Incentivizing appropriate prescribing in primary care

Ramerman, Lotte; Hek, Karin; Cramer- van der Welle, Christine; Simons-Winters, Ellis; Middelweerd, Anouk; Lambooij, Anke; Verheij, Robert A.

Published in:
Health Policy

DOI:
10.1016/j.healthpol.2022.07.004

Publication date:
2022

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

Health policy 126 (2022) 1010–1017

Contents lists available at ScienceDirect

Health policy

journal homepage: www.elsevier.com/locate/healthpol

Incentivizing appropriate prescribing in primary care: Development and first results of an electronic health record-based pay-for-performance scheme

Lotte Ramerman [a,*], Karin Hek [a], Christine Cramer- van der Welle [a], Ellis Simons-Winters [a], Anouk Middelweerd [a,b], Anke Lambooij [c], Robert Verheij [a,d]

[a] Netherland Institute for Health Services Research (Nivel), Otterstraat 118, Utrecht 3513CR, the Netherlands
[b] Techmed Centre, Personalized eHealth Technology, ZuidHorst, P.O. box 217, Enschede 7500AE, the Netherlands
[c] Dutch Institute on Rational Use of Medicine, Postbus 3089, Utrecht 3502GB, the Netherlands
[d] Tranzo, Tilburg School of Social and Behavioral Sciences, Tilburg University, Postbus 90153, Tilburg 5000LE, the Netherlands

ARTICLE INFO

Keywords:
Pay-for-performance
Primary care
Medication
Electronic health records
Formularies
General practitioner

ABSTRACT

Objective: Part of the funding of Dutch General Practitioners (GPs) care is based on pay-for-performance, including an incentive for appropriate prescribing according to guidelines in national formularies. Aim of this paper is to describe the development of an indicator and an infrastructure based on prescription data from GP Electronic Health Records (EHR), to assess the level of adherence to formularies and the effects of the pay-for-performance scheme, thereby assessing the usefulness of the infrastructure and the indicator.
Methods: Adherence to formularies was calculated as the percentage of first prescriptions by the GP for medications that were included in one of the national formularies used by the GP, based on prescription data from EHRs. Adherence scores were collected quarterly for 2018 and 2019 and subsequently sent to health insurance companies for the pay-for-performance scheme. Adherence scores were used to monitor the effect of the pay-for-performance scheme.
Results: 75% (2018) and 83% (2019) of all GP practicesparticipated. Adherence to formularies was around 85% or 95%, depending on the formulary used. Adherence improved significantly, especially for practices that scored lowest in 2018.
Discussion: We found high levels of adherence to national formularies, with small improvements after one year. The infrastructure will be used to further stimulate formulary-based prescribing by implementing more actionable and relevant indicators on adherence scores for GPs.

1. Introduction

Formularies, guidelines on medication, may play a role in stimulating appropriate prescribing by General Practitioners (GPs). In the Netherlands, formularies are available on a national level and regional level, the latter is sometimes disease-specific. The national formularies provide a concise overview of recommendations on prescribing medication, taking into account both the effectiveness and cost of medication and are aimed at a national audience. These national formularies cover the whole specter of medications and diseases, while regional formularies encompass *additional* and sometimes *other* prescribing recommendations. They may also focus on one or several specific diseases. Regional formularies are generally more specific than formularies on the national level [1]. Adherence to formularies varies between prescribers [2] suggesting there is room for improvement. However, implementing evidence-based practices, such as the recommendations provided in formularies that stimulate appropriate prescribing, is not an easy process [1,3–5].

A pay-for-performance scheme as part of the general practice reimbursement system, supplementary to for example a fee-for-service model or capitation fee [6,7], is one way to stimulate formulary adherence and thus improve the quality of prescribing while reducing health care costs [8]. Reimbursement systems in general have proven to have an effect on guideline adherence. Furthermore, pay-for-performance schemes have

* Corresponding author.
E-mail address: l.ramerman@nivel.nl (L. Ramerman).

https://doi.org/10.1016/j.healthpol.2022.07.004
Received 2 July 2021; Received in revised form 29 April 2022; Accepted 13 July 2022
Available online 17 July 2022

been implemented before, with varying results, to improve health care performance or quality [6,8–14]. By making quality of care a direct component of the finances of healthcare providers, it is expected that pay-for-performance will stimulate the improvement of the quality of care [10]. Studies showed an improvement in outcomes on quality indicators and positive effects on the variation between clinicians after the introduction of smaller and larger scale pay-for-performance schemes in different clinical settings, including primary healthcare [8,9,13–18] and for prescribing outcomes [9] based on formularies [3].

There are several aspects of pay-for-performance schemes that promote the effectiveness of the scheme. First, the scope of the pay-for-performance scheme (number of indicators) should be considered. Narrow schemes result in a focus on a small part of care, while broad schemes draw too much attention to performance [13]. Second, GPs are more likely to be committed and motivated when they are involved in designing the scheme, when there is professional consensus on indicators, and if the data used to calculate indicators is representative [9,13,19]. Third, pay-for-performance schemes need clearly defined, actionable indicators and constant adjustments to account for the effect of reaching a certain optimum in performance measurements [9,19,20]. Last, according to van Herck [9], pay-for-performance is most effective when it is aimed at quality improvement, with clear targets for clinicians [9,13].

Following the philosophy of value-based healthcare and Learning Health Systems (LHS), routine healthcare data from Electronic Health Records (EHR) may be used to stimulate adherence to formularies. By using routine healthcare data, the learning capacity of the health system is stimulated as a whole and the generation of knowledge is made an integrated part of the healthcare delivery process [21–23], while keeping the administrative burden low. In an LHS, routine healthcare data is used to enable cycles of continuous learning and improvement, enhancing the value of healthcare for patients and our health system in general, including the costs for our health system [23,24]. These ideas have been brought into practice and studied in primary care in several studies [25–28]. Using data from EHR systems may be an efficient way to establish clinically relevant indicators for pay-for-performance schemes [29–32]. Functions of EHR systems to support GPs, such as decision support or feedback tools, can facilitate GPs in improving daily practice, keeping in mind the GPs' own professionalism and freedom in choosing the right treatment for their individual patient [17]. However, previous studies also showed pitfalls in using EHR systems as the basis for pay-for-performance indicators. GPs experienced a negative impact on the continuity of care, as data collection for the scheme was time-consuming and the EHR software lacked extraction possibilities [17,33,34].

Incitives for formulary based prescribing have been part of the reimbursement system before. However, before 2018, the incentive for appropriate prescribing in the Netherlands was based on dispensing data from pharmacies, rather than prescribing data from GP practices, which diminishes the involvement of the GP's and reduced the feeling of ownership and control of the GP. Therefore, there was a need for a new infrastructure using routine healthcare data from EHRs to establish indicators on appropriate prescribing for the pay-for-performance scheme.

This paper aims (1) to describe the development of an indicator for appropriate prescribing based on formularies and GP practices' EHRs, (2) to describe the results of the pay-for-performance scheme on the outcomes of the indicator, and (3) to provide recommendations for further improvements on the indicator, infrastructure and pay-for-performance scheme.

2. Methods

2.1. Population and setting

Both the indicator and infrastructure were developed in collaboration with formularies, four major health insurance organizations, the national general practice trade union (LHV), the national organization for health insurance (ZN), and an umbrella organization for EHR suppliers (NedXis). The latter four were organized in a steering committee for the pay-for-performance scheme.

GP practices in the Netherlands, contracted by one of the four main health insurers (jointly covering 90% of the total number of practices), were eligible for the pay-for-performance reimbursement. Ten different EHR systems are available for GPs to choose from, all of which provide access to e-prescribing applications, provided by the EHR supplier or by a separate supplier for e-prescribing applications. All ten EHR systems were included for the data collection in this study.

The e-prescribing modules providing prescription suggestions are incorporated in all EHR systems and encompass any of the national formularies and may also support the use of regional formularies. They can be consulted as part of the prescribing process. Two national formularies were included in the study: the NHG formulary (i.e., the formulary of the Dutch National College of GPs), and the Health Base formulary, a national formulary specific to one EHR. Depending on the EHR system the GP used, they had access to either the NHG formulary or the Health Base formulary. The NHG formulary included 326 different Anatomical Therapeutic Chemical Classification (ATC) codes in 2018. The Health Base formulary included 316 of the NHG ATC-codes and an additional 239 ATC-codes, resulting in a total of 555 individual ATC-codes in 2018 for the same number of indications.

2.2. Indicator for formulary-conform prescribing

An indicator was developed to calculate the percentage of formulary-conform prescriptions for each GP practice every three months. A score of 100% adherence to the used formulary is explicitly not the aim of the study or the pay-for-performance scheme, as GPs may deviate from the formulary to meet the needs of individual patients.

Denominators were defined as the total number of new first prescriptions issued in the GP practice, for patients registered as patients in the practice only. This was possible because virtually every Dutch citizen is listed as patient in a particular practice. Prescriptions were regarded as first prescriptions if the same medication (same ATC7-code) had not been prescribed in the preceding 12 months. Prescriptions without a valid ATC-code, and codes starting with V, Y, or Z, as well as vaccinations (ATC-codes starting with J07) were excluded. *Numerators* were defined as the total number of first prescriptions by the GP of which the ATC7-code was included in the national formulary used within their EHR system.

2.3. Infrastructure for data collection

Both denominators and numerators were calculated within each practices' EHR system or the additional application from the e-prescribing supplier, thus avoiding the need to share individual patient-level data with researchers or health insurers. Denominators and numerators were calculated in the EHR system of the GP practice and sent to the researchers. After checks on correctness and validity, scores were sent to health insurance companies, to set off the reimbursement process. Inclusion of practices in the project was part of the standard reimbursement agreements negotiated annually between practices and insurance companies. However, EHR supplier only provided the scores after explicit consent from the GP practices.

2.4. Data validation

To ensure the accuracy of the indicator for each practice, the methodology and operationalization of the specifications of the denominator and the numerator by the EHR systems was validated with individual patient level data using a limited number of test practices for each of the ten EHR system suppliers. Denominators and numerators were calculated by the EHR supplier and then reviewed by the

researchers, by comparing them with their own calculations. When possible, already available data from the Nivel Primary care database was used for the validation [35].

2.5. Data collection

For each GP practice, denominators and numerators were collected quarterly for 2018 and 2019. The scores were calculated and validated on completeness and correctness by Nivel, before sending them to the health insurance companies for the pay-for-performance reimbursement. Fig. 1 illustrates the infrastructure of the project, starting with first prescriptions from the GP and ending at GPs receiving a pay-for-performance incentive for formulary-conform prescribing.

2.6. Analyses

Practices with fewer than 100 first prescriptions quarterly were excluded from the analyses, as this was regarded as implausible and likely to be a result of data flow failures. Due to differences in included ATC-codes between the national formularies, scores were compared between practices using the same formulary. For each formulary the mean score and standard deviation (SD) were determined quarterly and annually. In addition, practice variation was expressed by describing scores between the 10th and 90th percentile. Changes in formulary-conform prescribing after the pay-for-performance incentive were tested with three two-level multilevel model, using an unstructured covariance matrix for the residuals of the repeated measures (quarters). The dependant variable and the first level in all models were the quarterly scores on formulary-based prescribing (outcome of the indicator). The second level included the GP practices.

The first model studied the effect of the pay-for-performance incentive on the score by comparing 2018 and 2019. Therefore, a main effect for year was added to the model. The second model analysed the effect of lower-scoring practices (practices with a score in the 25th-percentile in 2018) on the level of change from 2018 to 2019. Therefore we included main effects for year, lower or higher-scoring practice and an interaction between the two. With the third model, we analysed the effect of the EHR suppliers on the level of changes in scores between 2018 and 2019. Therefore, main effects for year, EHR supplier and an interaction between the two were added to the model. EHR suppliers were excluded from the analysis when the number of general practices using that EHR was less than 10. The significance threshold for all analyses was $p < 0.05$. Stata version 15.0 was used to perform the analyses.

3. Results

In 2018, 75% ($n = 3755$) and in 2019, 83% ($n = 4125$) of all eligible GP practices in the Netherlands participated in the pay-for-performance scheme and provided formulary-adherence scores. Approximately 60% of GP practices used the NHG formulary, while 40% used the Health Base formulary. Table 1 shows the results for these GP practices. In 2018 168 (1%) GP practices and 131 (0.8%) in 2019 were excluded from analysis, because they had too few prescriptions.

Table 1
Results of GP practices participating in the pay-for-performance scheme for formulary-conform prescribing, including the adherence scores and practice variance in 2018 and 2019.

	2018		2019	
	NHG	Health Base	NHG	Health Base
Number of GP practices (n)	2337	1562	2599	1742
Number of first prescriptions (mean (SD))	1121	1260	1117	1309
Number of first prescriptions not-conform formulary (mean (SD))	166	54	159	58
Adherence scores (Mean (SD))	85.4% (2.94)	95.7% (2.00)	86.1% (2.72)	95.6% (1.45)
Adherence scores (10–90% range)	81.5–88.8%	93.8–97.3%	82.3–89.1%	93.9–97.0%
Adherence scores lowest-scoring practices* (mean SD))	81.4% (2.09)	93.2% (2.56)	82.9% (2.39)	94.0% (1.69)
Adherence scores lowest-scoring practices* (10–90% range)	78.6–83.4%	89.7–95.1%	79.7–85.5%	91.7–95.8%
Adherence scores highest-scoring practices** (mean SD))	86.7% (2.31)	96.4% (1.23)	86.8% (2.70)	95.8% (1.39)
Adherence scores highest-scoring practices** (10–90% range)	83.8–89.6%	95.2–97.6%	83.6–89.9%	94.4–97.3%

* all practices with a mean score within the 25th-percentile in 2018.
** all practices with a mean score within the highest 75th-percentile in 2018.

3.1. Improvements in formulary-conform prescribing

Fig. 2a presents the mean quarterly adherence scores on both national formularies for 2018 and 2019. The scores of GP practices that use the NHG formulary improved slightly and significantly, with a mean increase of 0.7%-point ($p < 0.001$). The mean scores for the Health Base formulary decreased slightly, but significantly, (0.09%-point; $p = 0.022$) between 2018 and 2019 (Table 2).

3.2. Improvement in formulary-conform prescribing of lowest scoring practices

Subsequently, the practices that scored within the 25th percentile in 2018 were selected (Tables 1, 2 and Fig. 2b) and compared with the improvement in scores made by the higher scoring practices between 2018 and 2019, showing a significant higher level of improvement. GP

Fig. 1. Infrastructure for collecting adherence scores on the indicator for formulary-conform prescribing and providing pay-for-performance incentives to GPs by the health insurers: GPs (~5000) prescribe medication for their patients (**1**), using their own EHR system, with access to formularies (NHG or Health Base) (**2**). GPs have access to feedback on formulary-conform prescribing in their own EHR (**3**), while Nivel receives aggregated prescription data from 10 different EHR-suppliers for calculating indicator scores (**4**). Nivel sends the scores to health insurance companies (**5**). GPs receive an incentive-based on their score (**6**).

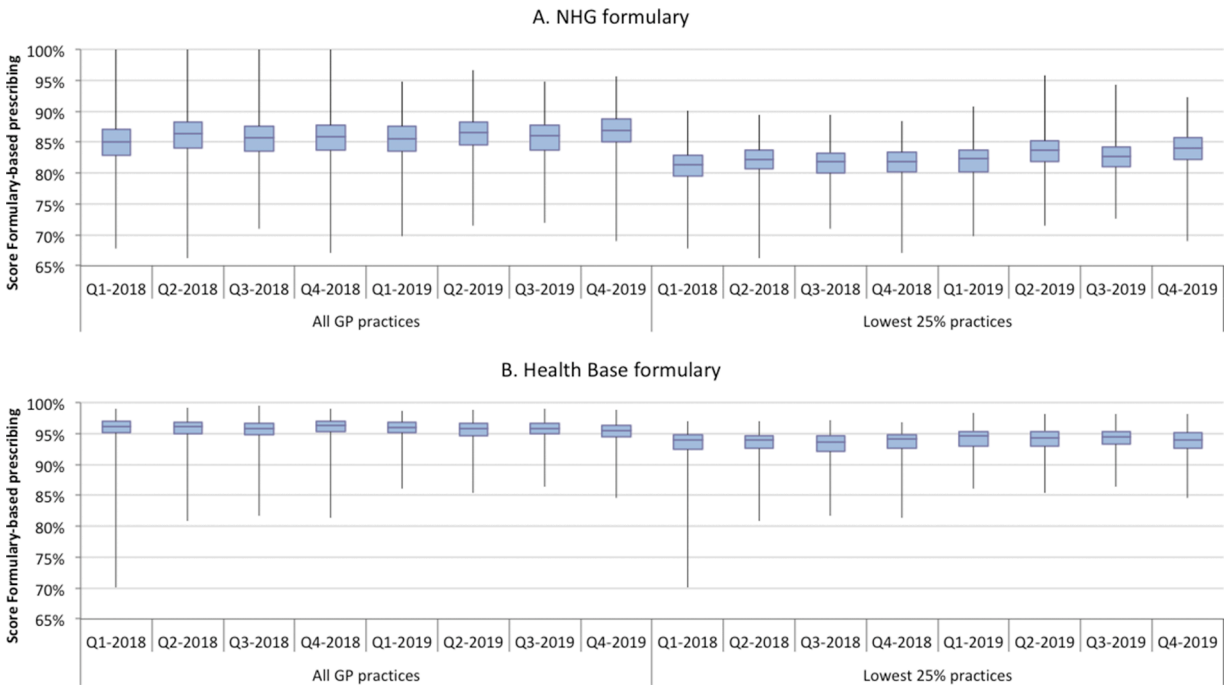

Fig. 2. a Boxplots of the adherence scores on formulary-conform prescribing of all GP practices on the NHG formulary and the Health Base formulary for each quarter of 2018 and 2019. b. Boxplots of the adherence scores on formulary-conform prescribing of GP practices scoring in the 25th percentile on the NHG formulary and the Health Base formulary for each quarter of 2018 and 2019.

practices using the NHG formulary increased their score significantly with an average of 1.5%-point in 2019 ($p < 0.001$), compared to an average increase of 0.8%-point for GP practices using the Health Base formulary ($p < 0.001$).

3.3. Differences between EHR suppliers

Fig. 3 shows the mean scores over time (2018–2019) for GP practices strafied by EHR supplier. EHR-3 offers a different formulary (Health Base formulary) compared to the other EHR suppliers and is therefore not included in further analyses. There were limited, but significant differences between the mean scores of practices of most of the EHR suppliers (Table 2; $p < 0.001$). However, the interaction between EHR supplier and year was not significant, except for EHR7 ($p = 0.001$). This indicates an effect of the EHR supplier on the mean scores of GPs for formulary-conform prescribing, but not on the ability of the GP to improve the level of formulary-conformity.

4. Discussion

This study describes the development and first results of a new national infrastructure and indicator based on prescription data from the EHR of the GP, for measuring appropriate prescribing using formularies. The level of formulary-conform prescribing was calculated using this indicator and infrastructure and was applied in a pay-for-performance scheme and subsequently evaluated.

The results of both 2018 and 2019 showed that GPs often prescribe medication conform the formulary they used, resulting in high percentages of formulary-conform prescribing (85.4–95.6%). This indicates a high level of formulary-conform prescribing which is in line with previous studies [2,5,36–39]. Even though scores were generally high, there was a small but significant improvement from 2018 to 2019 in all GP practices using the NHG formulary (i.e. one of the two national formularies in Dutch general practice). These small improvements were also found in other studies evaluating the effects of pay-for-performance schemes [12]. In general, previous studies found that narrow outcome models for pay-for-performance schemes (with few indicators), as presented in the current study, have more limited effects than pay-for-performance schemes, with a wider range of indicators [6,13]. Furthermore, the level of improvement may also be impacted by the cut-off value standard for reimbursement used by the health insurance companies: both a relative standard (using quartiles based on practice variance) and an absolute standard (using a single pre-set cut-off) were applied, depending on the health insurance organizations reimbursement policy. Health insurance organizations are not allowed to make agreements on their reimbursement policy's, resulting in the use of different cut-off standards. Previous pay-for-performance studies used similar standards for allocating the incentives, finding mixed results for both relative and absolute models [9,13,14,18,19,40,41]. A relative standard is more likely to promote gradual improvement, but were found to penalized GP practices unjustifiably when all score high [13,18]. An absolute threshold is perceived as more "fair" [33], but a drawback can be that once the target values have been reached, the stimulating effect of the indicator stops [18]. Further studies should focus on the variation between health insurer's policies in the level of improvement to assess the impact of different types of cut-off values.

Moreover, the initial scores of GPs on the indicator were relatively high, making improvements considerably more difficult, compared to starting with lower baseline scores [18]. This was especially true for GPs that used the Health Base formulary (i.e. one of the two national formularies in Dutch general practice). They scored approximately 10%-point higher than GPs using the NHG formulary, due to differences in the included number of drugs (ATC-codes); the Health Base formulary included 555 ATC-codes and the NHG formulary 316 ATC-codes. GPs who used the Health Base formulary thus have fewer restrictions in the medication they prescribe. Therefore, higher scores are easier to obtain, but improvements more difficult to achieve. With both relative and absolute standards used for the reimbursement, ceiling effects should be considered when establishing cut-off points for reimbursement, leaving room for GPs to deviate from the formulary to choose the right treatment for a specific patient (a score of 100% is not the goal). According to Roland and Olesen [13], ceiling effects may be a reason to discontinue

Table 2
Results of two-level multilevel analyses, including separate models for the general improvement between 2018 and 2019 (model 1), the improvement of lower scoring practices (model 2) and the impact of EHR suppliers (model 3).

	NHG formulary			Health Base formulary		
	Coef.	SE	Sig.	Coef.	SE	Sig.
Model 1: Main effect year						
Constant	−10.104	0.7000	*p < 0.001*	2.837	0.820	*p = 0.022*
Year	0.005	0.0003	*p < 0.001*	−0.001	0.0004	*P = 0.001*
Model 2: Lower scoring practices*						
Constant	0.865	0.0005	*p < 0.001*	0.964	0.0004	*p < 0.001*
Year	0.002	0.0004	*p < 0.001*	−0.004	0.0004	*p < 0.001*
Lower scoring practices	−0.049	0.0010	*p < 0.001*	−0.032	0.0009	*p < 0.001*
Year*lower scoring practices	0.011	0.0008	*p < 0.001*	0.0132	0.0009	*p < 0.001*
Model 2: Impact of EHR supplier**						
Year	0.005	0.0007	*p < 0.001*			
EHR1/constant	0.860	0.0011	*p < 0.001*			
EHR2	−0.008	0.0017	*p < 0.001*			
EHR4	−0.007	0.0017	*p < 0.001*			
EHR5	0.001	0.0017	*P = 0.652*			
EHR6	−0.032	0.0025	*p < 0.001*			
EHR7	−0.012	0.0024	*p < 0.001*			
EHR8	0.0004	0.0022	*P = 0.855*			
Year*EHR2	0.0009	0.0011	*p = 0.414*			
Year*EHR4	0.0001	0.0010	*p = 0.888*			
Year*EHR5	−0.0013	0.0011	*p = 0.210*			
Year*EHR6	0.0012	0.0015	*p = 0.402*			
Year*EHR7	0.0052	0.0015	*p = 0.001*			
Year*EHR8	−0.0024	0.0014	*p = 0.085*			

*All practices with mean score within the 25th-percentile in 2018.
**EHR3 was excluded from the analysis because only the Health Base formulary was available to its users.

the measurement of an indicator within a pay-for-performance scheme [13].

A more substantial improvement in formulary-based prescribing was found in the GP practices that scored lowest in 2018 (mean score in first quartile), for both national formularies. A higher potential for improvement resulted in more steep improvements [6,18]. However, it is uncertain if these changes are due to the incentive, more awareness of the use of formularies (signalling power of a pay-for-performance scheme) [6], a combination of both, or due to other unknown factors, such as differences in EHR system functions [42], access to real-time feedback [1] or changes in the formularies. In addition, for both the higher and lower scoring practices, policy makers should consider if the improvements represent a meaningful reform in appropriate prescribing, that is in line with the costs of the pay-for-performance scheme [18]. It is unknown whether the improvement in formulary-based prescribing of GPs will continue, stabilize or decline.

The indicator for formulary adherence did have a number of drawbacks similar to other pay-for-performance models [6,10,12–15,41]. At the same time, however, we managed to avoid a number of other problems often associated with pay for performance schemes. First, Roland and Olesen [13] reported that excluding specific patients, with a negative effect on the outcomes, were purposely excluded by providers to improve their scores [13]. In the current pay-for-performance model this was not possible due to the nature of the data collection infrastructure that used routine health data covering the whole patient population. However, for practices with a specific patient population, that were more prone to receive medication outside the formulary, this had a disproportionate negative effect on the score of these practices. To accommodate these practices they could appeal their awarded incentive with their health insurance organization. Second, a distinct hampering effect of the Quality of Framework (QOF) from the United Kingdom was the added administrative burden and increased workload from participating for both GPs and health insurance companies [12,14]. This administrative burden and added workload were kept to a minimum in the pay-for-performance model described in this study: the indicator was based on routine healthcare data in EHRs of general practices, recorded as part of the healthcare process. Furthermore, formularies were already available during prescribing via electronic prescribing systems built into the EHR systems and were already used by GPs. The pay-for-performance scheme could therefore be incorporated into normal practice of GPs realaltively easy.

4.1. Strengths and suggested improvements of the developed indicator and infrastructure

An infrastructure was developed for the measurement of the indicator, involving a large variety of stakeholders, such as EHR suppliers, health insurance companies, GPs, formulary developers and researchers. The involvement of relevant stakeholders, in the set-up of the pay-for-performance scheme is one of the factors named by van Herck et al. [9] in creating an effective pay-for-performance scheme [9]. Furthermore, the collaboration between all involved stakeholders and the created infrastructure may be used for the implementation of future indicators for stimulating formulary-conform prescribing. We will continue to develop the infrastructure with the intent to improve actionability for GPs and improve the fitness for purpose of the re-use of EHR data.

However, the current infrastructure and indicator leave room for improvement. First, the EHR system used by the GP seems to impact the adherence score for formulary-conform prescribing. GPs can choose from a variety of EHR suppliers in the Netherlands. These EHR systems show differences in software, decision support options during prescribing and feedback options. The present study shows differences in the mean scores of GP practices between EHR systems; however, they do not seem to affect the ability of GP practices to improve their score. Opondo et al. also found that the EHR system used by the GP affects prescribing of medication, based on the levels of co-prescription of gastro-protection in combination with NSAIDs [42]. There can be many reasons why practices vary on how they prescribe medication, not necessarily related to prescribing behaviour of the GP, such as how medication is recorded or how the data is extracted from the EHR. Policy makers and health insurance companies should consider these when developing pay-for-performance models based on routine healthcare data and even more so when interpreting the results and assigning the level of financial incentive.

Second, the simultaneous implementation, measurement and application of the indicator in the pay-for-performance scheme in the first year of the project negatively affected the possibilities for feedback on the quarterly scores of formulary-conform prescribing for GPs. Therefore, GPs had limited time to improve their level of appropriate prescribing. According to Roland and Olesen , feedback and public reporting on the outcomes, may have more impact than a financial incentive, especially if the incentive is relatively low [13]. The options to seize this opportunity have not (yet) been exploited enough within the

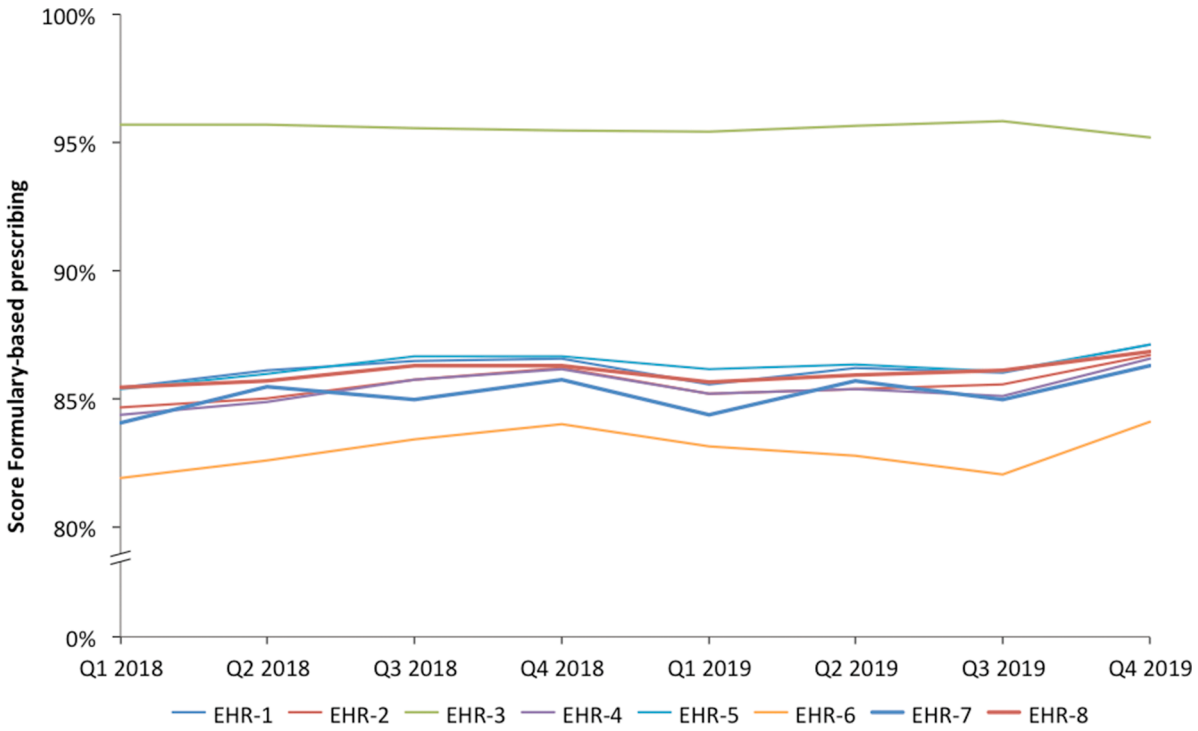

Fig. 3. Mean adherence scores of practices categorized for each EHR system.

presented pay-for-performance scheme. Eriksen et al. found that the improvement in formulary-conform prescribing was stimulated by providing feedback to GPs and education of GPs about the pay-for-performance incentive and its outcomes [1,5]. Although the timeliness of feedback improved during the second year (2019), this was not readily available for all GP practices. Furthermore, the infrastructure of the current study only provides feedback on practice level. Palin et al. favoured feedback on both practice and patient level for an effective use of a LHS infrastructure to optimise appropriate prescribing [28]. When designing a new pay-for-performance model or further developing existing ones, the development of a feedback infrastructure should be a main priority. Using routine healthcare data from EHRs provides opportunities for integrated feedback options (in the EHR of the GP), using their own data. However, besides easy access to feedback also timeliness, the level of feedback (patient vs aggregated at the level of the GP practice) and support in interpreting feedback should be considered.

Third, the indicator used in this study showed high adherence scores from start of implementation, with limited variation between practices and limited room for improvement. In combination with the slow start of feedback, the actionability of the indicator for GPs remained limited. According to literature, a successful pay-for-performance scheme is dependant on a clearly defined outcome (or indicator) that leaves sufficient room for improvement [9,19,20]. Additionally, the specificity of the indicator was limited, because the diagnosis for which the medication was prescribed was not included. As a consequence, prescriptions were labelled as conform formulary only based on ATC-code, even though it was prescribed for a diagnosis not in line with the formulary. New indicators should therefore include further levels of specificity, for example by including the diagnosis in the measurement or the device choice (for example type of inhalers for asthma medication). The development of new indicators will be dependant on the fitness for purpose of the recorded data by the GP from the EHRs, keeping in mind the possible sources of bias [43]. By increasing the level of specificity, the actionability for GPs will increase, possibly the practice variation will become larger and the indicator will show more possibilities for improvement.

Another point of improvement on the developed indicator may be found in the formularies used for testing. The current study and pay-for-performance scheme were limited to two national formularies. Both national and regional formularies are developed in cooperation between GPs, pharmacies and insurance companies, however there is less direct involvement from individual GPs in the national formularies compared to more regional formularies, which might have affected the formulary-conformity. Clinicians are more likely to adhere to guidelines that they developed themselves, due to a greater sense of commitment and ownership [3,4,44]. More than 20 regional and disease specific formularies existed in the Netherlands in 2019. Care groups, who represent a group of GPs, or regional pharmacotherapeutic meetings comprising both GPs and pharmacists, often initiate these regional formularies. These regional formularies contain their own set of recommendations, often resulting in more specific recommendations compared to national formularies. However, they may also contain medication that are not part of national recommendation but do fit more regional needs of the GPs, pharmacists and the own patient population. Erikson et al. contend that GPs are more likely to follow agreements that were made regionally and that suit the local need of GPs and pharmacists, than wider national level agreements that are supposed to address everyone's needs [1]. For future research and use of indicator(s) for the pay-for-performance scheme, regional formularies should be included.

4.2. *Strengths and limitations of the study*

Strength of this study is the inclusion of almost all GP practices in the Netherlands. Furthermore, the developed indicator for formulary-conform prescribing was calculated with prescription information directly derived from the EHR system of the GP practice, instead of the former method, using pharmacy dispensing data. Additionally, data was collected and validated by the researchers and scores were forwarded to the health insurance companies without sharing privacy sensitive information on individual patients. However, a limitation of the study was that only aggregated data on GP practice level was collected. Therefore we were unable to study the effects of certain patient characteristics of GP practice populations on the level of formulary-based prescribing, such the type of practice, region, level of urbanisation, socioeconomic

position or migrant status of the practice population.

5. Conclusion

Aim of this paper was to describe and assess the results and the usefulness of an infrastructure developed to stimulate formulary based prescribing in Dutch general practice. The level of formulary-conform prescribing improved slightly, especially in practices with the lowest scores during the introduction year (2018). However, there should be considered if the invenstments made are justified by the limited potential for improvement. Furthermore, it remains unclear whether the improvement is a direct effect of the incentive, since various factors may have influenced scores (i.e. GPs' recording behaviour in EHR systems). In order to promote formulary-conform prescribing through a pay-for-performance model more effectively, more specific and a wider range of indicators should be developed, including also regional, rather than just national formularies. Also, in order to improve actionability, GP practices should receive more and more timely feedback and future research is needed to determine what type of feedback for GPs may be most effective to stimulate appropriate prescribing.

Furthermore, some factors related to the Dutch healthcare system influenced the establishment of an infrastructure like the one we describe. First, health insurance organizations are not allowed to make price agreements, preventing uniformity in cut-off values and reimbursement policy. Second, the presence of *ten* different EHR suppliers makes it difficult to achieve uniformity of measurements and the provision of feedback, impacting the actionability of the current indicator. These differences may be obstacles for further development of indicators, as they may be more dependant on the structure of the EHR systems involved. In spite of these challenges, we also demonstrated that it was possible to develop an infrastructure and indicator aimed at improving prescribing behaviour, avoiding the privacy risks for individual patients, without increasing the administrative burden in general practice, and by cooperating with organisations representing health insurers, GPs as well as software suppliers. We believe this has been an important step in the development of a learning health system in which data collection and the further use of data is an integral part of the health care delivery process.

Declaration of Competing Interest

None of the authors had a conflict of interest.

Acknowledgments

We would like to thank the "Landelijke Huisartsen Vereniging", the health insurance companies involved (CZ, Menzis, VGZ and Zilveren Kruis) and "Zorgverzekeraars Nederland" for initiating and/or funding the project. Furthermore, we would like to thank the EHR and e-prescribing suppliers, without whom we would not have been able to build the infrastructure. We would also like to acknowledge that data specialist of Nivel for building the database and facilitating the data collection. Last, we would like to acknowledge GPs and pharmacist for their constructive input.

References

[1] Eriksen J, Ovesjö M, Vallin M, Juhasz-haverinen M, Andersén-karlsson E, Ateva K, et al. Primary care physicians report high trust in and usefulness of the Stockholm drug and therapeutic committee ' s list of recommended essential medicines (the ' Wise List '). Eur J Clin Pharmacol 2018;74(131):131–8.
[2] Van dijk L, de jong JD, Westert GP, de bakker DH. Variation in formulary adherence in general practice over time (2003–2007). Fam Pract 2011;28(6):624–31.
[3] Grol R. Successes and failures in the implementation of evidence-based guidelines for clinical practice. Med Care 2001;39(8):II-4–54.
[4] Grol R, Grimshaw J. From best evidence to best practice: effective implementation of change in patients' care. Lancet 2003;362(9391):1225–30.
[5] Eriksen J, Gustafsson LL, Ateva K, Bastholm-Rahmner P, Ovesjö ML, Jirlow M, et al. High adherence to the "Wise List" treatment recommendations in Stockholm: a 15-year retrospective review of a multifaceted approach promoting rational use of medicines. BMJ Open 2017;7(4):1–10.
[6] Vlaanderen FP, Tanke MA, Bloem BR, Faber MJ, Eijkenaar F, Schut FT, et al. Design and effects of outcome-based payment models in healthcare: a systematic review. Eur J Health Econ 2019;20(2):217–32. https://doi.org/10.1007/s10198-018-0989-8 [Internet]Available from:.
[7] Van Dijk CE, Verheij RA, Spreeuwenberg P, Van Den Berg MJ, Groenewegen PP, Braspenning J, et al. Impact of remuneration on guideline adherence: empirical evidence in general practice. Scand J Prim Health Care 2013;31(1):56–63.
[8] Himmelstein DU, Ariely D. Pay-for-performance : toxic to quality ? Insights from behavioral economics. Int J Health Serv 2014;44(2):203–14.
[9] Van Herck P, De SD, Annemans L, Remmen R, Rosenthal MB. Systematic review: effects, design choices, and context of pay-for-performance in health care. BMC Health Serv Res 2010;10(247):1–13.
[10] Mullen K, Frank R, Rosenthal M. Can you get what you pay for? Pay-for-performance and the quality of healthcare providers. Rand J Econ 2010;41(1):64–91.
[11] Allen T, Mason T, Whittaker W. Impacts of pay for performance on the quality of primary care. Risk Manag Healthc Policy 2014;7:113–20.
[12] Gillam SJ, Niroshan Siriwardena A, Steel N. Pay-for-performance in the United Kingdom: impact of the quality and outcomes framework-a systematic review. Ann Fam Med 2012;10(5):461–8.
[13] Roland M, Olesen F. Can pay for performance improve the quality of primary care? BMJ 2016;354(August):1–4. https://doi.org/10.1136/bmj.i4058 [Internet] Available from:.
[14] Peckham S, Wallace A. Pay for performance schemes in primary care: what have we learnt?. PubMed. In: Qual Prim Care 2010:137–46 [Internet][cited 2021 Mar 30]Available from: https://pubmed.ncbi.nlm.nih.gov/20529472/.
[15] Kirschner K, Braspenning J, Akkermans RP, Annelies Jacobs JE, Grol R. Assessment of a pay-for-performance program in primary care designed by target users. Fam Pract 2013;30(2):161–71 [Internet]Apr [cited 2021 Mar 31]Available from: https://pubmed.ncbi.nlm.nih.gov/22997223/.
[16] Ellegård LM, Dietrichson J, Anell A. Can pay-for-performance to primary care providers stimulate appropriate use of antibiotics ? Health Econ 2018;27:39–54.
[17] Khan N, Rudoler D, McDiarmid M, Peckham S. A pay for performance scheme in primary care: meta-synthesis of qualitative studies on the provider experiences of the quality and outcomes framework in the UK. BMC Fam Pract 2020;21(142) [Internet]Jul 13 [cited 2021 Mar 30]Available from: https://pubmed.ncbi.nlm.nih.gov/32660427/.
[18] Rosenthal MB, Dudley RA. Pay-for-performance: will the latest payment trend improve care? J Am Med Assoc 2007;297(7):740–4.
[19] Kirschner K, Braspenning J, Jacobs JEA, Grol R. Design choices made by target users for a pay-for-performance program in primary care: an action research approach. BMC Fam Pract 2012;13(25) [Internet][cited 2021 Mar 31]Available from: https://pubmed.ncbi.nlm.nih.gov/22453028/.
[20] Lester H, Matharu T, Mohammed MA, Lester D, Foskett-Tharby R. Implementation of pay for performance in primary care: a qualitative study 8 years after introduction. Br J Gen Pract 2013;63(611) [Internet]Jun [cited 2021 Mar 30] Available from: https://pubmed.ncbi.nlm.nih.gov/23735412/.
[21] Porter ME. What is value in health care? N Engl J Med 2010;363(26):2477–81.
[22] Porter ME. A strategy for health care reform — toward a value-based system. N Engl J Med 2009;361:109–12.
[23] Menear M, Blanchette MA, Demers-Payette O, Roy D. A framework for value-creating learning health systems. Health Res Policy Syst 2019;17(1):1–13.
[24] Friedman C, Rubin J, Brown J, Buntin M, Corn M, Etheredge L, et al. Toward a science of learning systems: a research agenda for the high-functioning learning health system. J Am Med Inform Assoc 2015;22(1):43–50.
[25] Platt JE, Raj M, Wienroth M. An analysis of the learning health system in its first decade in practice: scoping review. J Med Internet Res 2020;22(3) [Internet] Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7118548/.
[26] Peterson KA, Lipman PD, Lange CJ, Cohen RA, Durako S. Supporting better science in primary care: a description of practice-based research networks (PBRNs) in 2011. J Am Board Fam Med 2012;25(5):565–71 [Internet]Sep [cited 2021 Apr 1] Available from: https://pubmed.ncbi.nlm.nih.gov/22956691/.
[27] Delaney BC, Peterson KA, Speedie S, Taweel A, Arvanitis TN, Hobbs FDR. Envisioning a learning health care system: the electronic primary care research network, a case study. Ann Fam Med 2012;10(1):54–9.
[28] Palin V, Tempest E, Mistry C, Van Staa TP. Developing the infrastructure to support the optimisation of antibiotic prescribing using the learning healthcare system to improve healthcare services in the provision of primary care in England. BMJ Health Care Inform 2020;27(1):1–5.
[29] Urech TH, Woodard LD, Virani SS, Dudley RA, Lutschg MZ, Petersen LA. Calculations of financial incentives for providers in a pay-for-performance program: manual review versus data from structured fields in electronic health records. Med Care 2015;53(10):901–7 [Internet]Oct 1 [cited 2021 Mar 31] Available from: /pmc/articles/PMC4570867/.
[30] Bardach NS, Wang JJ, De Leon SF, Shih SC, Boscardin WJ, Goldman LE, et al. Effect of pay-for-performance incentives on quality of care in small practices with electronic health records: a randomized trial. JAMA J Am Med Assoc 2013;310 (10):1051–9 [Internet].Sep 11 [cited 2021 Apr 1]Available from: https://pubmed.ncbi.nlm.nih.gov/24026600/.
[31] Kruse CS, Stein A, Thomas H, Kaur H. The use of electronic health records to support population health: a systematic review of the literature. J Med Syst 2018;42:214. Springer New York LLC.

[32] Weiner JP, Fowles JB, Chan KS. New paradigms for measuring clinical performance using electronic health records. Int J Qual Health Care 2012;24(3):200–5 [Internet]Jun [cited 2021 Mar 31]Available from: https://pubmed.ncbi.nlm.nih.gov/22490301/.
[33] Kirschner K, Braspenning J, Jacobs JEA, Grol R. Experiences of general practices with a participatory pay-for-performance program: a qualitative study in primary care. Aust J Prim Health 2013;19(2):102–6 [Internet][cited 2021 Mar 30] Available from: https://pubmed.ncbi.nlm.nih.gov/22954169/.
[34] Eijkenaar F. Pay for performance in health care: an international overview of initiatives. Med Care Res Rev 2012;69:251–76. Med Care Res Rev[cited 2021 Mar 31]Available from: https://pubmed.ncbi.nlm.nih.gov/22311954/.
[35] Nivel. Nivel Primary Care Database [Internet]. December 2 2021. [cited 2022 Apr 29]. Available from: https://www.nivel.nl/nl/nivel-zorgregistraties-eerste-lijn/nivel-zorgregistraties-eerste-lijn.
[36] Beardon PH, Brown SV, Mowat DA, Grant JA, McDevitt DG. Introducing a drug formulary to general practice effects on practice prescribing costs. J R Coll Gen Pract 1987;37(July):305–7.
[37] Stewart RE, Kamps GB, Jong BM. Factors influencing adherence to guidelines in general practice. Int J Technol Assess Health Care 2003;19(3):546–54.
[38] Gustafsson LL, Wettermark B, Godman B, Andersøn-karlsson E, Bergman U, Hasselstrçm J, et al. The ‘ Wise List ’ – a comprehensive concept to select, communicate and achieve adherence to recommendations of essential drugs in ambulatory care in Stockholm. Basic Clin Pharmacol Toxicol 2011;108:224–33.
[39] Kamps GB, Stewart RE, Van Der Werf GT, Schuling J, Meyboom-de Jong B. Adherence to the guidelines of a regional formulary. Fam Pract 2000;17(3):254–60.
[40] Herrin J, Nicewander D, Ballard DJ. The effect of health care system administrator pay-for-performance on quality of care. Jt Comm J Qual Patient Saf 2008;34(11):646–54. https://doi.org/10.1016/S1553-7250(08)34082-3 [Internet]Available from:.
[41] Rosenthal MB, Frank RG, Epstein AM. Early experience with pay-for-performance. JAMA 2013;294(14):1788–93.
[42] Opondo D, Visscher S, Eslami S, Verheij RA, Korevaar JC, Abu-Hanna A. Quality of Co-prescribing NSAID and gastroprotective medications for elders in the netherlands and its association with the electronic medical record. PLoS One 2015;10(6):1–10.
[43] Verheij RA, Curcin V, Delaney BC, McGilchrist MM. Possible sources of bias in primary care electronic health record data use and reuse. J Med Internet Res 2018;20(5).
[44] Forsner T, Hansson J, Brommels M, Wistedt AÅ, Forsell Y. Implementing clinical guidelines in psychiatry: a qualitative study of perceived facilitators and barriers. BMC Psychiatry 2010;10(1):8.