]. However, to understand (future) empirical evidence, there is a need to gain insight into the key design elements of the APMs. As far as we are aware, such an overview of APMs in maternity care and their effects on health and spending, is currently lacking. Therefore, this study focused on the following research questions:

1) What are the key elements of APMs currently implemented in maternity care?
2) What evidence is available with regard to the empirical effects of such payment models on the maternal and neonatal health outcomes and perinatal spending?

Conceptual framework

Based on previous research [4, 8, 16, 17], we developed a framework in which the key elements of APMs gain insight into their level of integration of financial accountability and the level of integration of providers over domains of care. We defined APMs as initiatives that include changing the financing of care delivery that aimed to improve maternal and/or infant health outcomes and reduce utilization and/or health spending. Our definitions of the types of APMs are shown in Table 1. Frakt and Mayes (2012) showed that the level of financial risk varies with the payment model For example, under FFS, payers bear more risk than providers and under global payment providers bear more risk than payers (Frakt and Mayes 2012). This creates an incentive to collaborate with providers in other domains in order to reallocate and coordinate care efficiently. Yet, the level of financial risk a provider is bearing can be mitigated by 'risk mitigation strategies'. Such risk mitigation strategies prevent providers from bearing too much risk (i.e. insurers' risk) and include high-risk population exclusions and/or risk adjustment to account for the severity of the population in the episode price or benchmarks for shared savings. Another example of a risk mitigation strategie is a 'stop-loss provision', which is a threshold that caps the maximum amount to which the provider is at risk [18] and can be used at the individual level to limit the risk of e.g. high-cost Neonatal Intensive Care Unit (NICU) care or at the aggregate level [19].

METHODS

Search strategy and information sources

To identify as many initiatives as possible that implemented APMs for maternity care, we reviewed the international literature. The process was guided by the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) to ensure validity and reliability [22]. In collaboration with a librarian, we developed a search strategy for

Table 1: APM types and definitions (based on De Vries et al. [17])

APM type	Definition
Pay-for-performance	In pay-for-performance, a bonus/malus is paid for attaining certainty quality thresholds on top of the base FFS payment. Under FFS, providers are paid a fee for each service delivered [4]. The additional payments can be employed for improving coordination, care efficiency, quality of care or accessibility of care [20].
Shared savings	In a shared savings model, individual providers are each paid on a FFS basis and then there is reconciliation between the target episode price and the actual episode average price after a period of time across all the episodes attributed to a provider. Based on a specific formula, which is either negotiated or established by the payer, the accountable provider can share in gains and/or losses with the payer. Shared saving models that only share in gains are called *one-sided*. In t*wo-sided* models also risks are shared.
Bundled payments	Bundled payments are defined for a specific set of activities tied to an episode of care, such as maternity care, that includes more than one provider or organization. The entity receiving the bundled payment earns a higher margin if a patient has utilized less care, but also bears the financial risk of complications. In our definition, the main difference with shared savings is that savings or losses are not shared with the payer. There are two types of bundled payments, retrospective and prospective. In *retrospective* bundled payment, there is a virtual budget negotiated upfront, providers are paid by FFS and retrospectively, the target price is reconciled [21]. *Prospective* bundled payment pays a prospectively defined prize that is paid as one payment to the accountable entity that in turn pays the individual providers [16].
Global payments	In global payments the entire population and the entire continuum of care is included. The accountable provider is paid a fixed fee per head of the population.

APM: Alternative Payment Model; FFS: fee-for-service

PubMed, Embase and Scopus databases for articles published from January 2007 through November 2018. Search key words included 'payment', 'funding', 'alternative payment', 'value based payment', 'maternity ward', 'obstetric', etc. The full search strategy is displayed in the Appendix.

Eligibility criteria

English or Dutch language articles were found eligible if they described and/or empirically evaluated, APMs in maternity care in high-income countries. Additional relevant articles and grey literature (e.g. government reports, white papers) were

identified through reference tracking and recommendations from experts. Articles were excluded if they were commentary articles.

Study selection

First, duplicates were removed. Based on title and abstract, the remaining articles were screened for eligibility by two researchers (EdV and MdBK) independently. Differences were discussed and, if there remained any doubt, the full-texts were retrieved to reach consensus on whether or not to include the article.

Data extraction and synthesis

From the full-text articles, the following information was extracted: first author, year of publication, country, publication type, name of initiative and key elements of the implemented APMs:

- Type of APM
- Care providers that participated in the model
- Accountable entity
- Care activities that are covered by the model
- Link of the model with quality of care.

From studies that empirically evaluated APMs, we additionally extracted information on the research method, data collection period and results of the payment model on health outcomes and healthcare spending. We assessed the quality of the evidence using the Effective Public Health Practice Project Quality Assessment Tool for Quantitative Studies [23]. This tool provides an overall methodological rating of the article: strong, moderate or weak based on assessment of six components (selection bias, study design, confounders, blinding, data collection methods and withdrawals and dropouts). The quality appraisal was performed by EdV and MdBK, independently. Discrepancies were resolved by discussion until consensus was reached. The evidence of the studies was not pooled.

If blanks or uncertainty remained, the authors of the articles were requested to provide additional information through e-mail. EdV subtracted the data. JS and MdBK checked this randomly.

RESULTS

Study selection and characteristics

Figure 1 shows the study selection flow diagram. We identified 149 articles through a search in the peer-reviewed international literature. Reference tracking yielded an additional 31 documents of which also non-scientific articles, white papers, government documents and blogs. After removal of duplicates and non-eligible articles, our final sample consisted of six articles from peer-reviewed journals and 30 government documents, white articles or other documents. Four peer-reviewed articles performed an empirical evaluation of the APM on health outcomes and health spending. The documents and document types included in this review are fully listed in the Appendix.

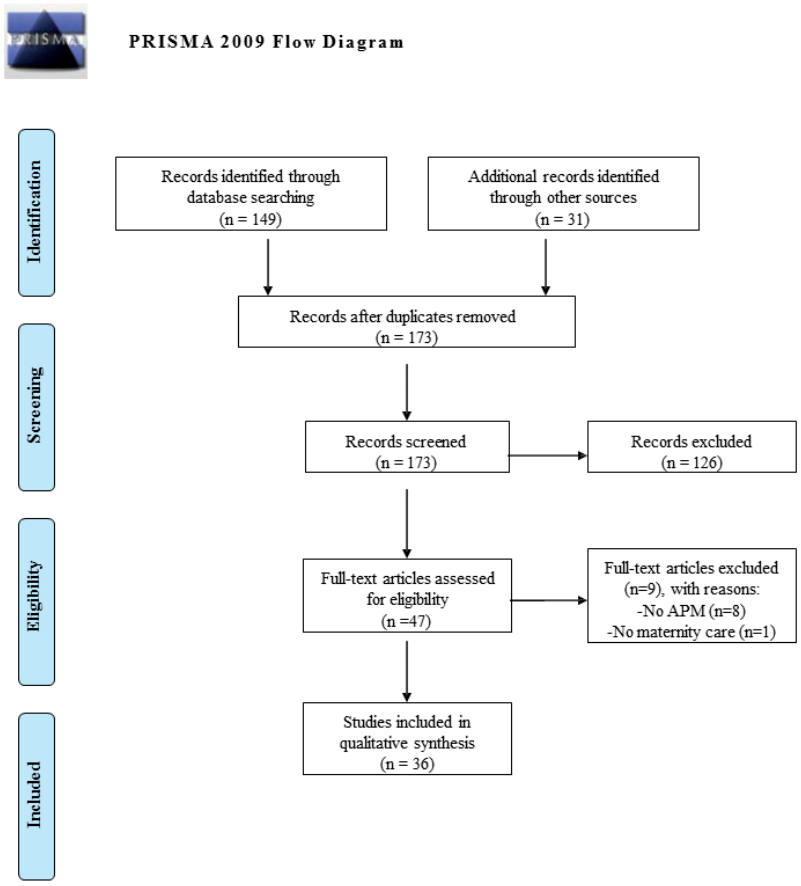


Figure 1: Study selection flow diagram according to Preferred Reporting Items for Systematic Review and Meta-Analysis (PRISMA)

Key elements of APMs in maternity care

General characteristics

In the 36 articles, we identified 17 initiatives that had implemented an APM (Table 2). Most of them are from the United States (n=13) and further initiatives are found in the United Kingdom (n=2), New Zealand (n=1) and the Netherlands (n=1). The earliest APMs were implemented in 2007 (GHS (16) and LMC (17)) and the most recent in 2017 (Dutch BP (13)). Most initiatives are established on a permanent basis (n=11); five others were pilots, and for one the status is unknown.

Type of APM

Table 2 also shows the type of APM of the 17 initiatives. The APMs are classified into three categories: pay-for-performance (n=2), shared-savings models (n=7) and bundled payment models (n=8). For a detailed overview, see Appendix B.

Pay-for-performance is applied in two initiatives (1,2). In the CQUIN initiative in England (1), hospitals are paid bonuses if they satisfy specified scores on a set of quality indicators pertaining to elective and emergency Caesarean sections. This pay-for-performance system is superimposed onto the existing FFS model, hence not replacing the existing payment structure. In the Texas Medicaid Program (2) there is a penalty for hospitals that delivered neonatal deliveries before 37 weeks gestation that are not medically necessary; these billing codes are ineligible for reimbursement.

Shared savings models are implemented in seven initiatives (Horizon (3), Baby+Company (4), TennCare (5), Arkansas (6), Ohio (7), CHC (8) and New York (9)). Those are contracts whereby any achieved savings is shared between providers and payers. Such savings are calculated by comparing the health care spending for the risk adjusted population included in the payment model either with the spending for a predefined control group (i.e. concurrent accountable providers) (TennCare (5), Arkansas (6), Ohio (7)) or with the spending for the intervention population in years preceding the implementation of the APM (historical benchmark) (Horizon (3), CHC (8)). If savings are achieved for the intervention population in comparison with the control group or the historical benchmark, those savings are partially distributed to the providers, resulting in for example 50% for the providers and 50% for the payers (TennCare (5)).

The shared-savings contracts vary in the degree to which health care providers bore financial risks in the event of spending overruns. Two initiatives, Horizon (3) and Baby+Company (4), operated a one-sided shared-savings model, whereby providers bear no downside risks if budgets are exceeded but share in any savings achieved. Five

Table 2: key elements of implemented alternative payment models in maternity care (more detailed information in Appendix B)

No	Initiative (abbreviation)	General characteristics			Type of APM	Risk mitigation strategies	References
		Country	Year of implementation	Status			
1	Commissioning for Quality and Innovation, Payment Framework (CQUIN)	UK	2007	Permanent implementation	Pay for performance	Unclear	[24]
2	Texas Medicaid Program (Texas)	US (Texas, and later on also Georgia, Michigan, New Mexico, New York and South Carolina)	2011	Developed into initiative no. 3	Pay for performance	Unclear	[25, 26]
3	Horizon Blue Cross Blue Shield of New Jersey, Pregnancy Episodes of care program (Horizon)	US (New Jersey)	2013	Permanent implementation	Shared savings, upside only.	Exclusions of high-risk and types of care.	[26]
4	Baby+ Company	US (North Carolina, Tennessee, Colorado, Arkansas)	2013	Permanent implementation	Shared savings (one-sided).	Exclusions of high-risk and types of care.	[19]

Table 2: Key elements of implemented alternative payment models in maternity care (more detailed information in Appendix B) (continued)

No	Initiative (abbreviation)	General characteristics			Type of APM	Risk mitigation strategies	References
		Country	Year of implementation	Status			
5	TennCare	US (Tennessee)	2013	Permanent implementation	Shared savings (two-sided)	Exclusions of high-cost and high-risk and types of care. Total averages for benchmark for shared savings were risk-adjusted.	[11, 27-31]
6	Arkansas Health Care Payment Improvement Initiative (Arkansas)	US (Arkansas)	2013	Permanent implementation	Shared savings (two-sided).	Exclusions of high-cost and high-risk and types of care. Total averages for benchmark for shared savings were risk-adjusted. Stop gain for savings.	[11, 12, 19, 32-35]

Table 2: Key elements of implemented alternative payment models in maternity care (more detailed information in Appendix B) (continued)

No	Initiative (abbreviation)	General characteristics			Type of APM	Risk mitigation strategies	References
		Country	Year of implementation	Status			
7	Ohio Episode-Based Payment Model (Ohio)	US (Ohio)	2015	Permanent implementation	Shared savings (two-sided).	Exclusions of high-cost and high-risk and types of care. Total averages for benchmark for shared savings were risk-adjusted.	[11, 19, 30, 36]
8	Community Health Choice, Bundled Payment Pilot (CHC)	US (Texas)	2015	Pilot	Shared savings (two-sided)	Exclusions of high-risk care. Risk adjustment. Individual stop loss provision.	[19, 37, 38]
9	New York State's Medicaid Maternity Care Value Based Payment Arrangement (New York)	US (New York state)	2016	Pilot	Shared savings (two-sided)	Exclusions of high risk. Risk adjustment and stop loss provision for savings and losses.	[39-41]

Table 2: Key elements of implemented alternative payment models in maternity care (more detailed information in Appendix B) (continued)

No	Initiative (abbreviation)	General characteristics			Type of APM	Risk mitigation strategies	References
		Country	Year of implementation	Status			
10	Pacific Business Group on Health, Blended Case Rate (Pacific)	US (Southern California)	2014	Pilot	Bundled payment (only delivery phase)	Exclusions for high-risk. No prospective risk-adjustment.	[19, 33, 42-44]
11	Minnesota Blended Payment (Minnesota BP)	US (Minnesota)	2009	Permanent implementation	Bundled payment (only delivery phase)	Exclusions of high-risk.	[45, 46]
12	Minnesota Birth Centers, BirthBundle (BirthBundle)	US (Minnesota)	2015	Pilot stopped	Bundled payment (retrospective).	Exclusions of facility fees.	[19, 47]
13	Dutch Bundled Payment for maternity care (The Dutch BP)	The Netherlands	2017	Pilot	Bundled payment (prospective)	Exclusions of types of care and bundle breakers*. Depending on the contract; risk corridors. Risk stratification for bundle tariff.	[13, 48-50]

Table 2: Key elements of implemented alternative payment models in maternity care (more detailed information in Appendix B) (continued)

No	Initiative (abbreviation)	General characteristics			Type of APM	Risk mitigation strategies	References
		Country	Year of implementation	Status			
14	Maternity Pathway Bundled Payment (Maternity Pathway BP)	England	2013	Permanent implementation	Bundled payment (prospective)	Exclusions of types of care. Risk stratification for bundle tariff.	[51]
15	Providence Health and Services, Pregnancy Care Package (Providence)	US (Oregon)	2013	Permanent implementation	Bundled payment (prospective)	Unclear	[19]
16	Geisinger Health System, Perinatal ProvenCare Initiative (GHS)	US (Pennsylvania)	2007	Permanent implementation	Bundled payment (prospective)	Exclusions of high-risk and types of care.	[19, 33, 52]
17	Lead Maternity Care Model (LMC)	New Zealand	2007	Permanent implementation	Bundled payment (prospective)	Exclusions of high-risk and types of care.	[53-55]

* bundle breakers (i.e. when pregnant woman decides to see care professionals outside of the integrated care network.

initiatives, TennCare (5), Arkansas (6), Ohio (7), CHC (8) and New York (9), agreed in their contracts that the providers must reconcile any spending overruns (two-sided models). In the more recently launched initiatives CHC (8) and New York (9), the shared-savings contracts employed one-sided models in the first year but were converted to two-sided contracts in the second year, thus gradually shifting more financial risks towards the providers. Four of the shared saving models (Horizon (3), TennCare (5), Arkansas (6) and Ohio (7)) appoint the provider (group) that delivers the baby as the accountable entity. In the shared-savings models, the distribution of savings is contingent on achieved improvements in quality. In the New York scheme (9), provider penalties for exceeding budgeted spending are reduced or eliminated for those scoring high on the quality indicators.

Eight initiatives (10, 11, 12, 13, 14, 15, 16, 17) implemented variants of bundled payment models that vary highly in terms of shifting the accountability. Two initiatives (Pacific (10) and Minnesota BP (11)) only cover care in the delivery phase for which the hospital is accountable. A fixed fee is negotiated for deliveries, irrespective of whether vaginal or Caesarean. In the Minnesota BP (11) program, complicated vaginal deliveries are excluded. Although the payment model is limited to the hospital (i.e. one provider only), this model contains a financial incentive to perform fewer Caesarean sections. Therefore, one may conclude that the financial risk in this model is partially shifted from the payer to the care provider(s). That is why we decided to classify this model as an APM and included it in our overview.

One initiative (BirthBundle (12) implemented a retrospective model. The integrated fee that is charged for a maternity care episode is in fact a 'virtual' fee, which is reconciled at the end of the episode by totaling the FFS for all the services delivered. If the spending turned out lower or higher than the virtual fee, the difference is transferred to the accountable entity, which is the birth center (BirthBundle (12)).

Five initiatives (The Dutch BP (13), Maternity Pathway BP (14), Providence (15), GHS (16) and LMC (17)) implemented prospective bundled payment models. All the services specified in the entire maternity care program (or split up into three or four phases (prepartum, delivery and postpartum) (The Dutch BP (13), Maternity Pathway BP (14), LMC (17)) are contracted, delivered and claimed as a single product by the accountable nurse or midwife or obstetrician. These prospective bundled payment models replaced the existing FFS models. No reconciliations are performed as in the retrospective models.

Risk mitigation

The pay-for-performance models (n=2) employ no risk mitigation strategies.
In the one-sided shared savings models (n=2), the Horizon program (3) includes both low- and high-risk pregnancies, but excludes several comorbidities in pregnancy such as HIV and cancer and neonatal care to set the benchmark. Baby+Company (4) only includes low-risk pregnancies. Other risk mitigation strategies were not found.

The two-sided shared savings models (n=3) (TennCare (5), Arkansas (6) and Ohio (7)) includes only low-risk pregnancies and excludes pregnancies ending in stillbirth. In addition, women with several comorbidities are excluded from the model, as is care for the newborn. These three initiatives (5, 6, 7) include a stop-loss provision for individual cases that exceeds more than three standard deviations above the risk-adjusted mean. CHC (8) and New York (9) includes also high-risk pregnancies in addition to newborn care, and applies an aggregate stop-loss provision.

For the retrospective bundled payment model (n=1), BirthBundle (12), we did not find any risk mitigation strategies.

The prospective bundled payment models (n=5) (The Dutch BP (13), Maternity Pathway BP (14) and LMC (17)) includes low- and high-risk pregnancies. Providence (15) and GHS (16) only includes low-risk pregnancies. GHS (16) excludes late referrals or members not enrolled at least 12 continuous weeks of the prenatal phase. The Maternity Pathway BP (14) and the Providence (15) initiative also include care for the newborn, although health problems in the newborns are excluded in the Maternity Pathway BP (14). In the other initiatives (16, 17), care for the newborn is excluded. In the Netherlands (13), in most regions, stop-loss provisions are applied to which new negotiations would start to decide what to do with the gains or losses. Other risk mitigation strategies were not found. As a quality assurance measure with respect to the care delivered, payers generally required providers to file yearly reports on quality indicators (accountability data).

Table 3 provides a schematic overview of the various payment models, showing the level of integration and the level of financial accountability. APMs that employ the highest level of integration and highest level of financial accountability are Maternity Pathway Bundled Payment, GHS and LMC.

Table 3: Level of integration and financial accountability of the alternative payment models in maternity care

	Fee for service				Fee for service			Bundled payment (prospective)		Global Payment
	A	B	C	D	A	B	C	A	B	C
	Additional fee for coordination	Pay for reporting	Pay for Performance (Rewards for attaining quality threshold)	Pay for performance (penalties for not attaining quality threshold)	Shared savings (one sided)	Shared savings (two-sided)	Bundled payment (retrospective)	Per episode	Per diagnose	Per population
One provider organization			CQUIN	Texas	Baby+ Company		BirthBundle	Pacific*, Minnesota BP*, Providence***		
One echelon (including multiple providers)										
One domain (including multiple echelons and multiple providers)					Horizon	TennCare, Arkansas, Ohio, CHC, New York		The Dutch BP, Maternity Pathway BP**, GHS, LMC**		
Multiple domains (including multiple echelons and multiple providers)										

*episode is limited to the delivery phase in the hospital only, **episode is divided into three or four phases (prenatal, delivery, postnatal, maternity community care), *** integrated birth center; CQUIN: Commissioning for Quality and Innovation Payment Framework; Texas: Texas Medicaid Program; BirthBundle: Minnesota Birth Centers BirthBundle; Pacific: Pacific Business Group on Health Blended Case Rate; Minnesota BP: Minnesote Blended Payment;; Providence: Providence Health and Services Pregnancy Care Package; Horizon: Horizon Blue Cross Blue Shield of New Jersey, Pregnancy Episodes of care Program; Arkansas: Arkansas Health Care Payment Improvement Initiative; Ohio: Ohio Episode-Based Payment Model; CHC: Community Health Choice, Bundled Payment Pilot; New York: New York States' Medicaid Maternity Care Value Based Payment Arrangement; The Dutch BP: The Dutch Bundled Payments for Maternity Care; Maternity Pathway BP: Maternity Pathway Bundled Payment; GHS: Geisinger Health System Perinatal ProvenCare Initiative; LMC: Lead Maternity Carer.

Effects of the APMs on maternal and neonatal health outcomes and health spending

Table 4 shows results of the available evaluations from four of the 17 initiatives that implemented an APM in maternity care (Texas, Arkansas, CHC and GHS). In two studies, a beneficial effect of the APM on the health outcomes was observed [25, 52]. The other two studies that did evaluate the effects of the APM [32, 45] did not show improvement on health outcomes. To studies [32, 45] gauged the effects of the scheme on health care expenditures, reporting positive effects.

Three out of four studies reporting evaluations were assigned moderate rating for the quality of the evidence. No studies evaluating APM in maternity care received a strong rating. Appendix D summarizes the details of quality of the evidence assessments that we conducted.

DISCUSSION

As APMs are increasingly implemented in maternity care, more insight in APMs and empirical evidence about the effects of APMs on maternal and infant health outcomes and spending is needed. We identified 17 initiatives implementing APMs in maternity care: pay-for-performance models (n=2), shared savings models (n=7) and bundled payment models (n=8). APMs that shifted more financial accountability towards providers seemed to include more strategies that mitigated those risks. Risk mitigation strategies included population and care exclusions, stop loss provision and risk adjustment. Of the seventeen initiatives, we found four empirical effect studies evaluating the APMs on health outcomes and health spending. Two studies found an association of the APM with an improvement in health outcomes and two studies described an association with a reduction in health spending.

Although the first studies examining the effects of APMs on health outcomes and health spending in maternity care seem tentatively positive, extensive conclusions on the effects of the APMs cannot be drawn; we did not find a 'one size fits all'. That is because it may take considerable time for the impact of an APM to be fully observable. As these studies captured early results, it is unclear what effects are in the longer run. Moreover, multiple types of APM with multiple key elements were implemented of which four were evaluated using varying indicators. Consequently, which element of which APM has which effects in which type of health system is therefore unknown. Further research should address these matters, but results cannot be expected in short term.

Table 4: Effects of the alternative payment model in maternity care on health outcomes and health spending

Initiative (abbreviation)	Type of APM	Study design	Data collection period	Results – health outcomes	Results – spending	Quality Appraisal of the Study	Ref
Geisinger Health System, Perinatal ProvenCare Initiative (GHS)	Bundled payment (prospective)	Observational study design. Pre-intervention period (n = 101); post-intervention period (n = 1,010)	Pre-intervention population: January 2008 – October 2008, Post-intervention population: April 2009 – June 2010.	Improvement on nearly all 103 indicators: e.g. 25% reduction in neonatal intensive care admissions and screening and prevention activities for smoking increased from 45% to 88%.	-	Weak	[52]
Arkansas Health Care Payment Improvement Initiative (Arkansas)	Shared savings (two sided)	Difference-in-difference design. Pre-intervention: n= 2,454 (intervention); n=20,824 (controls). Post-intervention: n= 1,737 (intervention); n=15,291 (controls)	Pre-intervention: 2010 to 2012, Post-intervention: 2013 - 2014	-	Perinatal spending decreased by 3.8% overall. The decrease was driven by the prices paid for inpatient facility care.	Moderate	[32]

Table 4: effects of the alternative payment model in maternity care on health outcomes and health spending (continued)

Initiative (abbreviation)	Type of APM	Study design	Data collection period	Results – health outcomes	Results – spending	Quality Appraisal of the Study	Ref
Minnesota Blended Payment (Minnesota BP)	Bundled payment (only delivery phase)	Interrupted time series design. Experiment group n= 25,080; Control group n=646,097	2006-2012	There were no significant effects on maternal morbidity.	Spending dropped by $425.80. It continued to decrease in by $95.04 per quarter.	Moderate	[45]
Texas Medicaid Program (Texas)	Pay-for-performance	Difference-in-difference design. Experiment group n=438,429. Control group 1 n=895,543; control group 2 n=1,691,896; control group 3 n=573,382	2009-2013	Gains of five days in gestational age and six ounces in birthweight.	-	Moderate	[25]

Yet, to deal with the current demand for payment reform [14], we identified two issues that should be addressed in order to design an APM that fits the health care setting at hand and works towards the desired goals.

First, a detailed understanding of the specific elements of the APMs is required with providers and payers that work to implement APMs. In this review, we found that currently there is a multiplicity of complex terminology and ambiguous definitions that confuses the understanding of APMs. For example, we found that descriptions of APMs often used the terms 'shared savings' and 'bundled payments' rather interchangeably. That was notably the case in initiatives that employed a two-sided shared-savings model or a retrospective bundled payment model. Presumably, the conflation of the two notions arises from the conceptual similarities between the two payment models. For example, the proportion of risk borne by the care providers working under the model is 100% in bundled payment contracts, and it is smaller (perhaps 50% or 70%) in shared-savings contracts. Such a distinction was not clearly made in several of the descriptions, and as such confusing the two terms. A clear definition of terminology will contribute to a better understanding of the key elements of the APMs.

Second, key elements of the APM including risk mitigation strategies may be best designed from the providers' perspective. For example, feasibility considerations may play an important role in designing the APM. We found, for example, that in some initiatives had been superimposed onto the existing FFS model for the single reason that retrospective reconciliation is easier to administer within the current FFS environment (also see [56]). This aligns with the theoretical notion that the 'best' APM shifts only that part of the accountability towards the provider that actually can be influenced by the provider. Careful deliberation that is based on the level of accountability the provider is willing and able to bear, should lead to which type of APM is the most optimal one to apply in order to deliver optimal care for the best attainable health for mothers and children.

Limitations of this study include that we might have missed relevant initiatives implementing APMs in maternity care by, for example, the use of different terminology. Nevertheless, as the aim was to provide an exhaustive list of these initiatives, we deliberately chose to use other sources than peer-reviewed journals, such as government documents and white papers. Therefore, we are confident that we captured the majority of the initiatives implementing APMs in maternity care in high-income countries. Another limitation is that we did not include grey literature in the search for the evaluations on the effects of APMs in maternity care. Although we may have missed relevant insights

in the effects of APMs on health outcomes and spending, we are convinced that we were able to assess the quality of the included studies by using the Effective Public Health Practice Project Quality Assessment Tool for Quantitative Studies (on peer-reviewed evaluation studies only).

Conclusions and implications

In our review, we identified maternity care APMs in the United States, New Zealand, United Kingdom and the Netherlands. All such APMs intended to improve health outcomes and reduce the spending level of maternity care by shifting financial accountability from payers to providers. We found that key terms describing the models, for example 'shared savings' and 'bundled payments', were used interchangeably and key design elements of the models varied highly. Although first evaluations of APMs in maternity seem tentatively positive, due to a variety of model elements and health system characteristics they operate into, extensive conclusions could not be drawn. Further research, clearly defining the key elements and an in-depth understanding of key elements and their effects under the influence of unique characteristics of health systems, is required to understand future evidence and shape payment reform that aligns with its goals.

REFERENCES

1. Euro-Peristat Project, *European Perinatal Health Report. Core indicators of the health and care of pregnant women and babies in Europe in 2015.* 2018.
2. World Health Organization (WHO), *WHO statement on caesarean section rates*. 2015.
3. Euro-Peristat Project, *European Perinatal Health Report. Core indicators of the health and care of pregnant women and babies in Europe in 2015.* 2018.
4. Frakt, A.B. and R. Mayes, *Beyond capitation: how new payment experiments seek to find the 'sweet spot' in amount of risk providers and payers bear.* Health Aff (Millwood), 2012. **31**(9): p. 1951-8.
5. Organisation for Economic Co-operation and Development (OECD), *Health at a Glance 2013: OECD Indicators*. 2013, OECD Publishing: Paris.
6. Robinson, J.C., *Theory and Practice in the Design of Physician Payment Incentives.* Milbank Q, 2001. **79**(2): p. 28.
7. Miller, H.D., *From volume to value: better ways to pay for health care.* Health Aff (Millwood), 2009. **28**(5): p. 1418-1428.
8. Ruwaard, S., *Purchasing healthcare; Beyond setting the financial incentives right*. 2018, Tilburg University: Tilburg.
9. Miller, H. *An alternative payment model for maternity care*.
10. Carroll, C., et al., *Effects of episode-based payment on health care spending and utilization: Evidence from perinatal care in Arkansas.* J Health Econ, 2018. **61**: p. 47-62.
11. Jarlenski, M., et al., *Episode-Based Payment for Perinatal Care in Medicaid: Implications for Practice and Policy.* Obstet Gynecol, 2016. **127**(6): p. 1080-4.
12. Arkansas Health Care Payment Improvement Initiative (AHCPII). (2017). *Principle Accountable Provider Manual: Program Overview.* 2017.
13. Struijs, J.N. and Hargreaves, D.S., *Turning a crisis into a policy opportunity: lessons learned so far and next steps in the Dutch early years strategy.* Lancet Child Adolesc Health, 2019. **3**(2): p. 66-68.
14. Nijagal, M.A., Shah, N.T., and Levin-Scherz, J., *Both patients and maternity care providers can benefit from payment reform: four steps to prepare.* American Journal of Obstetrics and Gynecology, 2018. **218**(4): p. 411.e1-411.e6.
15. Wing, D.A., *The Future of Obstetrics and Gynecology: MACRA, Electronic Health Records, and More.* Clin Obstet Gynecol, 2017. **60**(4): p. 840-852.
16. Alternative Payment Model Framework and Progress Tracking (APMFPT) Working Group, *Alternative Payment Model (APM) Framework Final White Paper*, 2016.
17. De Vries, E., et al., *Barriers to payment reform: experiences from nine Dutch population health management sites.* Health Policy, 2019. 123(11).
18. Porter, M.E. and Kaplan R., *How should we pay for healthcare?* 2015, Harvard Business School.
19. Clinical Episode Payment (CEP) Working, *Accelerating and aligning clinical episode payment models: maternity care*. 2016, The MITRE Corporation.
20. Organisation for Economic Co-operation and Development (OECD), *Better Ways to Pay for Health Care*. 2016, OECD Publishing: Paris.
21. Struijs, J.N., Hayen, A., van der Swaluw, K., *When Designing Bundled Payments, Don't Ignore The Lessons Of Behavioral Economics*, in *Health Affairs Blog*. 2018.
22. Moher, D., L.iberati, A,., Tetzlaff, J., Altman, D.G. The PRISMA Group, *Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement.* PLoS Med 6(6), 2009. **6**(6).
23. Effective Public Health Practice Project (EPHPP). Quality Assessment Tool for Quantitative Studies. 1998. Retrieved from https://merst.ca/ephpp
24. Ertok, M., *Evaluation of a Pay for Performance Scheme in Maternity Care: The Commissioning Quality and Innovation Payment Framework in England.* Procedia Social and Behavioral Sciences, 2015(195): p. 93-102.

25. Dahlen, H.M., et al., *Texas Medicaid Payment Reform: Fewer Early Elective Deliveries And Increased Gestational Age And Birthweight.* Health Aff (Millwood), 2017. **36**(3): p. 460-467.

26. Butcher, L., *Bundled Payment for Bundles of Joy.* Managed care (Langhorne, Pa.), 2018. **27**(2): p. 18-22.

27. Conrad, D.A. and Perry L., *Quality-based financial incentives in health care: can we improve quality by paying for it?* Annu Rev Public Health, 2009. **30**: p. 357-71.

28. Smith, D. and Hanlon C., *Case Study: Tennessee's Perinatal Episode of Care Payment Strategy Promotes Improved Birth Outcomes.* 2017, National Institute for Childen's Health Qulity NICHQ.

29. DeRoche, M., Guerette, P.J., and Janjan, N. , *The perinatal episode of care model: how will it affect you and your high-risk patients?* Contemp Obstet Gynecol, 2015. **1-9**.

30. Healthcare Payment Learning & Action Network (HCP LAN). *Establishing Maternity Episode Payment Models: Experiences from Ohio and Tennessee,* 2018. Retrieved from https://hcp-lan.org/workproducts/mac/Establishing-Maternity-Episode-Payment-Models.pdf.

31. Tennessee Division of Health Care Finance & Administration. *Executive summary: perinatal episode.* 2017. Retrieved from https://www.tn.gov/content/dam/tn/tenncare/documents2/Perinatal.pdf.

32. Carroll, C., et al., *Effects of episode-based payment on health care spending and utilization: Evidence from perinatal care in Arkansas.* Journal of Health Economics, 2018. **61**: p. 47-62.

33. Lally, S., *Transforming Maternity Care: A Bundled Payment Approach. Issue Brief.* 2013, Integrated Healthcare Association.

34. *Arkansas Health Care Payment Improvement Initiative: 2nd Annual Statewide Tracking Report.* 2016, ACHI, Arkansas Center for Health Improvement.

35. Chernew, M., et al., *The Arkansas Payment Improvement Initiative: Early Perceptions of Multi-Payer Reform in a Fragmented Provider Landscape.* The American Journal of Accountable Care, 2015.

36. State of Ohio, *Detailed business requirements perinatal episode.* 2015. Retrieved from: http://medicaid.ohio.gov/Portals/0/Providers/PaymentInnovation/DBR/Perinatal.pdf.

37. De Brantes, F. and Love, K., *A process for structuring bundled payments in maternity care.*, in *NEJM Catalyst*. 2016.

38. De Brantes, F., and Love, K., *Case study: first-year results of a pilot effort at a nonprofit Medicaid HMO in Texas*. 2016.

39. Fish, D., *Introduction to Value Based Payments: Maternity Care in Measurement Year 2017.* 2017, New York Department of Health: New York.

40. New York State Department of Health. *Maternity Care Value Based Payment Arrangement. Measurement Year 2017 Fact Sheet.*, in *NYS Medicaid Value Based Payment.* 2017, New York Department of Health: New York.

41. New York State Department of Health. *A Path toward Value Based Payment: Annual Update. June 2016: Year 2. New York State Roadmap for Medicaid Payment Reform. Delivery System Reform Incentive Payment (DSRIP) Program.* 2016, Department of Health. Medicaid Redesign Team.

42. Rubinstein, B., *Case Study: Maternity Payment and Care Redesign Pilot.* 2015, Pacific Business Group on Health.

43. Rubinstein, B., *Call to Member Action: Transform Maternity Care. Four strategies to decrease the C-section rate and reduce hospital C-section variation in California*. Pacific Business Group on Health.

44. Integrated Health Association, *Bundled Episode Payment and Gainsharing Demonstration*. 2016, The Agency for Healthcare Research and Quality.

45. Kozhimannil, K.B., et al., *Cesarean delivery rates and costs of childbirth in a state medicaid program after implementation of a blended payment policy.* Medical Care, 2018. **56**(8): p. 658-664.

46. Minnesota Health Services and Medical Management Division, *Perinatal Practices*, 2010. Retrieved from https://www.leg.state.mn.us/docs/2010/mandated/100219.pdf.

47. Calvin, S. and T. Balazovic, *A Single Bundled Payment for Comprehensive Low-Risk Maternity and Newborn Care Provided by Independent Midwife-Led Birth Center Practices that Are Clinically Integrated with Physician and Hospital Services.* 2017, Minnesota Birth Center.

48. Struijs, J.N., De Vries, E.F., De Bruin-Kooistra, M., Baan, C.A. *Payment reform in maternity care. What can we learn from initiatives abroad?* (in Dutch). 2017, Bilthoven: The Dutch National Institute for Public Health and the Environment (RIVM). Retrieved from https://www.rivm.nl/sites/default/files/2018-11/009961_105550_IVM_001_A4_Brochure_V2_TG.pdf

49. Struijs JN, De Bruin-Kooistra M, Heijink R, Baan CA. *The transition towards bundled payments for maternity care* (in Dutch). 2016, Bilthoven: The Dutch National Institute of Public Health and the Environment (RIVM). Retrieved from https://www.rivm.nl/publicaties/op-weg-naar-integrale-bekostiging-van-geboortezorg

50. Struijs, J.N., De Vries, E.F., Van Dorst, H.D.C.A., Over, E.A.B., Baan, C.A. *Insight in outcomes, utilization and medical spending of maternity care and first experiences with bundled payments* (in Dutch). 2018, Bilthoven: The Dutch National Institute for Public Health and the Environment (RIVM). Retrieved from https://www.rivm.nl/publicaties/geboortezorg-in-beeld-nulmeting-en-eerste-ervaringen-met-werken-met-integrale

51. Henderson, J., *Maternity Pathway Bundled Payment. Maternity Pathway Bundled Payment*. 2016, Department of Health. Retrieved from http://www.oecd.org/health/better-ways-to-pay-for-health-care-9789264258211-en.htm.

52. Berry, S.A., et al., *ProvenCare perinatal: a model for delivering evidence/ guideline-based care for perinatal populations.* Jt Comm J Qual Patient Saf, 2011. **37**(5): p. 229-39.

53. Grigg, C.P., et al., *Evaluating Maternity Units: A prospective cohort study of freestanding midwife-led primary maternity units in New Zealand - Clinical outcomes.* BMJ Open, 2017. **7**(8).

54. Bailey, D.J., *Birth outcomes for women using free-standing birth centers in South Auckland, New Zealand.* Birth (Berkeley, Calif.), 2017. **44**(3): p. 246-251.

55. Clark, *Primary Maternity Services Notice Amendment Notice 2018*, In *Gazette New Zealand*. 2018. Retrieved from https://www.health.govt.nz/system/files/documents/publications/primary-maternity-services-amendment-notice-2018.pdf

56. Berenson, R., et al., *Refining the framework for payment reform.* In *Health Affairs Blog*, 2016.

ADDITIONAL FILES

A1. Search strategy

Embase session results (comparable strategy voor Scopus and Pubmed)

105

#16

#15 AND [1-12-2016]/sd

125

#15

#14 AND [2016-2019]/py

797

#14

#13 NOT (**'asia'**/exp OR **'africa'**/exp OR **'south and central america'**/exp)

937

#13

#12 AND (**english**:la OR **dutch**:la)

982

#12

#10 NOT **#11**

5,901,894

#11

letter:it OR **note**:it OR **erratum**:it OR **news**:it OR **'conference abstract'**:it OR **'conference paper'**:it OR **'conference review'**:it

1,326

#10

#3 AND (**#4** OR **#5** OR **#6** OR **#7** OR **#8** OR **#9**)

29,268

#9

'new zealand'/mj OR **'new zealand'**:ti

64,271

#8

'prospective payment'/mj OR **'bundled payment'**/mj OR **'reimbursement'**/mj OR **'healthcare financing'**/mj OR **'financing, organized'**/mj OR **'financing, government'**/mj OR (**'integrated health care system'**/de AND **pay***:ti,ab)

156

#7

'individual funding':ti,ab OR **'personal funding'**:ti,ab OR **'individual budget*'**:ti,ab OR **'personal budget*'**:ti,ab

427
#6
'value based reimbursement':ti,ab OR **'funding model'**:ti,ab OR **'integrated funding model*'**:ti,ab OR **'case rate payment*'**:ti,ab OR **'individualised funding'**:ti,ab OR **'individualized funding'**:ti,ab OR **'personalised funding'**:ti,ab OR **'personalized funding'**:ti,ab OR **'individualised budget*'**:ti,ab OR **'individualized budget*'**:ti,ab OR **'personalised budget*'**:ti,ab OR **'personalized budget*'**:ti,ab
673
#5
((**bundle*** NEAR/5 **pay***):ti) OR ((**episode*** NEAR/5 **pay***):ti,ab)
45,105
#4
((**payment*** NEAR/3 **model***):ti,ab) OR ((**alternative** NEAR/3 **payment***):ti,ab) OR **'funding source*'**:ti,ab OR **'funding strateg*'**:ti,ab OR **'funding system*'**:ti,ab OR **'service funding'**:ti,ab OR **'financing source*'**:ti,ab OR **'financing strateg*'**:ti,ab OR **'financing system'**:ti,ab OR ((**pay** NEAR/3 **expenses**):ti,ab) OR **'bear cost*'**:ti,ab OR **'payment system*'**:ti,ab OR **reimbursement**:ti,ab OR **'payment reform*'**:ti,ab OR **'payment policy change*'**:ti OR **payment***:ti
308,705
#3
#1 OR **#2**
303,021
#2
'maternity ward'/de OR **'obstetric delivery'**/de OR **'home delivery'**/de OR **'obstetrics'**/de OR **'obstetrical nursing'**/de OR **'nurse midwifery'**/de OR **'midwife'**/de OR **'newborn nursing'**/de OR **'neonatal nurse practitioner'**/de OR **'perinatal care'**/de OR **'perinatal nursing'**/de OR **'perinatal mortality'**/de OR **'perinatal outcome'**/de OR **'postnatal care'**/de OR **'newborn care'**/de OR **'newborn assessment'**/de OR **'newborn death'**/de OR **'newborn intensive care'**/de OR **'newborn intensive care nursing'**/de OR **'newborn intensive care unit'**/de OR **'newborn monitoring'**/de OR **'maternal care'**/de OR **'maternal health service'**/de OR **'prenatal care'**/de OR **'prenatal diagnosis'**/de OR **'prepregnancy care'**/de OR **'pregnancy outcome'**/de
25,175
#1
'birthing centers':ti OR **'birthing centres'**:ti OR **'obstetric care'**:ti OR **'obstetric services'**:ti OR **'obstetric health services'**:ti OR **midwifery**:ti OR **midwives**:ti OR **'birth attendants'**:ti OR **deliveries**:ti OR **'prenatal care'**:ti OR **'antenatal care'**:ti OR **'natal care'**:ti OR **'birth care'**:ti OR **'neonatal care'**:ti OR **'perinatal care'**:ti OR

'perinatal health':ti OR **'perinatal services'**:ti OR **'perinatal population*'**:ti OR **'perinatal outcome*'**:ti OR **'maternal care'**:ti OR **'maternal health services'**:ti OR **'maternal services'**:ti OR **'maternity care'**:ti OR **'pregnancy care'**:ti OR **'pregnancy services'**:ti OR **'pre-pregnancy care'**:ti OR **'prepregnancy care'**:ti OR **'pregnancy related services'**:ti OR **homebirth***:ti

A2. Detailed characteristics of the 17 initiatives employing APMs in maternity care

No.	Initiative (abbreviation)	General characteristics			Type of APM							Risk mitigation strategy	Ref
		Country	Year	Status	Main type	Care providers participating in the payment model	Account-able entity	Eligible popu-lation	Episode time span	Care activities covered by payment model	Linkage to quality		
1	Commissioning for Quality and Innovation, Payment Framework (CQUIN)	United Kingdom	2007	Permanent implementation	Pay for performance	Single organisation	Hospital	Pregnancies ending in elective or emergency Caesarean section.	Delivery	All perinatal hospital care.	Yes: proportion of NHS providers income linked to achievement of specified thresholds on specified self-defined quality indicators	Unclear	28
2	Texas Medicaid Program (Texas)	US (Texas, and later also Georgia, Michigan, New Mexico, New York and South Carolina)	2011	Developed into initiative no. 3	Pay for performance	Single organisation	Hospital	Penalty for neonatal delivery before 37 gestation that was not medically necessary	Delivery	n.a.	n.a.	Unclear	22, 29
3	Horizon Blue Cross Blue Shield of New Jersey, Pregnancy Episodes of care program (Horizon)	US (New Jersey)	2013	Permanent implementation	Shared savings, upside only. Benchmark based on historical data.	Multiple organisations representing different levels of care	Provider	Low-risk pregnancies and high-risk pregnancies from 2015.	Till 30 days after birth	All prenatal outpatient care, all delivery-related care (vaginal/cs), postpartum care not known.	n.a.	Exclusion of pregnancies with comorbidities such as HIV and cancer. Exclusion of neonatal care.	29

No.	Initiative (abbreviation)	General characteristics			Type of APM							Risk mitigation strategy	Ref
		Country	Year	Status	Main type	Care providers participating in the payment model	Accountable entity	Eligible population	Episode time span	Care activities covered by payment model	Linkage to quality		
4	Baby+ Company	US (North Carolina, Tennessee, Colorado, Arkansas)	2013	Permanent implementation	Shared savings (one-sided). Separate rates if transferred before/during labor.	Single health care organisation	Free standing birth center.	Low-risk pregnancies.	From initial obstetrics consultation at birth centre to 6 weeks postpartum	Mother and newborn. Prenatal care including birthing plan, classes, postpartum care, newborn exam, metabolic screen and medications. Includes facility and professional fees.	Yes, but method unclear	Exclusion of members not enrolled during entire pregnancy, high-risk pregancies and complicated deliveries. Exclusions of lab testing and ultrasounds.	19
5	TennCare	US (Tennessee)	2013	Permanent implementation	Shared savings (two-sided): spli 50/50	Multiple organisations representing different levels of care	Principle Accountable Provider is the healthcare provider who delivers the neonate	Low- to medium-risk pregnancies with live births	40 weeks before delivery to 60 days postpartum	All inpatient and outpatient medical services as well as ancillary services such as pharmacy and imaging services. All related prenatal care and medication or emergency department visits. All care during delivery. Postpartum care till 30 days: non-inpatient readmissions, non-admission ED care, other covered medication, practitioner and facility costs. Postpartum care till 60 days: all postpartum-related medical care and medications.	Yes: distribution of savings contingent on quality improvements achieved	Exclusion of episode based on clinical and cost-based exclusion criteria. Total averages for benchmark for shared savings/ losses are risk-adjusted based on comorbidities, demographics and other indicators of patient variation. Exclusions, preconception and neonatal care.	10, 30-34

No.	Initiative (abbreviation)	General characteristics			Type of APM							Risk mitigation strategy	Ref
		Country	Year	Status	Main type	Care providers participating in the payment model	Accountable entity	Eligible population	Episode time span	Care activities covered by payment model	Linkage to quality		
6	Arkansas Health Care Payment Improvement Initiative (Arkansas)	US (Arkansas)	2013	Permanent implementation	Shared savings (two-sided). PAPs may be rewarded, penalized, or remain financially neutral based on how the average costs for their episodes compare with thresholds predetermined by payers.	Multiple organisations representing different levels of care	Principle Accountable Provider is the provider or provider group that performs the delivery.	Low- to medium-risk pregnancies with live births	40 weeks before delivery to 60 days postpartum	All inpatient and outpatient medical services as well as ancillary services such as pharmacy and imaging services. All prenatal care, care related to labor and delivery, postpartum maternal care.	Yes: distribution of savings contingent on quality improvements achieved	Exclusion of members not enrolled during the entire pregnancy period and of episodes based on clinical and cost-based exclusion criteria. Total averages for benchmark for shared savings/losses are risk-adjusted based on comorbidities, demographics and other indicators of patient variation. Cost saving payments are capped beyond a limit determined by each payer. Exclusions: costs not related to maternity care, neonatal care and preconception care.	10, 11, 19, 23, 35-37

No.	Initiative (abbreviation)	General characteristics			Type of APM							Risk mitigation strategy	Ref
		Country	Year	Status	Main type	Care providers participating in the payment model	Accountable entity	Eligible population	Episode time span	Care activities covered by payment model	Linkage to quality		
7	Ohio Episode-Based Payment Model (Ohio)	US (Ohio)	2015	Permanent implementation	Shared savings (two-sided). Historically determined threshold.	Multiple organisations representing different levels of care	Principle Accountable Provider, which is the healthcare provider who delivers the neonate	Low- to medium-risk pregnancies with live births,	280 days before delivery to 60 days postpartum	Mother only. All inpatient and outpatient medical services as well as ancillary services such as pharmacy and imaging services, including office visits, supportive services such as psychosocial evaluation, counseling, eduction (e.g. breat feeding), vaginal delivery or caesarean sections, induction, epidural, additional screening, neonatal support in the delivery period, transportation, follow-up visits for the mother, readmission, postpartum depression, and genetic screening.	Year 1: Yes: distribution of savings contingent on quality improvements achieved and the reporting of data on specified indicators, with threshold designed to make 75% of providers eligible for bonuses. From year 2: Yes: distribution of savings contingent on quality improvements achieved and the reporting of data on specified indicators, with threshold designed to make the highest quartile of providers eligible for bonuses	Exclusion of members not enrolled for the entire pregnancy or episodes based on clinical and cost-based exclusion criteria. Risk adjustment for the calculation of the shared savings /losses. Exclusions: prenatal medications, neonatal care and preconception care.	10, 19, 33, 38

No.	Initiative (abbreviation)	General characteristics			Type of APM							Risk mitigation strategy	Ref
		Country	Year	Status	Main type	Care providers participating in the payment model	Accountable entity	Eligible population	Episode time span	Care activities covered by payment model	Linkage to quality		
8	Community Health Choice, Bundled Payment Pilot (CHC)	US (Texas)	2015	Pilot	Year 1: Shared savings (one-sided), split 50/50. Year 2: shared savings (two-sided), split (greater for gains for providers, smaller for providers for losses). Benchmark based on historical average costs adjusted with yearly growth trends. In three components; pregnancy, delivery and newborn.	Multiple organisations representing different levels of care	Obstetricians from multi-specialty provider groups.	Low-risk and high-risk pregnancies.	270 days before delivery to 60 days postpartum	Newborn. Three components: pregnancy, delivery and newborn care. All prenatal care and services related to delivery (with blended rate for vaginal and Caesarean delivery). Postpartum: maternal and neonatal care (levels 1 to 3) till 30 days postpartum.	Year 1: Yes: Providers were to report accountability data, though outcomes had no financial consequences. Year 2: Yes: distribution of savings contingent on quality improvements achieved	Risk adjustment based on risk factors (age, comorbidities, clinical severity markers). Individual stop loss provision. Exclusions: level 4 neonatal intensive care.	19, 39, 40

No.	Initiative (abbreviation)	General characteristics			Type of APM							Risk mitigation strategy	Ref
		Country	Year	Status	Main type	Care providers participating in the payment model	Accountable entity	Eligible population	Episode time span	Care activities covered by payment model	Linkage to quality		
9	New York State's Medicaid Maternity Care Value Based Payment Arrangement (New York States)	US (New York state)	2016	Pilot	Year 1: shared savings (one-sided). Year 2: shared savings (two-sided).	Multiple organisations representing different levels of care	Value Based Payment contractor (usually a hospital and/or professionals involved in maternity care)	All-risk pregnancies	270 days before delivery till 60 days postpartum	Newborn. Prenatal services including visits, lab tests, medication, ultrasounds. Services associated with delivery whether vaginal or caesarean, up to 60 days postpartum for the mother. Facility costs, professional services and associated complications for mother and child are included. Newborn care up to 30 days postpartum.	Year 1: Yes: distribution of savings contingent on achievement of specified quality improvements. Year 2: Yes: distribution of savings contingent on achievement of specified quality improvements, with penalties partially or fully written off if quality improvements achieved	Exclusions of mothers aged <12 or > 64 at the time of the delivery, maternal death, stillborn and multiple live births, medicaid members for whom medicaid is not the sole payer, members eligible for inclusion in another program (e.g. hiv/aids or intellectually or developmentally disabled). Risk adjustment and stop loss provision for deciding on the shared savings. Bundles in three episodes: pregnancy, delivery till 60 days post discharge and neonatal care)	41-43

No.	Initiative (abbreviation)	General characteristics			Type of APM							Risk mitigation strategy	Ref
		Country	Year	Status	Main type	Care providers participating in the payment model	Accountable entity	Eligible population	Episode time span	Care activities covered by payment model	Linkage to quality		
10	Pacific Business Group on Health, Blended Case Rate (Pacific)	US (Southern California)	2014	Pilot	Bundled payment light. Fee for service (one blended fee per pregnant woman for all hospital deliveries irrespective of type).	Single health care organisation	Hospital is accountable for the facility blended rate. Medical group practice is accountable for the professional blended rate.	All risk hospital deliveries.	Hospital labour and delivery period only	All facility and professional activities during labor and delivery for both vaginal and caesarean section births.	n.a.	Exclusions of pregancies that left against medical advice, transferred during labor, various comorbidities (e.g. hiv/aids, cancer, also gestational age <37 weeks, multi gestation 3+).No prospective risk adjustment.	19, 35, 44-46
11	Minnesota Blended Payment (Minnesota BP)	US (Minnesota)	2009	Permanent implementation	Bundled payment light. Fee for service (one blended fee per pregnant woman for all hospital deliveries irrespective of type).	hospital	Hospital	Uncomplicated births	Delivery	professional services and facility fees for vaginal or casarean delivery and prenatal, postnatal care	no	Exclusion of complicated vaginal deliveries	24, 47

No.	Initiative (abbreviation)	General characteristics			Type of APM							Risk mitigation strategy	Ref
		Country	Year	Status	Main type	Care providers participating in the payment model	Accountable entity	Eligible population	Episode time span	Care activities covered by payment model	Linkage to quality		
12	Minnesota Birth Centers, BirthBundle (BirthBundle)	US (Minnesota)	2015	Pilot stopped	Bundled payment (retrospective). Benchmark based on historical data. If all care is within the birth center, facility fees and professional fees are included in the bundle. No birth center coordination fee in event of hospital delivery. Professional fees only are included if delivered in a hospital.	Single health care organisation	Minnesota Birth Center	Low-risk pregnancies / medium / high -risk	270 days before delivery to 56 days postpartum	Mother and newborn and facility fees. Prenatal care including ultrasounds, lab testing. Perinatal including doula. Postpartum neonatal assessment within 24-hours postpartum and consultation at 1-2 weeks and 6 weeks and lactation support.	n.a.	Facility fees are FFS outside of bundel.	19, 48

No.	Initiative (abbreviation)	General characteristics			Type of APM							Risk mitigation strategy	Ref
		Country	Year	Status	Main type	Care providers participating in the payment model	Accountable entity	Eligible population	Episode time span	Care activities covered by payment model	Linkage to quality		
13	Bundled payment for maternity care (The Dutch BP)	The Netherlands	2017	Pilot	Bundled payment (prospective)	Multiple organisations representing different levels of care	Integrated care organization	All risk pregnancies.	From pregnancy identification to 6 weeks post-partum.	Bundle defined in four phases (prenatal, natal, postnatal and kraamzorg), covering all necessary care - according to the national care standardization guidelines.	no (although possible to define links in contract)	Exclusions of bundle breakers (i.e. when pregnant woman decides to see care professionals outside of the integrated care network. Depending on the contract, risk corridors on where new negotiations are started - are negotiated. Each phase (except for kraamzorg) destinguishes between regular and complex cases.	12, 49-51

No.	Initiative (abbreviation)	General characteristics			Type of APM							Risk mitigation strategy	Ref
		Country	Year	Status	Main type	Care providers participating in the payment model	Accountable entity	Eligible population	Episode time span	Care activities covered by payment model	Linkage to quality		
14	Maternity Pathway Bundled Payment (Maternity Pathway BP)	England	2013	Permanent implementation	Bundled payment (prospective)	Multiple organisations representing different levels of care	Single lead provider per phase.	Low-, intermediate- and high-risk pregnancies	From 10th week of pregnancy to 6 weeks postpartum	All maternal and neonatal care in the prenatal, perinatal and postpartum phase.	n.a.	Exclusions, stratifying risks in several bundle prices. Supplementary payments for specific complications. in three components: prenatal, perinatal and postpartum. Prenatal and postpartum further stratified into standard, intermediate and intensive categories. Perinatal further stratified into with complications and co-morbidities and without complications and comorbidities. There are supplementary payments for specific complications. Exclusions: health problemes in neonates	52

No.	Initiative (abbreviation)	General characteristics			Type of APM							Risk mitigation strategy	Ref
		Country	Year	Status	Main type	Care providers participating in the payment model	Accountable entity	Eligible population	Episode time span	Care activities covered by payment model	Linkage to quality		
15	Providence Health and Services, Pregnancy Care Package (Providence)	US (Oregon)	2013	Permanent implementation	Bundled payment (prospective)	Single health care organisation	Nurse or midwife	Low-risk pregnancies	From pregnancy identification to 6 weeks postpartum	Mother and newborn. All prenatal care, including check-ups, prenatal screening, education, psychosocial support. All delivery-related care and all postpartum care for mother and neonate. Also includes mobile app (Doula) and patient navigators.	n.a.		19
16	Geisinger Health System, Perinatal ProvenCare Initiative (GHS)	US (Pennsylvania)	2007	Permanent implementation	Bundled payment (prospective)	Multiple organisations representing different levels of care	GHS provider	Low-risk pregnancies.	From pregnancy identification to a consultation 21–56 days postpartum	Mother only. Prenatal care including professional and outpatient services. Postpartum care including inpatient readmissions, outpatient and professional care.	No link, although quality metrics are incorporated, measured and tracked.	Exclusions: late referrals, high risk pregnancies, members not enrolled during the entire pregnancy (i.e. at least 12 continuous weeks of prenatal care and delivery must be performed by a GHS provider). Exclusions: neonatal care, care delivered by non-GHS providers.	19, 21, 35

No.	Initiative (abbreviation)	General characteristics			Type of APM							Risk mitigation strategy	Ref
		Country	Year	Status	Main type	Care providers participating in the payment model	Accountable entity	Eligible population	Episode time span	Care activities covered by payment model	Linkage to quality		
17	Lead Maternity Care Model (LMC)	New Zealand	2007	Permanent implementation	Bundled payment (prospective)	Single provider if the lead maternity caregiver was the actual practitioner, otherwise multiple providers	LMC midwife, GP or obstetrician	All risk pregnancies	From pregnancy identification to 6 weeks postpartum	Mother only. Including complications during postpartum period, including episode-related maternal readmissions,postpartum minimum of 5 home visits till 6 weeks postpartum. Exclusions: neonatal care, ultrasounds, consuklting obstetricians and pediatricians	n.a.	Exclusions of pregnancies that went elsewhere for further care. And exclusions of neonatal care. (fixed budget) in four modules: first and second trimester, third trimester, labour and birth and postnatal care.	53-55

A3. Full-text document types

Author, year	Document type	Initiative names
Miller et al., 2013	Interviews	LMC
Jarlenski et al., 2016	Commentary	Tennessee, Arkansas and Ohio
Kozhimannil et al., 2018	Evaluation	Minnesota Blended Payment
Carroll et al 2018	Evaluation	Arkansas
Butcher 2018	Blog	Horizon
Dahlen et al 2017	Evaluation	Texas Medicaid Program
HCPLAN	Review	Ohio, Minnesota birth bundle, baby company, AABC (proposal), Pacific, GHS, Providence, CHC, Arkansas, Tennessee
Calvin & Balazovic	White paper	Minnesota Birth Bundle
HCPLAN (experiences from ohio and tennessee)	Review - case report	Ohio and Tennessee
Fish 2017	Powerpoint Presentation	New York DSRIP
RIVM Factsheet	Factsheet	Dutch maternity care bundle
Smith and Hanlon 2017	Case Study	Tennessee
Berry et al., 2011	Observational study	GHS
Henderson 2016	White paper	UK
Ertok, 2015	Evaluation	CQUIN
Lally 2013	Issue brief	PBGH, Arkansas and GHS
New York State 2017	Factsheet	New York DSRIP
Ohio episode based payment update 2014	Powerpoint Presentation	Ohio

Author, year	Document type	Initiative names
Sutherland 2014	Commentary	Ohio
Chernew et al	Interviews	Arkansas
Arkansas Health Care Payment Improvement Initiative. Perinatal episode of care 2017	Program overview	Arkansas
State of Ohio. Detailed business requirements: perinatal episode. 2017	Perinatal Episode description	Ohio
DeRoche et al 2015	Commentary	Tennessee
Grigg and Tracy 2013	Description of NZ's maternity care system	LMC
Tennessee Detailed Business Requirement	Perinatal Episode description	Tennessee
De Brantes and Love 2016 NEJM Catelyst	Catalyst article	CMC Texas
Rubinstein 2015	Case study	Pacific business group,
HCIII Arkansas Episode Design Summary	Perinatal Episode description	Arkansas
De Brantes and Love Case Study		Texas Medicaid
Castellucci et al 2018	Blog	Humana
NZ Ministry 2007	Legal document	LMC
New York State 2016	White Paper	New York DSRIP
Perinatal Practices 2010	Legislative paper	Minnesota Blended Payment
New Zealand Gazette 2018	Legislative paper	LMC
Healthcare Payment Improvement Initiative	White paper	Arkansas
Integrated Healthcare Association 2016	Episode description	Pacific Business Group

A4. Quality appraisal of studies performing effect evaluations of APMs in maternity care

Reference	Initiative (abbreviation)	Payment model	Score[a]						
			Selection Bias	Study Design	Confounders	Blinding	Data Collection	Withdrawals and Dropouts	Global Rating
Berry et al 2011 [21]	Geisinger Health System, Perinatal ProvenCare Initiative (GHS)	bundled payment prospective	1	2	3	1	3	n.a.	3
Carroll et al 2018 [23]	Arkansas Health Care Payment Improvement Initiative (Arkansas)	shared savings two sided	1	2	1	1	3	n.a.	2
Kozhimannil 2018 [24]	Minnesota Blended Payment	FFS blended case rate	1	2	1	1	3	n.a.	2
Dahlen et al 2017 [22]	Texas Medicaid Program	FFS with penalties	1	2	1	1	3	n.a.	2

a Scores of 1, 2, and 3 correspond to strong, moderate, and weak quality, respectively, using the Effective Public Health Practice Project Quality Assessment Tool for Quantitative Studies 22 .

5


6

CHAPTER 6

Barriers to payment reform: experiences from nine Dutch population health management sites

Published as
De Vries EF, Drewes HW, Struijs JN, Heijink R, Baan CA
Barriers to payment reform: experiences from nine Dutch Population Health Management sites
Health Policy. 2019: 123(11)

ABSTRACT

Population health management (PHM) initiatives aim for better population health, quality of care and reduction of expenditure growth by integrating and optimizing services across domains. Theoretically, a precondition for PHM is to shift from traditional fee-for-service towards value-based payment models. We aimed to gain insight into payment reform in nine Dutch PHM sites. Specifically, we investigated 1) the type of payment models implemented, and 2) the experienced barriers towards payment reform. Between October 2016 and February 2017, we conducted 36 (semi-) structured interviews with program managers, hospitals, insurers and primary care representatives of the sites. We addressed the structure of payment models and barriers to payment reform in general. After three years of PHM, we found that four shared savings models for pharmaceutical care and five extensions of existing (bundled) payment models adding providers into the model were implemented. Interviewees stated that reluctance to shift financial accountability to providers was partly due to information asymmetry, a lack of trust and conflicting incentives between providers and insurers, but all the same to a lack of a sense of urgency. Small steps to payment reform have been taken in the Dutch PHM sites, which is in line with other international PHM initiatives. While acknowledging the autonomy of PHM sites, governmental stewardship (e.g. long-term vision, supporting knowledge development) can further stimulate value-based payment reforms.

INTRODUCTION

Increasingly, population health management (PHM) initiatives are being implemented across Western countries [1] striving for better population health, better quality of care and a reduction of the cost growth (i.e. Triple Aim). The essence of PHM is to integrate and optimize services across prevention, care, cure and social care for a pre-specified population within the region [2]. In regional networks of healthcare providers, insurers, local governments (municipalities) and other health organizations, several interventions, such as setting up integrated elderly care, substituting low-complex medical activities to a primary care setting and others, for improved collaboration and coordination of services, are implemented in PHM.

Theoretically, a key precondition for successful PHM is to shift away from the traditional fee-for-service (FFS) payment models to more value-based payment models [3, 4]. FFS models are known to incentivize each provider to increase the amount of services produced (as long as price is above marginal cost). As FFS models are designed for acute care specifically, they do not automatically align with the Triple Aim [5]. Payment models such as bundled payments or global payments are more aligned with the Triple Aim. They increase financial accountability for (groups of) providers and, in that manner, incentivize better coordination of care and support the integration of services over domains within the bundle or budget. At the same time, they reduce incentives for overtreatment and low-value care [6].

Nevertheless, there is uncertainty regarding how to successfully develop and implement value-based payment reform. Currently, several PHM initiatives are experimenting with payment reform [7], yet studies on their impact are few. Well-known examples are Gesundes Kinzigtal in Germany [8] and the Alternative Quality Contract (AQC) in the U.S. [9]. Gesundes Kinzigtal implemented a long-term shared savings contract [8], while in the AQC providers are being paid a risk-adjusted global budget [10]. Also in the Netherlands, PHM initiatives have been initiated [11, 12], which currently explore more value-based payment schemes as part of their efforts in moving toward the Triple Aim [12]. Few empirical studies evaluating payment reform have been conducted, showing some promising effects. For example, two studies [8, 13] showed beneficial outcomes of Gesundes Kinzigtal; savings and a lower mortality rate compared to the control group. In addition, AQC enrollees had lower spending growth and generally greater quality improvements after four years, as compared with similar populations in other states [10]. The authors conclude that payment reform in AQC's, i.e. global budget contracts with

quality incentives, contributed to these results, even though other factors also played a role [10].

Moreover, studies on the barriers and enablers to payment reform are scarce and are generally more theory-focused (e.g. [14, 15]). The few studies which have discussed experiences from the field highlighted that there is a need for a neutral convening party that maintains the commitment of providers and payers, a need of flexible, stable leadership, pressure from public and private payers, and an increased investment to support infrastructure, care management and human resources [16]. These findings were solely based on U.S. experiences. Since the organization of health systems differs, PHM initiatives in other countries may face different challenges. Therefore, it is important to broaden the scope of the literature including initiatives in other healthcare systems. The Netherlands has a system of regulated competition and includes mandatory health insurance and the expectation of competition between private insurers [17]. We aimed to gain insight into the process to payment reform in the Dutch PHM sites. More specifically, we aimed to 1) provide insight into the type of payment models that were introduced, and to 2) provide an overview of barriers during the implementation of payment reform based on the experiences of relevant stakeholders. This study is part of the National Monitor Pioneer sites (NMP) which monitors the progress of nine Dutch PHM sites from 2013 until 2018 [18].

MATERIALS AND METHODS

Study setting

Dutch PHM sites are regionally oriented network organizations varying in population size, participating organizations and interventions. Table 1 shows the general characteristics of the Dutch PHM sites. The population size ranged between 42,000 and 646,200 residents. Participating organizations included insurers, primary care providers, hospitals, municipalities, citizen representative organizations and employers [19]. All networks aim to integrate and optimize services across domains by implementing two types of interventions. First, PHM sites implement interventions to improve the organization of care through e.g. better data infrastructure. Second, PHM sites implement interventions to improve the delivery of care through substitution and integration of care and better self-management. Examples are substituting brand with generic medication, substituting secondary care with primary care, improving existent chronic care bundles, integrating care for elderly and prevention activities [19].

Table 1: Population health management site characteristics

Name	Population size	Participating organizations	Interventions
A	170.000	Citizen representative organization, GP, hospital, insurer, pharmacy	Self-management, integrated care (mental health), substitution (pharmacy, hospital care to primary care)
B	646.200	Citizen representatives organization, municipality, health promotion, home care facilities, GP, hospital, insurer	Integrated care (elderly, birth care), substitution (hospital care to primary care), data-infrastructure
C	106.500	Citizen representatives organization, municipality, health promotion, health promotion, home care facilities, youth care, mental care, GP, long-term care, hospital, pharmacy	Prevention, self-management, substitution (pharmacy)
D	270.000	Citizen representatives organization, municipality, health promotion, health promotion, home care facilities, GP, long-term care, hospital, pharmacy	Prevention, self-management, integrated care (elderly, diabetes), substitution (pharmacy, hospital care to primary care)
E	55.000	Citizen representative organization, GP, hospital, insurer, pharmacy	Integrated care (elderly, diabetes, COPD, CVRM), substitution (pharmacy), data-infrastructure
F	42.000	Citizen representative organization, GP, insurer	Integrated care (elderly), substitution (hospital care to primary care)
G	270.600	Citizen representative organization, mental care, GP, hospital, insurer, pharmacy	Integrated care (elderly, birth care, mental health, diabetes, COPD, CVRM), substitution (pharmacy, hospital care to primary care)
H	300.000	Citizen representative organization, GP, hospital, insurer, pharmacy	Self-management, integrated care (diabetes, COPD, CVRM), substitution (pharmacy, hospital care to primary care)
I	113.000	Citizen representatives organization, municipality, health promotion, long term care, home care facilities, GP, hospital, insurer, pharmacy	Prevention, self-management, integrated care (elderly), substitution (pharmacy)

GP: General Practicioner; COPD: Chronic Obstructive Pulmonary Disease; CVRM: Cardiac Vascular Risk Management

The PHM sites operate within the context of the Dutch health system where providers are predominantly being paid by FFS (for each visit or for each service) or diagnosis-related group (DRG). GPs are being paid through capitation augmented with FFS per visit and/or service. Basic mental health care, physical therapy, postpartum care and home care is being paid for through FFS. Additionally, services such as lab-testing and medication are being paid for through FFS. Hospitals are being paid through a type of DRGs which are episode-based bundles defined by the combination of diagnosis and treatment. As from 2007, bundled payments for integrated chronic care (diabetes, chronic obstructive pulmonary disease (COPD) and cardiac vascular risk management (CVRM)) have been introduced in the primary care setting [20]. Bundled payments for birth care, including primary and secondary care organizations, were introduced as from January 2017. These bundled payments are not mandatory; providers and health insurers can still choose to use the traditional FFS model. The decision to implement these bundled payments is left to the PHM sites. Furthermore, within the Dutch system there is room to experiment temporarily with new, locally defined, payment models. Also, providers and health insurers in PHM sites can make contractual agreements on e.g. shared savings or pay-for-performance.

Definitions – payment reform

We defined payment reform as PHM interventions that 1) made changes to the type of payment (i.e. moving away from traditional FFS) and/or 2) made changes in the number of participating providers in existent payment models.

Combining elements of previous work of the Health Care Payment Learning Action Network [21], Stokes, Struckmann [22] and Struijs JN [23], we distinguished between several types of payment models and the level of integration of the payment model. We distinguished between three base payment models: FFS, bundled payments and global payments. Under FFS providers are paid a fee for each services delivered and therefore incentivizes increasing the volume of care [24]. On top of the FFS structure, pay-for-performance, shared savings or retrospective bundles may be added to incorporate additional incentives. In pay-for-performance, on top of FFS, a bonus or a malus is paid for attaining some quality threshold. The additional payments can be employed for improving coordination, efficiency of care delivery or quality of care [25]. In a shared savings model, individual providers are each paid on a FFS bases and then there is reconciliation between the target episode price and the actual episode average price that is attributed to a provider (HCPLAN 2016), which is shared with the payer. Shared saving models that share in gains only are called one-sided. In two-sided models, also risks are shared. In our definition, shared savings models differ from bundled payments

as in bundled payments bonuses are not shared with the payer. In retrospective bundled payment, there is a virtual budget negotiated upfront, providers are being paid by FFS and retrospectively, the target prices is reconciled fully [23].

Prospective bundled payment is not based on the FFS base structure, as it pays a prospectively defined price that is paid as one payment with no reconciliation [21, 23].

In global payments, the entire population and the entire continuum of care is included. The accountable providers is paid a fixed fee per head of the population. Moving away from FFS (i.e. adding additional payment components (pay-for-performance or shared savings) or implementing another base payment structure (prospective bundled payment or global budget)) indicates increased financial provider risk.

The level of integration refers to the type and the scope of providers that are involved in the payment model. We distinguished between payments that are horizontally integrated (e.g. only covers providers within primary care) and payments that are vertically integrated (e.g. includes multiple types of providers – primary care and secondary care or over multiple domains, e.g. care and cure) in the payment model [22]. The latter refers to a higher level of integration.

Study design and sample

This qualitative study consisted of two parts. In the first part (October 2016), we conducted exploratory interviews with program managers (n=9) in order to identify and gain insight into the type of payment reforms within the PHM sites.

In the second part (November 2016 through February 2017), 27 relevant stakeholders concerning the identified payment reforms were interviewed to explore the experiences with payment reform in general (i.e. the interviews were not restricted to the payment reforms that we identified during the first part).

We used purposive sampling [26] in order to ensure that experiences of key stakeholders within the PHM regions were sufficiently reflected within the sample. Therefore, per site, we interviewed professionals who were involved in payment reform from varying sectors and organisations (program manager (8), health insurer representative (9) from six different insurer companies, care group or primary care organization representative (8), municipality civil servant (1) and hospital board member (1). The professionals were recruited through the network of the NMP. Whenever participants showed interest, they were emailed to schedule an appointment.

Data collection

In the first part, the interviews (by phone) were based on a topic list (see Appendix A) that included, among others, a description of the PHM interventions, funding and payment model. The answers were written down in a fixed format containing all interview topics. All program managers checked, adapted if needed and approved the content of the forms.

In the second part, participants were asked to sign a consent form. A semi-structured interview guide was used during the interviews (of +/- 1 hour) to identify different barriers to payment reform (see Appendix B). For the development of the guide, the authors build on issues raised in previous research (refs rapport). Interviews took place face-to-face (n=21) at the professionals work space or in a restaurant or by telephone (n=6).

Data analysis

In the first part, we selected payment reforms (see definition in conceptual framework) that had to be in effect before, or at the time of, the interviews. Following, we described the targeted condition, the participating providers and the type of payment model that was implemented. For definitions, see 'conceptual framework'.

In the second part, the 27 interviews were audio recorded, transcribed verbatim and approved by the interviewee. The interviews were coded inductively using MAXqda software. Based on the first two interviews, EdV coded the interviews inductively per interview guide topic. Next, the coding scheme was discussed with HD and adapted to focus on themes that were brought up by the interviewees. Following, the remaining interviews were coded by EdV and the coding was checked by HD. Finally, drafts of the summary results were discussed with all authors and approved by all authors.

RESULTS

Type of payment reform

Table 2 describes the nine PHM payment reforms by their type of payment model, the aim of interventions, the targeted conditions, the level of integration (i.e. the included providers) and the payment flow. We identified FFS (n=3), prospective bundled payment (n=2) and one-sided shared saving models (n=4). Generally, the aim of the payment reforms was to substitute secondary with primary care or substitute brand with generic medication. Targeted patients were chronic conditions such as diabetes, CVRM or

Table 2: Payment reforms in the nine Dutch PHM sites

Intervention no.	Payment model	Intervention aim	Targeted conditions	Level of integration - Involved providers	Payment flow
1	FFS*	To substitute secondary care to primary care when possible.	Musceloskeletal conditions	Vertically: GPs using the consultation of medical specialists in a primary care setting	The medical specialist has a freelance contract for 1 day a week in the GP-practice and is paid by FFS for a maximum of two consultations using GP-funds.
2	FFS*	To substitute non-complex care from hospital to primary care.	Non-acute patients.	Vertically: GPs using the consultation of medical specialists (internists, orthopedists, neurologists, dermatologists, cardiologists) in a primary care setting	The joint venture (GPs and hospital) gets paid by FFS of which GPs and the medical specialists receive a fixed ratio.
3	FFS*	To substitute and integrate mental health care.	Patients with non-complex, non-acute mental or social problems	Horizontally: GP, GP nurse and primary care psychologists in a primary care setting through an umbrella organization where GP-practices (with specialized mental health nurses) and primary care psychologists are part of.	The umbrella organization gets paid by FFS, for the mental health nurse and the psychologists separately. The GP and the psychologists, in turn, are getting paid per amount of time spend with the patient (which may include a number of visits).

Table 2: Payment reforms in the nine Dutch PHM sites (continued)

Intervention no.	Payment model	Intervention aim	Targeted conditions	Level of integration - Involved providers	Payment flow
4	Bundled payment (prospective)	To substitute non-complex care from hospital to primary care.	Diabetes, asthma and COPD	Vertically: GP using the consultation of medical specialists (by phone, or sent the patient in).	. The medical specialists are paid by GP-funds. The out-of-pocket payments for patients using secondary care are not addressed.
5	Bundled payment (prospective)	To substitute non-complex care from hospital to primary care.	Diabetes, COPD and CVRM	Vertically: GP using the consultation of medical specialists	The medical specialists are paid by GP-funds.
6	Shared savings (one sided)	To substitute brand medication with generic medication.	Diabetes and CVRM patients (cholesterol medication, ATII-blockers). Also patients using proton pump inhibitors.	Vertically: GP, medical specialists and pharmacists	The existing payment structures were not adjusted.
7	Shared savings (one sided)	To substitute brand medication with generic medication.	Diabetes and CVRM patients (cholesterol medication, ATII-blockers). Also patients using proton pump inhibitors.	Vertically: GP, medical specialists and pharmacists	The existing payment structures were not adjusted. Savings will be used to invest in new projects.
8	Shared savings (one sided)	To substitute brand medication with generic medication.	Diabetes and CVRM (cholesterol medication)	Vertically: GP, medical specialists and pharmacists	The existing payment structures were not adjusted. Savings are shared between insurer and the PHM site.

Table 2: Payment reforms in the nine Dutch PHM sites (continued)

Intervention no.	Payment model	Intervention aim	Targeted conditions	Level of integration - Involved providers	Payment flow
9	Shared savings (one sided)	To substitute brand medication with generic medication.	Diabetes and CVRM patients (cholesterol medication and ATII-blockers).	Vertically: GP, medical specialists and pharmacists	The existing formal payment structures were not adjusted. Part of the savings will be invested in new projects.

GP: General Practicioner; FFS: fee-for-service; COPD: Chronic Obstructive Pulmonary Disease; CVRM: Cardio Vasculair Risk Management; AT: Angiotensine; PHM: Population Health Management; *In FFS the reform consisted only of including more providers in the payment model.

mental health problems. One payment reform targeted musculoskeletal patients. The level of integration was mostly (n=8) vertical, as the payment reform general practitioners and medical specialists (n=4) and in four reforms also pharmacists were included. One reform was only horizontally integrated, it included mental health professionals, all in primary care.

Experienced barriers to payment reform

Table 3 presents a summary of the experiences of the participants (program managers, health insurer representatives, care group or primary care organization representatives, municipality civil servants and hospital board members) with the barriers to implement payment reform in general. Multiple barriers were mentioned per site and per participant. The following section discusses the experiences by theme.

Table 3: Experienced barriers to payment reform in nine Dutch PHM sites

	Experienced barrier
1	Information asymmetry.
2	Worsening reputation of insurers.
3	Lack of trust as a result of failed reform attempts.
4	Misaligned incentives in the hospital setting.
5	Hesitation to accept financial accountability.
6	Lack of start-up funding.
7	Lack of leadership and intrinsic motivation.

Information asymmetry

Health insurer, program manager and primary care organization participants mentioned *information asymmetry* as barrier towards payment reform. Insurers and program managers indicated information asymmetry at the favor of providers. Accordingly, some insurers stated that they need the transparency to be improved in future negotiations, as they are of the opinion that the prices are too high. They indicated that shifting accountability to providers would increase the information asymmetry at the benefit of providers.

"We need to remove information asymmetries. And make agreements with each other. It is not going to work without proper agreements. Providers need to realize that; they have freedom to organize care between provider and patient. But that doesn't mean that they don't have to justify themselves. So, as you get more freedom to act, your accountability also increases. That is, exactly, well, that is exactly what it's about in the pioneer site, in this contract. You get a

standard contract. And you are allowed to deviate, but then you have to accept that you increase your accountability. And that means that you need to provide data more often." (program manager)

At the same time, participants from primary care organizations expressed the need to limit the information asymmetry at the favor of insurers. As such, a couple of physician provider organizations said to use their knowledge on the prices in the negotiations with insurers, trying to incorporate some playing room for overhead and efficiency improvements by revealing the true costs at a certain point in time.

All interviewees agree on that the lack of data infrastructure and privacy issues are important barriers to reduce the information asymmetry. In one pioneer site, the respondents state that they are experimenting with innovative data sharing mechanisms, to provide real time data to insurers and providers as part of the contract. According to a program manager from another site, another solution would be to create an independent party who collects and analyses patient data to get around privacy issues.

Worsening reputation of insurers

Insurer interviewees also described the *worsening reputation of insurers* as a barrier towards payment reform. Some insurers stated that payment reform would be perceived as a 'cost cutting' measure and therefore will worsen their reputation.

"Yes, absolutely. That is one of the issues. […] The reputation that we have, that is another issue. We say: "It is not up to us to perfom a certain surgery 90 times, as opposed to 80 times, thats the minimum. Who are we to say?" Well, then you can picture the headlines in the papers already." (insurer)

Instead, insurers say they offer long-term contracts with the possibility to slowly decrease hospital production using a virtual budget, which aims to let professionals take the lead:

"It says, the medical personnel, well, most often it are doctors.. We have formulated it more broadly. Well, they actually initiate interventions. When we start them, they think it is not good." (insurer)

Lack of trust as a result of failed reform attempts

Another theme was a *lack of trust as a result of failed reform attempts*, that was mentioned by participating insurers, primary care providers and hospitals after failed attempts to implement shared savings programs.

"Because, at first, the idea was that the savings would go to the providers, as an individual bonus. That is something I was against, from the start. I always say, we are not going to refer patients over their heads. Or, realizing certain things, because it will come back at us, in such way, that we gain from it financially. So, we want to profit from it, but then to invest collectively in this new innovation […]. Well, the deep fear of insurers, they thought they couldn't get it from the hospitals. So, that they have to pay us savings, because we showed that we had savings, but they, they would lose even more." (physician organization)

Participating insurers, primary and secondary care providers disagreed on how to spend the potential savings, which negatively impacted levels of trust between stakeholders.

Misaligned incentives in the hospital setting

All interviewees mentioned *misaligned incentives*, especially in the *hospital setting*. These misaligned incentives were experienced between insurer and hospital boards and between hospital boards and medical specialists:

"[…] Payment systems need to be based on value, and not volume. […] But we have volume caps negotiated with the insurer. […] So it is vital for me to maximize revenues within the cap. […] The problem is that I have to work on World Peace, but I'm financed like I'm in truce. When I go out there, I know I'm in war with my specialists, because they will say: 'You are not taking good care of me'." (hospital board member)

All respondents stated that the volume-based payments do not stimulate to reduce production in hospitals in order to substitute care towards the primary care setting, as medical specialists and medical specialist parterships are incentivized by the FFS-structure to provide more services. The respondents agreed that volume caps between insurer and hospital increases the interest of hospital board members to maximize production within the cap, because future caps are generally determined by historical production.

Additionally, an insurer stated that the FFS incentive at the level of the medical specialists and the parternships in the hospital is barely limited by the cap. Insurer respondents state that they try to create comfort for hospitals by slowly reducing production using long-term contracts with virtual budgets.

"And then you have to think about it in this way, that, well, that they at the moment their volume declines, they implement flanking policies. That means that, at least for the fixed costs, that they do not disappear in it one by one. Because, at the moment you do not see any patients

anymore, you cannot claim the tariffs, so there is less revenue. […] While the fixed costs do not decline. […] So, well. So we try to make arrangements in order to reduce it more slowly." (insurer)

An insurer added that this is also influenced by the pressure of the cost growth reduction of the budgetary framework from the Dutch government that exerts similar incentives as the cap.

The combined threat of a reduction of the hospital budgets by the volume caps and the budgetary framework from the Dutch government hinders hospital boards to accept payment reform, as hospital boards employing medical specialist partnerships already experience difficulties to act on the volume caps in the current FFS structure.

Hesitation to accept greater financial accountability

Some providers stated to feel *hesitant to accept greater financial accountability*. Primary care representatives expressed worries that they are too small and have too little financial reserves to carry these risks. Additionally, a primary care representative worries that they are becoming accountable for outcomes they cannot control.

"[…] but, then you have to select those indicators of which you know you can influence as a provider, that is directly linked to the achievements providers can make, and then you could say, well, if you meet this and that threshold, you can get a bonus, this could work. The only thing is, that in these types of things, they say, we don't have any influence on what patients do. So it's very hard to select these things, I guess." (primary care representative)

At the same time, primary care representatives stated that they fear that budget from secondary care will not follow the substitution of care sufficiently towards primary care; while they do feel that more time or budget is necessary for the additional tasks they perform. This was confirmed by an insurer who stated that he feels reluctant to promise structural additional revenues to primary care, as long as secondary care payments have not reduced.

Lack of start-up funding

Most respondents of provider representatives stated that insurers should invest more in payment reform, while bearing in mind that experimenting with payment reform requires risk capital. Some insurers stated that they are reluctant to invest in those programs, since the financial benefits are often generated after several years only.

"[..] if they talk about shared savings, thats fine, but in 2017, we get 1/3 of it.. They make their business case that the benefits will flow to them first. But we say, you have to reimburse for the costs first! And when there are actually savings, ofcourse, we will share them. Otherwise we are waiting for 5, 6, 7 years on our money. So you see, to define what it is and understand each other's point of views and under which conditions, it is quite new ground." (insurer)

The same insurers said that if they would invest in payment reform, they would first want their investments back, before they pay out savings to the participants.

Lack of leadership and intrinsic motivation

In general, primary care representatives and some insurer representatives expect the government to stimulate payment reform and to provide with funding and knowledge to facilitate the reform. Conversely, (three) other insurer representatives and one primary care representative feel that risk appetite and providers' intrinsic motivation need to be stimulated directly, by giving providers more freedom to reorganize care delivery. Other insurer representatives feel that the intrinsic motivation to reform should stem from higher levels of trust between stakeholders, leadership and through quick wins in (e.g.) pharmacy-related shared savings programs, which were implemented as a reaction to the failed shared savings programs:

"Even more, if you can show that you can do something and it results in a small fund.. I have seen what the project, how it got people enthusiastic. Anyway, in two sites, people worked together for the first time, because of that project!" (Insurer)

Shifting a part of the financial accountability to providers, by tying quality performances to bonuses, is not perceived as a solution by both insurers and providers. They fear that focusing on a defined set of quality indicators will destroy the intrinsic motivation of professionals to deliver high quality care, and that it may result in avoiding complex patients.

DISCUSSION

In this study, we aimed to provide insight into the process towards payment reform in nine Dutch PHM sites by investigating to what extent payment reforms were implemented and the experienced barriers to payment reform. This study showed that payment reforms currently include expanding the number of providers in existing fee-for-service or bundled payment models and implementing pharmacy-related shared

savings programs in pharmaceutical care in the cure sector. Bundled payments or global payments including providers over the domains (cure, prevention or social care) have not (yet) been launched. The shared savings programs in pharmaceutical care mainly aimed to create trust to encourage motivation to reform. Previous shared savings attempts, that tried to include multiple providers, failed and negatively affected mutual trust. The interviews revealed reluctance to shift financial accountability to providers from both an insurer as well as a provider perspective. This was mainly due to a lack of trust and information asymmetry at the favour of providers. Furthermore, conflicting incentives at organization level and physician or specialist partnership level, a worsening reputation of insurers and a lack of funding and leadership hindered further reform. Above all, a lack of a sense of urgency hinders the payment reform in the Dutch PHM sites.

Small steps are taken in the progression towards payment reform in the Dutch PHM sites. This is in line with other PHM initiatives such as Accountable Health Communities in the U.S. [27] and the Vanguards in the UK, as they similarly focus on care redesign instead of payment reform. Examples of more disruptive payment reforms are few. The AQC in the U.S. and Gesundes Kinzigtal in Germany implemented global payments or shared saving programs including multiple providers, but those reforms are limited to the cure sector. Therefore, it seems in place to acknowledge that payment reform takes time, especially in PHM that aims to connect efforts over the domains of cure, care and prevention, as it is challenging to change within the complex and fragmented health system.

Although at the time of this study, the Dutch PHM sites were just three years in effect; it is interesting to explain the experienced barriers towards payment reform, because it helps to understand the problems PHM initiatives all over the world are currently facing. Our study corresponds with findings from earlier U.S. studies by showing that there is a need of stable leadership and increased funding to support infrastructure [16, 28]. Our findings extend to the existing literature by showing that there is a vacuum where various stakeholders expect others to take action. Especially, insurers and large providers are being accused of a lack of a sense of urgency to pursue the reform. This is illustrated by insurers' fear for an increased information asymmetry that will lead to an increased power imbalance between insurers and large provider organizations at the favour of providers (in accordance with Schut and Varkevisser [29]). The mechanisms at play are complicated and relate to factors such as leadership, alignment of goals and incentives, shared norms and values as well as the relations between the actors [30].

This might – in part, be explained by differences in the US and Dutch health systems. In the US, the Affordable Care Act (ACA) explicitly dictated to establish the Medicare Shared Savings Program that encourages various regional stakeholders to form ACOs and several types of bundled payment models [31]. In that way, the US government exerts pressure to payment reform of the public health system, to start the care delivery reform specifically. In the Netherlands, the role of the government is less clear. And even though it is the responsibility of the Dutch government to regulate the preconditions within the system of regulated competition with private payers [32], it is the question to what extent the government should interfere to increase the sense of urgency to payment reform. Inherently, PHM initiatives require decentralization to a certain extent. Specifically in the Dutch context, payment reform is seen as a part of the move towards PHM, instead of being the start of the reform. Yet, our study showed that some interviewees expect the government to provide more guidance or assistance on payment reform by creating a platform for knowledge. This platform could provide a long-term vision with information on the implementation and potential impact of different types of payment reform. Such governmental stewardship might resolve a part of the hesitance in moving towards value-based payment models. An example where the Dutch government exerted more pressure is in bundled payments for birth care. Here, the Ministry of Health strongly recommended the (voluntary) uptake of bundled payments [33] and provided subsidies to develop knowledge and tools and made available specific payment (infra)structures. It seems to have created a sense of urgency for implementing payment reform. However, as PHM sites need autonomy to operate, finding the optimal balance between top-down efforts from the government and bottom-up efforts from providers, insurers and sites seems essential for successful PHM [34]. Therefore, strong leaders who are aware of the need for change, with experiences in the health care market and who are able to create a new impulse in the insurer-provider relationship are required to align goals and incentives.

Aligning incentives with the Triple Aim seems to require a shift towards more value-based payment models. However, as internationally implemented payment reforms are scarce and do not (yet) stretch further than the cure sector, it is unknown how to align incentives for multi-stakeholder initiatives that services several patient groups over the domains of care. Payment models for chronic care demand other incentives than payment models for screening activities. Therefore, successful PHM probably requires a combination of value-based payment models adjusted to the complex and dynamic PHM setting.

This study has several limitations. First, not all Dutch payment reforms were included as they were not part of the PHM pioneer sites. For instance, the previously mentioned bundled payments for birth care were not included. Second, as only one researcher coded the interviews, the results may have been influenced by the researchers' subjective interpretation. To minimize the bias, another researcher checked the coding work and the summary results were discussed until consensus was reached with all authors. Third, as we inductively gathered experienced barriers, we might have missed insights from the PHM sites that were not monitored in this study, as for example the view of municipalities. Nevertheless, based on efforts to study other PHM initiatives [11] we are confident that we have shown a fair representation of the situation in Dutch PHM sites. We suggest future research to follow efforts to payment reform closely, to qualitatively investigate what works and what does not, to investigate potential solutions for barriers encountered and to quantitatively support those finding on the Triple Aim goals.

Conclusions

During the first three years of the Dutch PHM sites, payment reforms included paying for consultation of medical specialists in a primary care setting through traditional fee-for-service models, adding secondary care in existent bundled payment models for chronic care and shared savings programs in pharmaceutical care**.** Bundled payments or global budgets including providers over the domains (cure and prevention or social care) have not (yet) been launched as PHM intervention. PHM representatives stated that reluctance to shift financial accountability to providers was partly due to information asymmetry, a lack of trust between providers and insurers and conflicting incentives, but all the same to a lack of sense of urgency. Small steps to payment reform have been taken in the Dutch PHM sites, which is in line with other international PHM initiatives. While acknowledging the autonomy of PHM sites, governmental stewardship (i.e. long-term vision and supporting knowledge development) can further stimulate value-based payment reform.

REFERENCES

1. Organisation for Economic Co-operation and Development (OECD), *Fiscal Sustainability of Health Systems*. 2015, OECD Publishing: Paris.
2. Steenkamer, B.M., et al., *Defining Population Health Management: A Scoping Review of the Literature.* Popul Health Manag, 2017. **20**(1): p. 74-85.
3. Miller, H.D., *From volume to value: better ways to pay for health care.* Health Aff (Millwood), 2009. **28**(5): p. 1418-1428.
4. Nolte, E., McKee, M., *Caring for people with chronic conditions: a health systems perspective*, in *European Observatory on Health Systems and Policies Series*, 2008.
5. Tsiachristas, A., *Financial Incentives to Stimulate Integration of Care.* Int J Integr Care, 2016. **16**(4): p. 8.
6. Mechanic, R.E. and Altman S.H., *Payment reform options: episode payment is a good place to start.* Health Aff (Millwood), 2009. **28**(2): p. w262-71.
7. Hendrikx, R.J., et al., *Which Triple Aim related measures are being used to evaluate population management initiatives? An international comparative analysis.* Health Policy, 2016. **120**(5): p. 471-85.
8. Pimperl, A., et al., *Evaluating the Impact of an Accountable Care Organization on Population Health: The Quasi-Experimental Design of the German Gesundes Kinzigtal.* Popul Health Manag, 2017. **20**(3): p. 239-248.
9. Song, Z., et al., *The 'Alternative Quality Contract,' based on a global budget, lowered medical spending and improved quality.* Health Aff (Millwood), 2012. **31**(8): p. 1885-94.
10. Song, Z., et al., *Changes in health care spending and quality 4 years into global payment.* N Engl J Med, 2014. **371**(18): p. 1704-14.
11. Lemmens, L.C., et al., *Een populatiegerichte aanpak voor verbinding van preventie, zorg en welzijn: de beweging in beeld.* Ned Tijdschr Geneeskd, 2017(161: D849).
12. Drewes, H.W., Struijs, J.N., Baan, C.A., *How the Netherlands Is Integrating Health and Community Services.* In *NEJM Catalyst*, 2016.
13. Busse, R. and Stahl, J., *Integrated care experiences and outcomes in Germany, the Netherlands, and England.* Health Aff (Millwood), 2014. **33**(9): p. 1549-58.
14. Conrad, D.A., *The Theory of Value-Based Payment Incentives and Their Application to Health Care.* Health Serv Res, 2015. **50 Suppl 2**: p. 2057-89.
15. Christensen, T.J., *A framework for guiding efforts to reward value instead of volume.* Int J Health Econ Manag, 2016. **16**(2): p. 175-187.
16. Conrad, D.A., et al., *Emerging lessons from regional and state innovation in value-based payment reform: balancing collaboration and disruptive innovation.* Milbank Q, 2014. **92**(3): p. 568-623.
17. Van de Ven, W.P. and Schut, F.T., *Universal mandatory health insurance in the Netherlands: a model for the United States?* Health Aff (Millwood), 2008. **27**(3): p. 771-81.
18. Drewes, H.W., et al., *National Monitor Populationmanagament. Part 1: first description of the pioneer sites.* (in Dutch) 2014, National Institute for Public Health and the Environment (RIVM): Bilthoven.
19. Drewes, H.W., Heijink, R., Struijs JN, Baan CA, *Working together towards sustainable care. National Monitor Population Management* (in Dutch). 2015, National Institute for Public Health and the Environment (RIVM): Bilthoven.
20. Struijs, J.N. and Baan C.A., *Integrating care through bundled payments--lessons from The Netherlands.* N Engl J Med, 2011. **364**(11): p. 990-1.
21. Alternative Payment Model Framework and Progress Tracking (APMFPT) Working Group, *Alternative Payment Model (APM) Framework*. 2016. Retrieved from https://hcp-lan.org/workproducts/apm-whitepaper.pdf.
22. Stokes, J., et al., *Towards incentivising integration: A typology of payments for integrated care.* Health Policy, 2018. **122**(9): p. 963-969.
23. Struijs, J.N., Hayen, A., van der Swaluw, K., *When Designing Bundled Payments, Don't Ignore The Lessons Of Behavioral Economics*, in *Health Affairs Blog*. 2018.

24. Frakt, A.B. and Mayes, R., *Beyond capitation: how new payment experiments seek to find the 'sweet spot' in amount of risk providers and payers bear.* Health Aff (Millwood), 2012. **31**(9): p. 1951-8.

25. Organisation for Economic Co-operation and Development (OECD), *Better Ways to Pay for Health Care*. 2016, OECD Publishing: Paris.

26. Mays, N. and Pope, C., *Rigour and qualitative research.* Bmj, 1995. **311**(6997): p. 109-12.

27. Alley, D.E., et al., *Accountable Health Communities--Addressing Social Needs through Medicare and Medicaid.* N Engl J Med, 2016. **374**(1): p. 8-11.

28. Hildebrandt, H., et al., *Gesundes Kinzigtal Integrated Care: improving population health by a shared health gain approach and a shared savings contract*. Int J Integr Care, 2010. **10**: p. e046.

29. Schut, F.T. and Varkevisser, M., *Do healthcare insurers have too much power?* (in Dutch) Ned Tijdschr Geneeskd, 2016. **160**(0): p. D1133.

30. Drewes, H.W., et al., *Regions moving towards sustainable health systems: National Monitor Populationmanagement – Reflection on five years of pioneering* (in Dutch). 2018, National Institute of Public Health and the Environment (RIVM): Bilthoven.

31. Skinner, J. and A. Chandra, *The Past and Future of the Affordable Care Act*. Jama, 2016. **316**(5): p. 497-9.

32. Van de Ven, W.P., et al., *Preconditions for efficiency and affordability in competitive healthcare markets: are they fulfilled in Belgium, Germany, Israel, the Netherlands and Switzerland?* Health Policy, 2013. **109**(3): p. 226-45.

33. Struijs JN, De Bruin-Kooistra M, Heijink R, Baan CA. *The transition towards bundled payments for maternity care* (in Dutch). 2016, Bilthoven: The Dutch National Institute of Public Health and the Environment (RIVM). Retrieved from https://www.rivm.nl/publicaties/op-weg-naar-integrale-bekostiging-van-geboortezorg

34. Bengoa, R., *Transforming health care: an approach to system-wide implementation.* Int J Integr Care, 2013. **13**: p. e039.

ADDITIONAL FILES

A1. Interview topic list

1. Contract
   a. Level of contract (at the PHM site level and insurer?)
   b. Scope (financing sources (overhead and care related, included population)
   c. Preconditions (monitoring or IT?)
   d. Design of the model and incentives (tied to quality of care?)
   e. Inclusion criteria patients
   f. Duration
2. Intervention description (based on TIDIER Hoffman)
   a. Name
   b. Rationale
   c. Target population
   d. Involved organizations
   e. Involved care professionals
   f. Financing/contract
   g. Materials
   h. Procedures
   i. Status
   j. Tailoring
   k. Adjustments
   l. Evaluation
   m. Measuring

A2. Semi-structured topic list

- What is necessary for realizing the Triple Aim?
- What are your expectations regarding payment reform?
- What are your previous experiences with payment reform?
- What are your thoughts of payment reform in relation to:
  - o Market position
  - o Role of government
  - o Other stakeholders
  - o Competition
  - o Funding
  - o Structure of the PHM site

6


7

CHAPTER 7

General discussion

Background and thesis aims

The increasingly pressured sustainability of health systems challenges Western countries to improve the value of their health systems. Value can be increased by improving population health, reducing spending levels or both. One of the main responses across countries is to improve health care delivery by integration of services across medical care, social care and public health within regions, which is called population health management (PHM). This thesis aimed to contribute to the existing literature on PHM by:

A. exploring how to measure the concept of value for monitoring and budget allocation decision making within and across PHM regions.
B. gaining insight into experiences with, and the types and effects of, alternative payment models (APM) in order to assess its possible role in PHM.

The remainder of this chapter presents the main findings of this thesis, reflects on the main findings and discusses implications and further steps.

Main findings

Part A: Measuring value in population health management

Part A explored how to measure the concept of value in a PHM setting; by reviewing the current state of low-value care indicators (chapter 2) and by exploring how to operationalize the concept of value within observational datasets (chapter 3 and chapter 4).

Chapter 2 showed that the majority of low-value care indicators are in medical (primary or secondary) care (87 out of the 115 listed low-value care indicators). The remaining indicators were found in prevention (n=25) and in long-term care (n=3). No indicators were found in social care. Three indicators were assigned the highest level of evidence as they were underpinned by both guidelines and evidence from the literature. Other indicators were underpinned by clinical guidelines or Choosing Wisely recommendations [1]. Despite the fact that several indicators are used in provider-payer contracts, no information on the validity of the indicators was found in the literature.

Chapter 3 analyzed the drivers of regional variation in medical spending by looking at subgroups (i.e. individuals with diabetes and depression) in addition to the total population. Heterogeneity issues with regard to case-mix were aimed to overcome by using an extensive dataset (secondary health survey data linked with claims data, health care supply data and municipality registration data), in addition to the selection

of subgroups. The results showed that PHM regions with above (or below) average spending for the general population mostly showed above (or below) average spending for diabetes and depression as well. Yet, the drivers of the variation at the regional level varied between subgroups. Demand factors explained most variation across regions in medical spending for depression and explained 88% of the variation for diabetes. The variation left unexplained (12% for diabetes) indicates differences across regions due to inefficiencies. This suggests that the extent to which regional variation in medical spending can be considered as inefficiency may differ between regions and subgroups.

In **chapter 4,** we described the variation of six proposed value-indicators for maternity care across Dutch Maternity Care Networks (MCNs). Specifically, we used the association between low-value services utilization and maternal and neonatal health outcomes as measures of value across the MCNs. We found substantial variation across MCNs for the six value-indicators. The additional analyses suggested that the inputs (i.e. low-value care indicators) have captured a part of value. We could not rule out that these findings were due to differences in case-mix, despite the use of many case-mix variables.

Part B: gaining insight into alternative payment models

Part B gained insight into the current state of alternative payment models (APMs) using international literature (chapter 5) and experiences of PHM stakeholders in the Netherlands (chapter 6).

Chapter 5 reviewed what types of APMs have been implemented in maternity care in Western countries. Seventeen initiatives employed APMs in the United States (n=13), the United Kingdom (n=2), New Zealand (n=1) and in the Netherlands (n=1). Within these initiatives, pay-for-performance models (n=2), shared savings models (n=7) and bundled payment models (n=8) were found. Key design elements (such as eligible population, episode time span, care providers that participated in the model, care activities covered by the model, risk mitigation strategies) varied highly. Key terms describing the type of payment model (e.g. shared savings and bundled payments) were used interchangeably. APMs that shifted more financial accountability toward providers tended to include more strategies that mitigated financial risks. The first evaluations (n=4) on the effects of APMs are tentatively positive on different indicators of health and spending. Two studies found a positive association with health outcomes and two studies found a positive association with a reduction of medical spending.

Chapter 6 investigated what types of payment models were implemented in the Dutch PHM regions and what barriers were experienced towards payment reform**.** After three

years of PHM**,** shared savings models for pharmaceutical care (n=4) and extensions of existing (bundled) payment models adding providers into the model (n=5) were adopted. The experienced barriers included information asymmetry between providers and insurers, worsening reputation of insurers, lack of trust as a result of failed reform attempts, misaligned incentives in the hospital settings, hesitation to accept financial accountability and lack of start-up funding, leadership and intrinsic motivation to reform. According to the interviewees, these experienced barriers were partly due to a lack of a sense of urgency.

Reflections on the main findings

Part A: Measuring value in population health management

Although the concept of value originated in a disease-specific context, it gained popularity in integrated services settings, such as PHM. The concept of value, which is defined as health outcomes that matter to the patient relative to spending, is important in PHM as it is linked with the interdependent PHM goals (i.e. increasing population health outcomes and quality of care while slowing the cost growth). Following Struijs' framework [2] (see Chapter 1, Figure 2) it is essential to continuously monitor progress on these goals in order to (re)allocate resources in such a way that the needs of the population are met. The following section discusses the several conceptual and methodological issues the we encountered in our exploration to measure value.

Gaps in availability of indicators of value and data across domains

This thesis has shown that there are currently gaps in the availability of indicators of value in two ways. First, low-value care indicators are currently found in medical care mainly, and domains such as prevention, long-term care and social care are underrepresented (chapter 2). This is not surprising, considering the origin of value and Value-Based Health Care (VBHC) initiatives that are particularly popular in, and designed for, medical care [3]. Yet, previous research [4-6] has indicated (e.g. through measuring practice variations) that low-value care is also present in other sectors along the care continuum. Second, within medical care, most health outcome indicators are observational outcomes (e.g. Apgar scores, admission rates, mortality), and currently information on the experiences of patients is lacking. Patient Reported Outcome Measures (PROMs) and Patient Reported Experience Measures (PREMs) are important for measuring value in PHM, because they reflect outcomes and experiences that matter to the patient. Therefore, such indicators comprise an essential part of the value concept [3]. The International Consortium of Health Outcome Measurement (ICHOM) continually works on developing PROMs and PREMs.

The spending part of value is endogenous

Operationalizing the spending part of the value-equation can be problematic due to endogeneity. The problem is that spending encompasses both the treatment and the complications that may be caused by the treatment itself. Moreover, health care spending may reflect the health status of persons rather than the contribution of health services to health outcomes, even when adjusted for case-mix (chapter 4). Because we had access to an extensive set of case-mix variables from multiple data-sources, the question arises whether the presence of even more case-mix variables in the data would help to limit bias. Although this problem is acknowledged by multiple authors (e.g. [7, 8]), spending is still being used in studies regarding efficiency and value [8, 9]. Chapter 4 aimed to avoid endogeneity by replacing medical spending with low-value care indicators, but this strategy also has methodological and conceptual drawbacks. For instance, using a single low-value care indicator reduces the sensitivity of value dramatically, and therefore, is not suitable for assessing value across services or even across multiple domains (like in PHM). Another drawback is that, based on available data, it is sometimes hard to disentangle whether a service is of high-value or of low-value [10]. For example, a caesarean section is of high-value when fetal and/or maternal distress occurs. When there is no medical need, a caesarean section is of low-value. Instead, the caesarean section may even impose harms, because such invasive medical procedures are not risk free and may cause, for example, serious infections [11]. Furthermore, using solely low-value care indicators instead of spending implicitly filters out the contribution of high-value services, which is equally relevant in determining value [12]. For instance, effective identification of what type of care is necessary for whom and continuity of care professional are services of high-value in maternity care that should be acknowledged in the conceptualization of value.

Considering these issues, it is essential to align the operationalization of value to the research question and the level of measurement. As such, despite the endogeneity problems, it may be more feasible to use spending in the value-equation if the question is to monitor progress towards value over time in a country. In this example, the endogeneity problems may have less harmful consequences, than when value is measured across regions to determine the quality of care for consumer information or tying payment to. In addition, depending on whether the goal is to improve value by reducing spending (growth) and maintaining health outcomes or improve value by maintaining spending (growth) and improving health outcomes, it makes sense to include either low-value or high-value indicators.

Can value be influenced at the regional level?

Conceptually, value can be captured at each decision-making level [13], which includes the system-level, regional level, provider organization level, and of course, the level of the communication between care provider and patient. This thesis implicitly assumed that regional level policies affect value by influencing provider behavior [14]. More specifically, the assumption was made that PHM may lead to improved value by increased collaboration and integration among providers. This assumption was based on examples from the literature where regional policies affected provider behavior, such as implementing APMs [15], making budget allocation decisions [16, 17], adopting integrated care delivery models [18], standardizing clinical practices by improving guideline adherence [19, 20], or adopting new practices as a consequence from contact with (new) colleagues [21]. This thesis showed (small) variation between subgroups within and across regions, that persisted after correction for case-mix. Therefore, we cannot rule out the possibility of differences at the regional (and subgroup) level due to inefficiencies [13, 22]. This is in line with our assumption and other studies showing small regional differences [13, 23, 24]. At the same time, we acknowledge that variation can occur at other levels, for example at the provider (organization) level (e.g. [25]) or at the system level (e.g. [23]), which are additionally important to study.

Part B: Gaining insight into alternative payment models

A frequently discussed theme in the quest of PHM initiatives to improve value is alternative payment models (APMs). Shifting away from the traditional FFS models to more value-based payment models is perceived by many to be a precondition for a successful implementation of PHM [26-28]. In practice, many PHM initiatives struggle on how to design and implement APMs, and there are doubts on whether APMs will turn out to be valuable. Below, we discuss some topics, which can be helpful in the discussion regarding to the role of APMs within PHM initiatives.

The complexity of alternative payment models

Although APMs generally search for the optimal allocation of risks between payer and provider [29], there is not a *one size fits all*. The optimal allocation of risks will differ, depending on the specific aims and settings of the APMs and care settings. Chapter 5 reviewed the use of APMs in maternity care and found that currently multiple types of APMs (pay-for-performance, shared saving models, bundled payment models) are adopted in practice. Moreover, key elements, such as the included population, included care professionals, time span, and the link with quality of care and risk mitigation strategies varied widely across the initiatives. The high variety of APM elements may indicate that initiatives actively try to adjust the APM to the contextual factors of the

health system at hand, such as scale (i.e. small practice, practice groups, hospitals, etc.) [30] and the level of risk that providers are able and willing to bear. As such, when relatively more financial accountability is shifted, more risk mitigation strategies are applied (chapter 5). This indicates that there is a lack of trust, probably of providers, in techniques to allocate insurance risk only with the insurer, and provider risk only with the provider [31, 32] (for more information on insurer risk and provider risk, see Chapter 1). In turn, providers become reluctant to adopt (increased) financial accountability (chapter 6) and have a desire for more comfortable adoption of APMs. Moreover, there is a variety of complex terms and ambiguous definitions that lead to key terms are being used interchangeably, and therefore, confuses understanding of stakeholders. For example, the finding that 'shared savings' and 'bundled payments' are used interchangeably (chapter 5) indicate that there is no common language for the APM typology, which further complicates the discussions about the adoption of APMs in practice.

Will alternative payment models help to improve value?

There is much debate on what to expect from APMs. Therefore, this thesis aimed to contribute to the discussion by reviewing the literature about the effects of APMs in maternity care on the two pillars of value; spending and health outcomes. Chapter 5 found that, for maternity care, only four evaluations in peer-reviewed journals were performed, which were tentatively positive. This is in accordance with literature evaluating the effects of APMs for other episodes of care, which results also seem tentative. Most studies find evidence for APMs to reduce spending [33, 34], but there are also studies suggesting that spending is being shifted outside the scope of the APM [35]. Commonly, studies do not find much evidence for APMs to improve health outcomes [33, 34]. The lack of conclusive evidence in the literature may be due to that it may take longer for initiatives to make changes visible. For example, Song et al. [36-38] showed increasing effects of their global payment model after 1, 4 and 8 years. Another reason for the lack of empirical evidence is the difficulty to isolate the impact of APMs from other (policy) changes at the same time, which is a common problem in the evaluation of policy changes. Moreover, it seems sensible to select indicators for an empirical evaluation based on what the APM is expected to contribute to what aspect in the health system. However, the complexity of the (terminology of the) APMs and the high variety of key elements complicates the selection of the indicators that measure the effect of the APM. Chapter 5 summarized evaluations of APMs in maternity care and found that evaluations used *different* indicators of health and spending. This underlines one of the main problems (which also concerns the above discussion in *measuring the concept of value in a PHM context*) as formulated by Glied [39]: we don't really know what we mean by value, and therefore, we don't really know how to pay for it [39] and how to measure it.

When value is measured differently across initiatives, it is not surprising that conflicting conclusions are drawn from different studies. This is problematic for both the evidence base of APMs, as the incentives that are intended to follow from the APMs.

Besides all that, it may be that initial expectations about the effects of APMs were too high or not correct. The articles that formulated that one of the solutions to keep health systems sustainable is the introduction of alternative payment models [40, 41] may have contributed to those expectations. In reality, the effects of APMs on value may be more nuanced. APMs can contribute to increased value by providing incentives to efficient, high-value care, which should be understood in the context of many other factors that play a role in the reform towards more integrated care [42], which could contribute to a better value health system.

Future research and policy implications

Aforementioned issues should be considered in order to operationalize value in a PHM context for monitoring and budget allocation purposes, and to design and adopt APMs that are aligned with the PHM goals. The following section discusses future research and policy implications.

Develop indicators of value along the care continuum at each relevant level

As the need to improve the sustainability of health systems in Western countries will continuously increase, the urge to measure the progression towards value will also increase. This thesis showed that there are currently gaps in the availability of indicators to address the full concept of value. At the same time, PHM initiatives such as the Accountable Health Communities (AHC) and the Dutch PHM regions develop interventions that include multiple domains along the care continuum. Therefore, it is important to further develop value-indicators, which are aligned with the goals and the activities of PHM initiatives. In the last couple of years, low-value care indicators gained increasing attention, but there should also be attention for the development of high-value indicators. Moreover, health indicators that reflect outcomes and experiences that *matter* should be developed in addition to the high- and low-value indicators. Future research should focus on the development of a valid set of value-indicators that reflect value along the entire care continuum according to the population, and for use at the point of care, for each individual specifically. Comprehensive sets of value-indicators should be developed that can be used for monitoring purposes, budget allocation decisions or direct provider feedback. The sets of value-indicators should not be too narrow and should not be too wide. Too narrow sets of indicators may lead to, for example in provider-payment contracts, in a simple cost-containment strategy or even adverse health effects. Too broad

or too complex sets of indicators may lead to complex interpretation of the 'scores' for the user [43]. Instead, a comprehensive small number of meaningful indicators with a solid evidence base and validity could incentivize value better [43]. Hereby, transparency is essential, as using indicators of low quality might lead to (depending on the goal) negative consequences, such as underuse of indicated services, patient- and provider dissatisfaction, adverse health outcomes or patient selection [10].

Carefully design alternative payment models to each specific situation

For APMs to be valuable, it is important to carefully design APMs based on the specific care setting at hand [30]. As such, APMs should be designed depending multiple factors, such as the scope and the type of population, the level of standardization of care, the type of providers and the level of risk providers are able and willing to bear. At first sight, and apart from discussions on terminology of the models, the type of APM should vary with the level of risk the payment model imposes on (groups of) providers (also see Figure 2 [29]). The exact amount of risk that is imposed on (groups of) providers can be finetuned by using risk mitigation strategies. Conceptually, this creates a distinction between *provider risk* and *provider accountability*. That is, because risk mitigation strategies limit the financial risk that providers bear for while holding providers still accountable. The actual level of *risk* should be adjusted to what providers are able and willing to bear, while the *accountability* of providers may be larger. Key elements of APMs and risk mitigation strategies within the APM must be designed in such a way that it supports the intrinsic motivation of providers to deliver high value care [44]. That means that the model should, among others, not lead to excessive administrative burdens.

In the search to discover what works in which situation, it is important that space is being created to learn from best practices and also to learn from mistakes. Transparency about the aspects of APMs that need further development is required to prevent that unwanted incentives result from the APMs. Best practices, learning points, and also theoretical knowledge could be shared by using platforms. One aspect that deserves more attention is the notion that measurement of value is critical in designing APMs [30] and that, currently, we don't know how to measure value fully. This implies that tying payments to value-indicators makes more sense when value-indicators are considered to be valid. Another related aspect is that risk mitigations strategies, and in particular risk-adjustment, needs further refinement to properly separate out insurance risk and provider risk. Research on how to deal with population heterogeneity may also inform these issues. This is important to protect providers from incurring too much risk, and also to lower barriers for providers to participate in APMs. But most importantly, the

adoption of APMs should be approached as a part of a broad transformation including many factors that together may lead to a more valuable health system.

Conclusions

This thesis aimed to grasp the concept of value in a population health management context. By exploring how to measure value, this thesis showed that there is a need to develop comprehensive sets of value-indicators capturing the full continuum of care and outcomes that matter to the population. By gaining insight into what types of APMs are currently implemented and what experiences are, this thesis showed that progress to validly measure value is essential for the design and adoption of APMs. In the meantime, APMs should be carefully designed to care context at hand, while being transparent about their stage of development, available evidence and realistic expectations.

REFERENCES

1. ABIM. *Choosing Wisely Master List*. 2013 [cited 2014 12-24]; Available from: http://www.choosingwisely.org/wp-content/uploads/2013/02/Choosing-Wisely-Master-List.pdf.
2. Struijs, J.N., et al., *How to evaluate population management? Transforming the Care Continuum Alliance Population Health Guide into a broadly applicable analytical framework*. Health Policy, 2015. 119(4): p. 522-529.
3. Porter, M.E., *What Is Value in Health Care?* New England Journal of Medicine, 2010. 363(26): p. 2477-2481.
4. Duell, D., Koolman, X., and Portrait, F., *Practice variation in the Dutch long-term care and the role of supply-sensitive care: Is access to the Dutch long-term care equitable?* Health Econ, 2017. 26(12): p. 1728-1742.
5. Duell, D., et al., *Practice variation in long-term care access and use: The role of the ability to pay.* Health Econ, 2019.
6. The King's Fund, *Better value in the NHS: The role of changes in clinical practice*. 2015, The King's Fund: London.
7. Gutacker, N., et al., *Truly inefficient or providing better quality of care? Analysing the relationship between risk-adjusted hospital costs and patients' health outcomes.* Health Econ, 2013. **22**(8): p. 931-47.
8. Portrait, F.R., Van der Galien, O., and Van den Berg, B., *Measuring Healthcare Providers' Performances Within Managed Competition using Multidimensional Quality and Cost Indicators.* Health Econ, 2015.
9. Romley, J., et al., *Geographic variation in the delivery of high-value inpatient care.* PLoS One, 2019. 14(3): p. e0213647.
10. Mathias, J.S. and Baker, D.W., *Developing quality measures to address overuse.* JAMA, 2013. 309(18): p. 1897-8.
11. Glantz, J.C., *Obstetric Variation, Intervention, and Outcomes: Doing More but Accomplishing Less.* 2012. 39(4): p. 286-290.
12. De Jonge, A., et al., *Value based maternal and newborn care requires alignment of adequate resources with high value activities.* BMC Pregnancy Childbirth, 2019. 19(1): p. 428.
13. Newhouse, J.P., et al., *Variation in Health Care Spending: Target Decision Making, Not Geography*. 2013, National Academies Press (US): Washington (DC).
14. Landon, B.E., Wilson, I.B., and Cleary, P.D., *A conceptual model of the effects of health care organizations on the quality of medical care.* Jama, 1998. 279(17): p. 1377-82.
15. Robinson, J.C., *Theory and Practice in the Design of Physician Payment Incentives.* The Milbank Quarterly, 2001. 79(2): p. 28.
16. Javitt, J.C.R., James, B. & Reisman, L., *Information technology and medical missteps: Evidence from a randomized trial.* Journal of Health Economics, 2008. **27**: p. 585-602.
17. Bright, T.J., et al., *Effect of clinical decision-support systems: a systematic review.* Ann Intern Med, 2012. 157(1): p. 29-43.
18. Jackson, G.L., et al., *The patient centered medical home. A systematic review.* Ann Intern Med, 2013. 158(3): p. 169-78.
19. De Vries, E.F., et al., *Unraveling the drivers of regional variation in healthcare spending by analyzing prevalent chronic diseases.* BMC Health Services Research, 2018. 18(1): p. 323.
20. Heijink, R., et al., *A Window on Geographic Variation in Health Care: Insights from EuroHOPE.* Health Econ, 2015. 24 Suppl 2: p. 164-77.
21. Landon, B.E., et al., *Variation in patient-sharing networks of physicians across the United States.* Jama, 2012. 308(3): p. 265-73.
22. Skinner, J., *Chapter Two - Causes and Consequences of Regional Variations in Health Care1*, in *Handbook of Health Economics*, T.G.M. Mark V. Pauly and P.B. Pedro, Editors. 2011, Elsevier. p. 45-93.
23. Lavergne, M.R., et al., *Examining regional variation in health care spending in British Columbia, Canada.* Health Policy, 2016. 120(7): p. 739-48.

24. Gopffarth, D., Kopetsch, T., and Schmitz, H., *Determinants of Regional Variation in Health Expenditures in Germany.* Health Econ, 2015.

25. Schwartz, A.L., et al., *Low-Value Service Use in Provider Organizations.* Health Serv Res, 2018. 53(1): p. 87-119.

26. Miller, H.D., *From volume to value: better ways to pay for health care.* Health Affairs, 2009. 28(5): p. 1418-1428.

27. Nolte, E., Knai, C., McKee, M., *Managing chronic conditions: experience in eight countries*, in *European Observatory on Health Systems*, WHO, Editor. 2008.

28. Huerta, T.R., Hefner, J.L., and McAlearney, A.S., *Payment models to support population health management.* Adv Health Care Manag, 2014. 16: p. 177-83.

29. Frakt, A.B. and Mayes, R., *Beyond capitation: how new payment experiments seek to find the 'sweet spot' in amount of risk providers and payers bear.* Health Aff (Millwood), 2012. 31(9): p. 1951-8.

30. Conrad, D.A., et al., *Implementing Value-Based Payment Reform: A Conceptual Framework and Case Examples.* Med Care Res Rev, 2016. 73(4): p. 437-57.

31. Ruwaard, S., *Purchasing healthcare; Beyond setting the financial incentives right*. 2018, Tilburg University: Tilburg.

32. Ruwaard, S., Struijs, J.N., Douven, R., Polder, J., *Key drivers of payment reforms: Transferable Lessons from the Alternative Quality Contract.* Submitted, 2017.

33. Struijs, J.N., De Vries, E.F.., Baan, C.A., Van Gils, P., Rosenthal, M.B., *Bundled-Payment Models Around the World: How They Work and What Their Impact Has Been*. Issue Brief (Commonw Fund), Accepted.

34. Colla, C.H. and Fisher, E.S., *Moving Forward With Accountable Care Organizations: Some Answers, More Questions.* JAMA Intern Med, 2017. 177(4): p. 527-528.

35. Struijs J.N., et al., *Effects of bundled payment on curative health care costs in the Netherlands : An analysis for diabetes care and vascular risk management based on nationwide claim data, 2007-20100. RIVM report 260013001*. 2012, Rijksinstituut voor Volksgezonheid en Milieu: Bilthoven.

36. Song, Z., et al., *Health Care Spending, Utilization, and Quality 8 Years into Global Payment.* N Engl J Med, 2019. 381(3): p. 252-263.

37. Song, Z., et al., *Health care spending and quality in year 1 of the alternative quality contract.* N Engl J Med, 2011. 365(10): p. 909-18.

38. Song, Z., et al., *Changes in health care spending and quality 4 years into global payment.* N Engl J Med, 2014. 371(18): p. 1704-14.

39. Frakt, A.B., *'Value' of Care Was a Big Goal. How Did It Work Out?*, in *The Incidental Economist, The Health Services Blog*. 2019.

40. Antos, J.R., Pauly, M.V., and Wilensky, G.R., *Bending the Cost Curve through Market-Based Incentives.* 2012. 367(10): p. 954-958.

41. Emanuel, E., et al., *A Systemic Approach to Containing Health Care Spending.* 2012. 367(10): p. 949-954.

42. Steenkamer, et al., *Population health management guiding principles to stimulate collaboration and improve pharmaceutical care.* J Health Organ Manag, 2018. 32: p. 224–245.

43. Joynt Maddox, K.E. and McClellan, M.B., *Toward Evidence-Based Policy Making to Reduce Wasteful Health Care Spending.* JAMA, 2019. 322(15): p. 1460-1462.

44. Casalino, L.P. and Khullar, D., *Value-Based Purchasing and Physician Professionalism.* JAMA, 2019.

7

DANKWOORD

Toen ik in november 2014 aan dit promotietraject begon, leek het schrijven van dit dankwoord nog erg ver weg. Maar nu is het dan 'ineens' zover. Als ik terugkijk op deze periode, dan ben ik dankbaar voor de mooie momenten die ik heb meegemaakt en de levenslessen die ik heb mogen leren. Graag maak ik van deze gelegenheid gebruik om mijn dank te uiten aan mensen die een belangrijke rol hebben gespeeld in de afgelopen jaren. Het is nogal een uitdaging om dat luchtig te houden en niet te vervallen in zoetsappig geouweneel. Hierbij een poging.

Allereerst wil ik de leescommissie, prof. dr. **Diana Delnoij**, prof. dr. **Job van Exel**, prof. dr. **Ank de Jonge**, prof. dr. **Patrick Jeurissen** en prof. dr. **Johan Polder,** bedanken voor het lezen en beoordelen van dit proefschrift en de bereidheid te opponeren tijdens de verdediging.

Dan wil ik graag mijn (co)promotoren bedanken. **Caroline**, bedankt voor de vrijheid die je me hebt gegeven om mijn weg te vinden als onderzoeker, maar ook als vrouw binnen een organisatie. Ik ben dankbaar voor je scherpe vragen en bewonder je heldere denken. Je hebt mij geleerd hoe je nuance kunt aanbrengen in het werk, en vooral hoe belangrijk de nuances zijn. **Richard**, jou wil ik graag bedanken voor je scherpte in je commentaar. Sinds je start bij de NZa was je wat meer op afstand, maar ondanks dat heb je je altijd goed kunnen inlezen en had je goede adviezen. Daarnaast was je door je kalmte een baken van rust, wat op sommige momenten erg fijn was. **Jeroen**, ik heb genoten van onze samenwerking en dat doe ik nog steeds. Ik overdrijf niet als ik zeg dat ik dit promotietraject niet had kunnen volbrengen zonder jouw begeleiding en persoonlijke contact. De manier waarop we met elkaar communiceren is oprecht, direct en hard op de inhoud. Heerlijk, want hierdoor kunnen we beter tot de kern komen in onze discussies. Ik heb meer dan eens aan omstanders van onze telefoongesprekken moeten uitleggen dat we geen ruzie hadden, maar dat we 'gewoon' een discussie hadden. Best geestig, eigenlijk. Bedankt voor de kans om de (DIAPER) dataset rondom zwangerschap en geboorte te ontwikkelen. Ik ben blij dat we ook in de toekomst zullen blijven samenwerken op het RIVM.

Een aanzienlijk deel van dit proefschrift is tot stand gekomen dankzij de beschikbaarheid van observationele data van nagenoeg alle **inwoners van Nederland**. Ik wil graag van deze gelegenheid gebruik maken om alle betrokkenen, maar in het bijzonder **Perined** en **Vektis,** te bedanken voor het beschikbaar stellen van data. Daarnaast wil ik graag het **Centraal Bureau voor de Statistiek** bedanken voor het beschikbaar stellen van een groot aantal datasets, zoals de Gemeentelijke Basisadministratie en inkomensgegevens van de Belastingdienst. Ook is het bieden van een platform om op een beveiligde en

verantwoorde wijze de veelheid aan data te kunnen koppelen en analyseren van wezenlijk belang geweest. De komst van de Remote Access op de eigen computer heeft mijn leven een stuk eenvoudiger gemaakt, al vond ik de periode dat ik bij de **afdeling Statistiek, Informatica en Modellering** zat erg leuk. Dank voor jullie gastvrijheid en hulp!

Dat brengt mij op de gezellige club van collega's binnen het **centrum Voeding, Preventie en Zorg**. Ik ben dankbaar voor de prettige werksfeer waarin ook simpele grappen gemaakt kunnen worden. In het bijzonder wil ik graag **Annerieke**, **Claudia, Ellen** en **Roy** bedanken voor de gezelligheid, koppen koffie en - ook - het delen van de frustraties die komen kijken bij het schrijven van een proefschrift. Daarnaast wil ik **Albert**, **Eelco**, **Hanneke**, **Mieneke** en **Roy** bedanken voor jullie adviezen en het meeschrijven aan mijn artikelen. **Joyce** en **Zoë**, ik ben blij dat jullie er zijn. Samen met Jeroen vormen we een echt team, en ik heb veel zin om samen met jullie verder de mogelijkheden van DIAPER te verkennen! **Josefien**, bedankt voor het aller gezelligste congres ooit; San Sebastian zal ik nooit vergeten. Super leuk dat je vanaf januari 2020 nu ook officieel mijn collega bent.

Lieve **vrienden**, ik wil jullie allemaal bedanken voor de gezelligheid, borrels en vriendschap. In het bijzonder wil ik MASSEL, **Maaike**, **Astrid**, **Sabine** (Zeebaain), **Suzanne** en **Laura**, bedanken. Dankzij jullie ben ik het gaan waarderen om onderdeel te zijn van een groep vrouwen. We hebben samen veel meegemaakt in de afgelopen jaren. Lol, feest en stapels nieuwe kinderen, maar ook verdriet en verlies. We zijn allemaal compleet anders, maar stuk voor stuk zijn jullie fantastisch. Door onze drukke levens lukt het vaak niet meer om elkaar wekelijks te zien, maar daar wordt de vriendschap niet minder door. Ik ben dankbaar dat jullie er voor me zijn. **Joanne**, jij bent mijn oudste vriendin. Ik kijk met veel plezier terug aan de te gekke momenten die we als pubers hebben beleefd (arme ouders) en geniet van de momenten nu dat we lekker een kop thee drinken onder het genot van spelende kinderen. Lieve **Eline**, ik ben blij dat we vriendinnen zijn geworden. Je bent echt heel grappig en het is heerlijk goede gesprekken met je voeren op een van onze day drink dagen. **Ward**, de volgende keer ga jij ook mee, hoor. Ook wil ik graag mijn paardenvriendinnen, **Anja**, **Christy**, **Myrthe**, **Stephanie** en **Yvette**, bedanken. Voor het delen van onze hobby, de gezellige lessen en de hilarische en – soms ook - levensgevaarlijke ritten door de duinen. Maar vooral ook voor de goede gesprekken en steun.

Sanne en **Suzanne**, dank dat jullie mij op deze dag willen ondersteunen! Sanne, we hebben samen veel meegemaakt, je hebt mij op m'n best gezien en op m'n slechtst. Bedankt voor je humor, oprechte interesse en je vriendschap. Ik kan niet wachten om

samen weer op avontuur te gaan. Suzanne, ik heb genoten van onze PhD-tijd samen op het RIVM. Tot nu toe is het alleen jou gelukt om mij structureel van mijn computer los te weken rond lunchtijd. Bedankt voor de komische intermezzo's, IT Crowd en de goede gesprekken. Sorry **Ben** en **Gerben**, volgende keer mogen jullie mijn paranimfo's zijn.

Beste **schoonfamilie**, ik heb het maar met jullie getroffen. **Arnold** en **Marja**, bedankt voor jullie interesse en jullie steun. Ik zal niet snel vergeten hoe lief jullie er voor mij waren, toen ik eind november alles op alles moest zetten om het manuscript van dit proefschrift nog vóór de overgang naar het nieuwe systeem in te leveren (en Ard natuurlijk nét op vakantie was met de mannen). Bedankt dat Floris mag opgroeien met een stapel aan neven, **Mark**, **Luuk**, **Tom** en **Mik**, die hij elke vrijdag bij jullie kan zien. **Bram**, **Merel**, **Job** en **Simone**, bedankt voor de gezelligheid op alle familiebijeenkomsten en de interesse in mijn eindeloze afstudeerproject.

Lieve **ouders met aanhang en broer**, 'de zorg' is mij met de paplepel ingegoten. Niet alleen vanwege ziekte thuis, maar ook omdat jullie met hart en ziel in de zorg werken. Jullie verhalen over bijzondere ontmoetingen met patiënten en ook frustraties over het disfunctioneren van het zorgsysteem hebben mij geïnspireerd om onderzoek te doen in de zorg. Het leven is lang niet altijd makkelijk geweest. Desondanks blijven jullie met passie doorgaan en is er altijd ruimte om op Floris te passen. Bedankt voor jullie levenslessen en alles wat jullie mij hebben gegeven om te worden wie ik nu ben. Ik wens jullie alle goeds en vele gezonde en gelukkige jaren toe. Ik ben wel enorm opgelucht dat jullie me nu niet meer gaan vragen of ik al 'afgestudeerd' ben. Lieve **zus**, jou kan ik natuurlijk niet weglaten in dit rijtje. Hoe is alle zorg die jij hebt ontvangen voor jou van waarde geweest? Voor veel ben ik dankbaar, maar in het kader van dit proefschrift wil ik je bedanken voor het inzicht dat een op zichzelf staande handeling in de zorg weinig kan betekenen. De zorg levert alleen echt iets op als alle zorg, op alle mogelijke gebieden rond het leven van een persoon, in samenhang georganiseerd wordt. Allerliefste **oma**, wat is het een voorrecht om zo'n band met je grootmoeder te hebben. Je bent er altijd voor mij geweest en met jouw 91 jaar ben je nog steeds van alles perfect op de hoogte. Ik vind het prachtig hoe je met Floris bent. Dank je wel voor alles.

Lieve **Ard** en **Floris**, ik bof maar met jullie. Jullie zijn mijn veilige thuis. Ard, ik ben onder de indruk van jouw discipline, rechtvaardigheidsgevoel en geduld. Ik hoop dat ik in de jaren samen wat van jou heb geleerd op die vlakken. Je geeft mij alle ruimte om mij te ontwikkelen en dat heeft mij enorm geholpen bij het afronden van dit proefschrift in combi met onze kleine boef. Bedankt dat je er bent en dat je zoveel geduld hebt met mij.

Lieve Floris, ik geniet elke dag van jouw aanwezigheid. Ik kijk uit naar onze toekomst, zonder al die extra proefschriftuurtjes!

Poging mislukt, het is een aardig zoetsappig dankwoord geworden..

ABOUT THE AUTHOR.

Eline Frouke de Vries was born in Haarlem on July 31th 1987. She started high school at *gymnasium Felisenum* in Velsen-Zuid. In 2005, she started studying Medicine at *Vrije Universiteit medisch centrum*. Because of personal experiences, she felt the importance and complexity of the concept of value in health care services. Eline learned that providing more health services is not always what is best and that it is sometimes more valuable to stop or change treatment. In 2007, she decided to quit her studies and started fulltime working at an online shop. In the evening hours, she followed a study in Management, Economics and Law with a specialization in Healthcare Management at the Amsterdam University of Applied Sciences and obtained her Bachelor of Business Administration (BBA) in 2012. Hereafter, she started her masters study in Health Economics, Policy and Law (HEPL) with a specialization in Health Technology Assessment at the Erasmus University Rotterdam. She was rewarded a certificate of achievement, meaning that she performed in the top 25% of cohort 2013-2014. Eline's interest in conducting research was gained during the writing of her masters' thesis on the relationship between health and happiness. In 2014, she started working as a researcher at Leiden University Medical Center, where she worked on a model for the early economic evaluation of treatments for chronic kidney disease. After finishing this project, Eline started a PhD project on value in population health management, which resulted in this dissertation. She was officially affiliated at Tranzo, Tilburg School of Social and Behavioral Sciences and in daily practice worked at the National Institute for Public Health and the Environment (Rijkinstituut voor Volksgezondheid en Milieu (RIVM)). Eline supervised MSc theses, lectured about integrated care and payment models in maternity care and served as a referee for multiple journals, among which Nephrology Dialysis Transplantation, Health Policy and the British Medical Journal. During her research, Eline has built experience in merging and handling large observational datasets and performed several econometrical analyses. She continues her work at the RIVM.

Eline lives in Haarlem with her husband Ard and son Floris.

PUBLICATION LIST

De Vries EF, Rabelink TJ, Van den Hout WB. *Modelling the cost-effectiveness of delaying end-stage renal disease.* Nephron. 2016;133(2): 89-97

De Vries EF, Struijs JN, Heijink R, Hendrikx RJP, Baan CA. *Are low-value care measures up to the task? A systematic review of the literature.* BMC Health Services Research. 2016;16(1):405

De Vries EF[1], Van de Wetering EJ[1], Tsiachristas A, Van Exel NJA. *How important is health for happiness?* In: Lawerman-van de Wetering, EJ (2016). Valuing and Refining Outcome Measures for Economic Evaluations in Health Care. Rotterdam: Optima.

De Vries EF, Heijink R, Struijs JN, Baan CA. *Unraveling the drivers of regional variation in healthcare spending by analyzing prevalent chronic diseases*. BMC Health Services Research. 2018:18(1):323

De Vries EF, Drewes HW, Struijs JN, Heijink R, Baan CA. *Barriers to payment reform: experiences from nine Dutch population health management sites.* Health Policy. 2019; 123(11):1100-1107

Struijs JN, De Vries EF, Baan CA, Rosenthal M. *Bundled payments around the world*. Issue Brief (Commonwealth Fund). (accepted).

De Vries EF, Over EAB, Wong A, Baan CA, Heijink R, Struijs JN. *Measuring value in maternity care using observational data. A first attempt with Dutch Maternity Care Networks.* (submitted).

De Vries EF, De Bruin-Kooistra M, Baan CA, Struijs JN. *Alternative payment models in maternity care and their effects on health and spending. A scoping review.* (submitted).

1 Shared first authorship.

Presentations at scientific congresses

De Vries EF, Rabelink TJ, Van den Hout WB[2]. *Modelling the cost-effectiveness of delaying end-stage renal disease.* Lowlands Health Economics Study Group 2015, Maastricht, the Netherlands.

De Vries EF[3], Struijs JN, Heijink R, Hendrikx RJP, Baan CA. *Low-value care measures, are they up to the task?* European Health Economics Association 2016, Hamburg Germany.

De Vries EF[3], Heijink R, Struijs JN, Baan CA. *Regional variation in healthcare spending, a disease-based approach.* European Health Economics Association 2016, Hamburg Germany.

De Vries EF[3], Struijs JN, Drewes HW, Heijink R, Baan CA. *Payment reform in population health management: lessons from Dutch pioneer sites.* International Congress of Integrated Care 2017, Dublin, Ireland. Abstract published in the International Journal of Integrated Care. 2017;17(5):A129.

De Vries EF[3], Struijs JN, Heijink R, Hendrikx RJP, Baan CA. *Low-value care measures, are they up to the task?* International Congress of Integrated Care 2017, Dublin, Ireland. Abstract published in the International Journal of Integrated Care. 2017;17(5):A130.

De Vries EF[3], Heijink R, Struijs JN, Baan CA. *Studying regional variation in healthcare spending using prevalent chronic diseases.* International Congress of Integrated Care 2018, Nieuwegein, the Netherlands. Abstract published in the International Journal of Integrated Care. 2018; 18(S2): A320, pp.1-8.

Struijs J, van Dorst H[3], de Vries E[3]. *Workshop integrating care by payment reform; the case of Dutch birth care.* International Congress of Integrated Care 2019, San Sebastian, Basque. Abstract published in the International Journal of Integrated Care. 2019;19(4):319.

De Vries EF, Over EAB, Wong A, Heijink R, Baan CA, Struijs JN.[4] *Identifying value in maternity care networks.* Lowlands Health Economics Study Group 2019, Almen, the Netherlands.

2 Work was discussed by Silvia Evers.
3 Presenter.
4 Work was discussed by Joaquim Vidella-Martin.