BMJ Open Quality

'Give me something meaningful': GPs perspectives on how to improve an audit and feedback report provided by health insurers – an exploratory qualitative study

P J G M de Bekker [1,2] V de Weerdt [1,3] M D H Vink,[1,4] A B van der Kolk,[5] M H Donker,[6] E J E van der Hijden[1,7]

For numbered affiliations see end of article.

Correspondence to
Mr P J G M de Bekker; pietdebekker@zorgvuldigadvies.nl

ABSTRACT

Background Audit and feedback (A&F) is a valuable quality improvement strategy, which can contribute to de-implementation of low-value care. In the Netherlands, all health insurers collaboratively provide A&F to general practitioners (GPs), the 'Primary Care Practice Report' (PCPR). Unfortunately, the use of this report by GPs is limited. This study examined the thoughts of GPs on the usability of the PCPR and GPs recommendations for improving the PCPR.

Method We used an interpretative qualitative design, with think-aloud tasks to uncover thoughts of GPs on the usability of the PCPR and semistructured interview questions to ask GPs' recommendations for improvement of the PCPR. Interviews were audiorecorded and transcribed ad verbatim. Data were analysed using thematic content analysis.

Results We identified two main themes: 'poor usability of the PCPR', and 'minimal motivation to change based on the PCPR'. The GPs found the usability of the PCPR poor due to the feedback not being clinically meaningful, the data not being recent, individual and reliable, the performance comparators offer insufficient guidance to assess clinical performance, the results are not discussed with peers and the definitions and visuals are unclear. The GPs recommended improving these issues. The GPs motivation to change based on the PCPR was minimal.

Conclusions The GPs evaluated the PCPR as poorly usable and were minimally motivated to change. The PCPR seems developed from the perspective of the reports' commissioners, health insurers, and does not meet known criteria for effective A&F design and user-centred design. Importantly, the GPs did state that well-designed feedback could contribute to their motivation to improve clinical performance.

Furthermore, the GPs stated that they receive a multitude of A&F reports, which they hardly use. Thus, we see a need for policy makers to invest in less, but more usable A&F reports.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Audit and feedback (A&F) is an often-used quality improvement strategy, with varying effectiveness.

WHAT THIS STUDY ADDS

⇒ This study strengthens findings on how to design A&F that professionals deem usable and add the perspective of general practitioners on claims data as the source for A&F.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ A call for action to make use of existing knowledge on A&F and user-centred design and invest in less, but better A&F.

strategies across healthcare settings, including primary care.[1] A&F is defined as 'a summary of clinical performance over a specific period of time, and the provision of that summary to individual practitioners, teams or healthcare organisations'.[1] Theory behind A&F is that, by providing clinicians feedback on clinical performance, they are stimulated to change their clinical behaviour and thereby improve clinical practice and patient outcomes.[2–4] While numerous trials show A&F is effective, the effectiveness of A&F interventions varies widely.[1]

There remains uncertainty regarding how to design A&F in such a way it is maximally effective in contributing to quality improvement.[5 6] Research shows that the effectiveness of A&F depends on the design of the intervention itself, recipient characteristics, and the context in which the intervention is performed.[1 7–9] Regarding the design of the intervention evidence shows that A&F is most effective when the 'frequency of feedback is more than once, the format is both written and verbal, feedback is given by a respected

BACKGROUND

Audit and feedback (A&F) is one of the most frequently used quality improvement

peer or colleague and feedback is accompanied by specific goals and an action plan'.[1] [10] Furthermore, A&F may be more effective if it is 'actionable', meaning the feedback is timely, individual and non-punitive.[11–13] Also, research suggests that performance comparators in A&F ideally include trends and specific targets and are compared with similar peers.[14]

Quality improvement strategies are necessary in many Western countries, as they struggle with the provision of low-value care (LVC), that is, care is not effective or not cost-effective.[15] [16] The provision of LVC reduces the overall quality of care, potentially causes patient harm and is ineffective use of scarce healthcare resources. In primary care, LVC entails prescribing of ineffective medications and unnecessary referrals, diagnostics and interventions.[17–19] A&F has been shown to be effective in de-implementing LVC in primary care.[20–22]

In the Netherlands, health insurers collectively commission the development of an A&F report, the 'Primary Care Practice Report' (PCPR)[i], to stimulate general practitioners (GPs) to improve their clinical behaviour. The PCPR presents benchmark information on consultations, diagnostics, interventions and prescriptions in primary care; and numbers of patients using medical specialist care and mental care. Unfortunately, the use of the PCPR by Dutch GPs is limited: only 60% of the GPs have ever downloaded the PCPR between 2015 and 2019 and repeated use of the PCPR is not customary.[ii] While literature gives some theoretical guidance regarding how to design A&F, large uncertainty remains on how to design A&F. The limited use of the PCPR clearly illustrates a gap between theory and how an A&F report is designed in practice. Therefore, we examined the thoughts of GPs on the usability of the PCPR and second, we examined GPs recommendations for improving the PCPR.

METHOD
For reporting of this study, we followed the COnsolidated criteria for REporting Qualitative research, the COREQ-checklist can be found in online supplemental file A.[23]

Design
We used an interpretative qualitative design to explore the perspectives of GPs on the usability of the PCPR.

We used think-aloud tasks to examine the thoughts of GPs on the usability of the PCPR and we used a semi-structured interview to ask GPs for their recommendations on how to improve the usability of the PCPR (online supplemental file B).

In think-aloud methodology, participants are asked to perform a task and verbalise any occurring thoughts while doing so.[24] [25] This methodology is valuable to get a complete and realistic view of users' perspective on tools, such as the PCPR.[26] In each interview, we showed GPs a representative selection of tables and graphs (online supplemental file C) from the PCPR and asked the participating GPs to 'think aloud' while interpreting these tables and graphs.

Patient and public involvement
Due to the nature of the study, patients and public were not involved in the design of this study.

The PCPR
The PCPR is a nationwide benchmark tool for GPs developed since 2015 by VEKTIS in collaboration with health insurers and GPs. VEKTIS is a business intelligence centre that is established and financed by the Dutch Association of Health insurers,[iii] in which all Dutch health insurers are represented.

The PCPR provides GPs insight in the care use of their patients and aims to contribute to GPs' insights in the quality and cost-effectiveness of the care they provide. The PCPR is based on claims data, received by VEKTIS from all Dutch health insurers on insured citizens which covers about 99% of the Dutch population. The feedback includes costs and/or volume indicators on seven main topics: (1) the practice patient population, which includes figures on the volume of the patient population, presented by age, gender and socioeconomic status (SES); (2) the practice total care costs, (3) the GP care provided, which includes cost and volume indicators on consultations, visits and minor surgical procedures performed in primary care; (4) costs of prescribed drugs; (5) referrals to mental healthcare; (6) referrals to medical specialist care and (7) primary care diagnostics. Excluding general descriptive information, the PCPR consists of 24 pages of feedback, with 86 figures and 22 tables.

Feedback is presented for the total GP practice, thus if multiple GPs work in one practice the feedback is on group level. Only one figure in the PCPR presents feedback on individual level, this is for the number of diagnostic requests per year. Furthermore, the feedback in the PCPR is given for a specific year, about 1.5 years prior to the release of the report, the PCPR contains some figures with historical trends, and the performance results of a specific practice are in many figures compared to an 'expected value'. This expected value is based on the national average for a practice with a similar patient population which is case-mix corrected for age, gender, income, pharmaceutical use, and history of Chronic Obstructive Pulmonary Disease (COPD), diabetes, and cardiovascular diseases. This expected value is given as a benchmark to enable GPs to compare their own clinical behaviour. The PCPR is updated once or twice annually, after which GPs are invited via email to freely download the report from the

---

[i]In Dutch, this report is called Praktijkspiegel

[ii]Personal data 2019 from VEKTIS, the Dutch research Centre responsible for the development and distribution of the PCPR

[iii]In Dutch: Zorgverzekeraars Nederland

VEKTIS website. Since 2015, seven versions of the PCPR were made available for download.

Participant recruitment strategy

We purposively recruited GPs that were diverse in respondent characteristics, such as gender, type of practice (health centre, group practice and duo practice) and geographical distribution of the Netherlands. We received contact details of 80 GPs from the medical faculty and via our personal network. GPs related to the medical faculty offer internships to medical students from the 'Vrije Universiteit, in Amsterdam, the Netherlands' and represent a diverse group that work in varying regions of the Netherlands (widely dispersed in the North and Middle region of the Netherlands) and work in various settings (solo practices, duo practices, group practices and health centres). Besides one GP, of whom the interviewer was enlisted in the GP practice, none of the contacted GPs were directly related to the personal network of the authors. Of the 80 GPs that were invited by email or phone for participation in this study, 32 responded to the invitation of which 18 GPs were interested in participation. Fourteen GPs declined; reasons for decline were high work pressure, retirement and acting as a locum. The study aimed to enrol 10–20 participants. This sample size was based on existing think-aloud studies, the likelihood of drop-out and the expected number required to meet data saturation and congruency.[27] Data saturation was reached after 10 interviews, a further two interviews were conducted to increase the rigour of the results. GPs participated voluntarily in the study and did not receive any kind of compensation.

Data collection

All interviews were conducted by a master student (ABvdK, male), with no clinical affiliation. The interviewer knew one GP prior to the interview, as he was enrolled at this GPs practice. He had no personal relationship with this GP or relation to any of the other GPs.

Interviews were held from May to July 2018 in the GP practice of the participant. ABvdK was supervised by two of the authors (PJGMdB, MD), with no clinical affiliations or existing relationships with the participants. All participants provided written informed consent.

Prior to the think-aloud, we asked GPs for descriptive information regarding whether and which types of A&F they used. Then, we used think-aloud tasks to examine the thoughts of GPs on the usability of the PCPR. At the start of each think-aloud session the interviewer briefly explained think-aloud as a method to uncover participants thoughts. The interviewer explained to GPs they would be shown different pages of the PCPR and asked them to reflect on anything they saw or thought when reviewing each page. We asked the GPs to perform the think-aloud task six times, for six different pages of the PCPR, covering 18 out of 86 graphs and 3 out of 22 tables. Since the PCPR does not require an active task but an interpretation and reflection on clinical performance, we used guiding questions to stimulate GPs to think-aloud. Thus, for each page of the PCPR shown, the GPs were asked to answer three subsequent questions (1) What do you see and do you understand what you see? (2) What are your thoughts on what you see? and (3) How does what you see influence you and/or your practice? The six pages for the think-aloud tasks were selected in collaboration with VEKTIS based on diversity in topic of the PCPR, type of healthcare presented and design of graphs. We selected pages that are representative of the total PCPR. The selection included three out of seven topics in the PCPR, namely: (1) patient population: showing the volume of the patient population, presented by age, gender and SES, (2) general practice care: showing volumes of use of care, volumes of various types of consultations and visits, and volumes of minor surgical procedures performed in the GPs practice and (3) general practice diagnostics: showing costs and volumes of diagnostic requests. To assess accuracy of the interpretation of the participant, two researchers discussed the transcripts of the think-aloud tasks and compared results to the legend of the PCPR.

Second, we asked the GPs semistructured interview questions to share their recommendations on how to improve the usability of the PCPR. Lastly, participants were given the opportunity to make additional statements or ask questions.

After the first interview, the wording of some questions was altered to more concrete and accurate wording.

All interviews were audiorecorded. Field notes were made during and immediately after the interviews to aid analysis. We sent a summary of each interview to the respective participant for a member check. None of the participants responded. Interviews lasted 30–55 min. No repeat interviews were carried out.

Data analysis

The interview recordings were transcribed ad verbatim. We used thematic content analysis to analyse the interviews.[27] The analysis was done in two phases. First, a deductive analysis was performed in the context of a master thesis internship, for which two coders independently coded a transcript after each interview was held. The resulting code book was discussed within the wider research team, but since this scheme closely reflected the topic guide it did not fully answer the research questions. To gain deeper understanding in the interviews the research team decided to recode the transcripts inductively, for which initially one reviewer recoded all transcripts. The resulting coding scheme of the inductive approach was briefly discussed in the wider research team to assess the added value of inductive coding. After the research team decided inductive coding was valuable, another researcher (who previously coded deductively) recoded all manuscripts inductively. Subsequently, the two coders discussed the codes per transcript, to ensure consistency in coding. Lastly, the coding scheme was discussed within a wider research team to identify recurrent themes and

Table 1 Respondent characteristics

GP	Sex	Age	Years of clinical experience	Years active in current practice	Type of practice	Work setting
R1	F	52	22	21	Health centre	Urban
R2	M	59	23	22	Group	Urban
R3	F	61	34	32	Health centre	Urban
R4	F	64	36	23	Group	Urban
R5*	M	59	28	27	Group	Rural
R6*	F	59	29	28	Duo	Rural
R7	F	50	17	17	Duo	Urban
R8	M	53	19	19	Duo	Urban
R9	M	62	30	10	Group	Urban
R10	F	35	6	6	Group	Urban
R11†	F	46	16	12	Duo	Rural
R12†	F	40	12	7	Duo	Rural

Type of practice=health centre (centre in which multiple types of healthcare providers reside, such as a physiotherapist, pharmacy, midwife, social worker, dietician, psychologist), group practice (GP practice owned by >2 GPs) or duo practice (practice jointly owned by two GPs).
Work setting=urban (GP practice in a city); rural (GP practice outside a city).
*R5 and R6 worked in the same practice.
†R11 and R12 worked in the same practice.
GP, general practitioner.

formulate the final themes. In the themes, we grouped results from the think-aloud tasks and recommendations, since these two approaches identified similar themes. Think-aloud identified negatively framed items, while recommendations identified similar items framed positively.

All coding was performed using qualitative computer analysis software package, Atlas.ti (V.9.1.3) software. The quotes that best illustrated the different themes and subthemes were selected and translated from Dutch to English. All quotes were then evaluated by the remaining authors and changes were made to reach agreement among all authors.

RESULTS
Participant characteristics
We interviewed 12 GPs, from 10 different practices, aged 35–64 years. The respondents varied in years of work experience, active years in current practice and digital experience, type of practice and work setting (table 1). All GPs were familiar with A&F reports in some form, although most GPs did not regularly and systematically used A&F to review their own performance. Respondents receive many types of A&F reports, such as reports from their own GP information system, accreditation programmes, a recognised health services research institute, health groups or regional networks, individual health insurers, pharmacotherapy networks and diagnostic laboratories. Eight GPs (67%) had heard of the PCPR before the invitation, but only three GPs (25%) had tried to use the PCPR for quality improvement purposes.

Themes
We identified two main themes from the interviews: 'poor usability of the PCPR', and 'minimal motivation to change based on the PCPR'. In the theme 'poor usability of the PCPR', we identified the subthemes 'desire for clinically meaningful feedback'; 'limitations of claims data for audit'; 'performance comparators are difficult to interpret'; 'desire to discuss results with peers' and 'unclear definitions and figures' (online supplemental file D). The second theme did not have subthemes. Recommendations for improvement of the PCPR, if provided from the interviews, are described per theme.

Poor usability of the PCPR
Desire for clinically meaningful feedback
The GPs stated that they did not find the feedback clinically meaningful, since feedback on quality of care is lacking, cost indicators reflect prices which GPs cannot influence and detail on conditions is lacking. The GPs did not see how the feedback shown coud help them to improve their clinical practice.

R7: "It is unclear to me how these figures might help the patient sitting at the other side of my desk."

R11: "We strive for high quality healthcare, but feedback usually reflects euros or percentages, instead of health gains. For how many people suffering from high blood pressure did my prescription result in a lower blood pressure? And how many of my patients have not had a stroke as a consequence? How do I relate to peers?"

The GPs found the cost indicators uninteresting and frustrating, since these include variation in prices or costs made in secondary care, which both cannot be influenced by the GP. Also, GPs found the feedback on high level, in which all types of conditions are combined, not clinically meaningful.

GPs recommended including quality indicators on either clinical outcomes regarding their treatment role, such as the blood pressure of their patients, or indicators on the effectiveness of their gatekeeper role, such as whether longer GP consultations lead to less referrals to secondary care. Furthermore, GPs recommended excluding feedback outside their control, such as costs influenced by prices. Lastly, GPs recommended including feedback with lower detail level, such as for common disorders.

R1: "I would like to see more figures on my referrals for separate conditions. And this information is mostly cost-related, so price increases have an effect. However, I can't control those costs."

Limitations of claims data as a source for A&F

The GPs found that the claims data on which the PCPR is based has several important limitations. First, half of the GPs indicated that the timeliness of the data was insufficient, with data dating back to 2 years prior to the report release date.

R11: "Drawback is that it is of course much too late, in the sense that I'm looking at numbers of 2016 (in 2018). Thus, it is not exactly management information. I can only see that I am at the bottom of the canyon, because I missed an exit in 2016"

R4: "Anyway this feedback is of over two years ago, I find that difficult. The longer ago, the more difficult it is to apply in practice. I would appreciate feedback on my figures from last year somewhere within the first three to four months of the current year. That would really help me."

Second, the GPs stated that the group level feedback makes it unclear whether deviations in performance are attributable to themselves or to a colleague. Even for the only figure with individual feedback, GPs stated that this may not be attributable to themselves, since diagnostic requests have been made in their name by physician assistants, residents GPs and part-time GPs.

Third, the GPs questioned the reliability and validity of the data.

R1: "Then my thoughts are: is it correct? Is it correct for our population?"

R5: "This seems to be untrue; I do not know how the benchmark is developed; this difference is too big. So, I have my reservations."

Furthermore, two GPs expressed a lack of trust in the source of the PCPR, they were sceptical about the role of health insurance companies as the supplier of data.

GPs recommended using recent, individual, reliable, and valid data as the source for feedback.

Performance comparators offer insufficient guidance to assess clinical performance

GPs found that the historical trends, the 'expected value' as the benchmark and the lack of targets in the PCPR offered insufficient guidance to assess their clinical performance.

Futhermore, GPs noticed unstable trends in their performance, with high outliers in one year and low outliers in the next. This caused confusion and GPs did not know how to interpret this or how to act on these unstable trends. Also, GPs indicated that policy changes in healthcare organisation or claim codes can make data incomparable over time, hindering interpretation of historical trends.

R5: "If one year I do too much and the other year I am below expected, then yeah, I think yeah, I really won't do my work differently."

Last, GPs found it difficult to compare their clinical performance to the 'expected value', since this value is derived from national data and thus includes incomparable practices. Eight of the 12 GPs attributed deviations in their clinical performance to casemix differences, while the figures were already case-mix corrected. In addition, several GPs missed specific targets in the PCPR.

R2: "Are a hundred colleagues providing too many treatments or am I providing too few? Hmmm, who knows?"

GPs recommended comparing performance to similar peers, rather than national data. GPs would like to compare to peers within their practice as well as peers from other practices. GPs found it important that the comparator is similar in setting (rural versus urban) and type of care services (whether an ultrasound is available, a Doppler, etc). Also, GPs recommended including regional comparisons and to account for integrated networks. Furthermore, GPs recommended including an advice or an action plan in the feedback, which helps them assess their clinical performance.

Desire to discuss results with peers

Based on the feedback, multiple GPs expressed they would like to discuss this feedback with peers from their own practice or peers from other practices. GPs wanted to discuss the causes of deviations in audit results with peers, to gain insights in other ways of organising and delivering care and thus on if and how to change their clinical behaviour.

Unclear definitions and figures

Every GP found one or more tables/graphs in the PCPR hard to interpret. The GPs stated that several definitions were unclear, such as 'primary care diagnostics', 'request for diagnostics'. GPs did not know how these indicators were defined and what classified under such categories.

> R12: "I also don't really know what primary care diagnostics entails, is that only blood tests or does it include more?"

> R11: "I don't even know about what kind of diagnostics this is now: special diagnostics, scopies, no clue what this is about."

GPs recommended including more explanations on definitions and case-mix correction in the tables/graphs instead of only in the preface of the report. Furthermore, GPs found there was an unclear presentation of numbers with varying scales in similar graphs, unclear units and indistinctive colours.

Minimal motivation for change based on the PCPR

Most GPs found that the PCPR did not provide the correct information to change, due to the content of the feedback, the type of audit data that was used or the lack of guidance of performance comparators.

> R5: "If you are higher than expected at one year and lower at the year after, then I think: 'the average seems good, there's no reason to work differently'. If I have a consultation with a patient who is coughing for half a year, I do not think 'oh, I've applied too much photos so far, let's not refer this patient.' You still look at the patient and his needs."

There were only a few cases in which GPs stated they were motivated to change their behaviour based on the PCPR. They were either motivated to change clinical behaviour, such as increasing intensive visits and increasing Dopplers, or motivated to change claims behaviour.

> R4: "This number strikes me. I think this is due to under registration on our part."

> R11: "We perform less 24hours blood pressure measurements than average. That is good to know, because then we will perform those a bit more, because they earn well. For Dopplers the same. Those we can nicely increase a bit."

> R6: "We should look at our administrative process. Probably we sometimes forget to submit declarations. That is no medical difference, but purely administrative."

Interestingly, most small deviations of performance did not give GPs motivation to change, while large deviations resulted in GPs questioning the data thus also did not evoke motivation to change. Also, many deviations in performance were explained by GPs as conscious practice management.

> R1: "ECG we indeed don't perform often, we only do this for rhythmic disorders, and we do that on purpose, we have them done in the hospitals because we believe our quality in assessing ECGs is inadequate."

> R4: "Well, as I said we perform less phone consultations, we agreed not to perform triage, that is conscious policy and we are very happy with that."

Importantly, the majority of the GPs stated that well designed A&F could potentially motivate them to change behaviour and thus have a meaningful impact on quality of care. Two GPs stated that A&F in general could not change their behaviour. Both GPs stated that this was because aggregated data could not influence how they provide care for an individual patient, one GP states this was because he applies critical thinking continuously and that he would find it difficult to weigh aggregated data more than his own critical thinking and the other that he provides care one patient at a time and that you do want your actions to be 'medically responsible', thus that aggregated data does not influence patient care.

DISCUSSION

This study examined the thoughts of GPs on the usability of the PCPR and GPs recommendations for improving the PCPR. GPs found the usability of the PCPR poor due to (1) the feedback not being clinically meaningful; (2) the audit data not being recent, individual and reliable; (3) the performance comparators offered insufficient guidance to assess clinical performance; and (4) the results are not discussed with peers and the definitions and visuals are unclear. GPs recommend improving these issues. The poor usability of the report led to minimal motivation to change based on the PCPR.

Our results are consistent with previous literature on effective A&F and user-centred design; and indicate that policy makers still do not draw sufficiently from existing evidence when developing A&F reports.[13 28–31] The PCPR contains information on process and cost indicators aggregated across all types of conditions, based on the healthcare procurement perspective of health insurers. However, this does not match the clinical perspective of the GPs. In our study GPs requested information on clinical outcomes regarding their treatment role and effectiveness of their gatekeeper role, which can be developed based on claims data. Also, the PCPR could present feedback for common disorders and common interventions specifically instead of aggregated data. Aligning feedback with recipient priorities is considered key to improve the effectiveness of A&F.[32] Furthermore, the PCPR does not specifically focus on LVC and lacks targets, which makes it difficult for GPs to interpret whether and how they can improve their quality of care. Since claims are not identifiable to one GP and claim processing has a lag-time, it is difficult to meet requirements of recent and individual A&F based on claims data.

Our study also adds to the knowledge on how to design effective A&F. First, GPs specifically asked for feedback on clinical outcomes and the effectiveness of their gatekeeper role, such as whether more long consultations prevent referrals or whether more email consultations in primary care prevent face-to-face consultations in primary care. It is interesting that GPs in our study requested clinical outcomes as feedback, even though outcomes are not a direct reflection of clinical performance.[11] GPs in our study stated that outcome indicators, such as irregulated blood pressures, are important indicators for them to review their clinical behaviour, such as whether they should change their prescriptions or referral behaviour. Second, although using trends as a performance comparator is recommended since trends can create positive motivation towards change, we found that unstable trends could cause confusion in interpreting the feedback.[14] To help clinicians interpret trends we advise to include descriptions of relevant policy changes or possible changes of claims code definitions in the caption of feedback. Third, we found that using claims data as the source for A&F in some cases motivated GPs to change claims behaviour instead of clinical behaviour. Since it is the aim of A&F to stimulate reflection on clinical behaviour, it is questionable whether using claims data as the source for feedback distracts from this purpose and whether patient record data or medical registries are better suited data sources. However, it is unknown whether the type of data source used for A&F influences the effectiveness of A&F interventions.[33] Lastly, some GPs were sceptical about the role of health insurance companies as the supplier of the data. Health insurers in the Netherlands have the legal task of guarding affordability, accessibility, and efficiency of healthcare. Therefore, professionals express a fear that health insurers will use audit data to reduce healthcare reimbursements. Using A&F in a punitive manner might lower professionals' motivation to change and thus result in less effective A&F interventions.[11]

A comparable qualitative study, which examined how the PCPR and group discussions contribute to GPs motivation to change, found many similar results, such as that feedback should be reliable, recent, individual and concern the own influence sphere.[34] Furthermore, this study found that group discussions on the PCPR contributed to GPs motivation to change. It could be that not the quality of the feedback itself, but feedback combined with peer discussion is a vital strategy in A&F effectiveness.[34 35] However, studies examining the effectiveness of peer-discussion in A&F interventions and quality improvement show varying results.[1 36–38]

Theoretical underpinning of using A&F for quality improvement purposes is that A&F motivates professionals to change.[2 4] In our study, most GPs stated they were a priori motivated to change to improve quality and that A&F could contribute to their motivation to change. However, the GPs stated that they find the quality of the PCPR insufficient, thus it did not motivate them to change. In A&F literature lack of quality of feedback is frequently mentioned as a barrier for motivation to change.[39] It is possible that lack of quality is mentioned as a barrier for change, because of cognitive dissonance or external attribution, in which professionals search for an external explanation for underperformance.[14 40] Experts state that the optimal design of A&F likely depends on recipient factors and context, but it is yet unclear which thresholds exists for quality of feedback to be 'sufficient' to contribute to motivation to change[6] Furthermore, Desveaux et al[35] describe that 'A&F may not be translated into actual change, due to incapability of participants to interpret A&F in an actionable way'. Thus, even when the quality of the feedback itself is sufficient, recipients may still need education on how to interpret and translate A&F into change.[41]

Multiple studies throughout the years describe that GPs find A&F reports unusable, while these same studies do find that GPs are motivated to use A&F for quality improvement.[41–43] In addition in our study, GPs state that they receive a multitude of A&F reports, many of which they hardly use. Using low-quality A&F creates waste of resources and leads to ineffective use of clinician's time. While the PCPR has important strengths, such as support from national policy makers, national coverage, it covers all healthcare provided and it is based on routinely collected data, not all concerns of GPs regarding usability of the PCPR can be solved, such as the timeliness of the data or individual feedback. Thus, it might be more effective to invest in another type of A&F for quality improvement. For more effective quality improvement, we see a need for policy makers to invest in less, but more usable A&F reports.[44]

Strengths and limitations

An important strength of this study was the think-aloud methodology, which enabled us to capture direct reactions to specific issues of the PCPR, instead of capturing general reflections on the PCPR or A&F. Hereby, we were able to identify items that explain the limited uptake and influence the usage of the PCPR. Another strength is that we examined an existing A&F report which is routinely available.

This study also had several limitations. First, we had a relatively limited recruitment strategy. While the group of GPs was diverse in several characteristics, they were relatively old which may have caused bias. Second, the interviewer was enlisted at the GP practice of one respondent, but since this prior relationship was limited, we do not believe this relationship has altered our results. Third, some of the respondents were familiar with the PCPR prior to the interview, which may have affected their responses to the think-aloud tasks. Lastly, since we did not work iteratively and completely independent during the second analysis phase this might have caused bias in the interpretation of our results. However, we tried to overcome this bias by discussing codes per transcript and discussing themes within the wider research team and since another study regarding the PCPR found similar results, we believe that our results are reliable.

CONCLUSIONS

The GPs in our study evaluated the PCPR as poorly usable and were minimally motivated to change behaviour based

on the PCPR. The PCPR seems developed from the perspective of the commissioner, the health insurer, and does not meet known criteria for effective A&F design and user-centred design. Importantly, the GPs did state that well-designed feedback would contribute to their motivation to improve clinical performance.

Furthermore, GPs in our study stated that they receive a multitude of A&F reports, which they hardly use. For more effective quality improvement, we see a need for policy makers to invest in less, but more usable A&F reports.

Author affiliations
[1]Department of Health Economics & Talma Institute, Vrije Universiteit Amsterdam, Amsterdam, Netherlands
[2]Zorgvuldig Advies, Utrecht, Netherlands
[3]Amsterdam University Medical Centres, Holendrecht, Netherlands
[4]Gynaecology, Amsterdam Universitair Medische Centra, Duivendrecht, Netherlands
[5]Talma Institute, Vrije Universiteit Amsterdam, Amsterdam, Netherlands
[6]Department of Health Sciences, Vrije Universiteit Amsterdam, Amsterdam, Netherlands
[7]Zilveren Kruis Health Insurance, Zeist, Netherlands

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Ethics approval According to the regulations of the BETHCIE, this study did not require full ethical review. Participants gave informed consent to participate in the study before taking part.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available on reasonable request.

ORCID iDs
P J G M de Bekker http://orcid.org/0000-0001-8560-9103
V de Weerdt http://orcid.org/0000-0003-3644-620X

REFERENCES
1 Ivers N, Jamtvedt G, Flottorp S, *et al*. Audit and feedback : effects on professional practice and healthcare outcomes (Review). *Cochrane Database of Systematic Review* 2012;6:1–217.
2 Kluger AN, DeNisi A. The effects of feedback interventions on performance: a historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychol Bull* 1996;119:254–84.
3 Gardner B, Whittington C, McAteer J, *et al*. Using theory to synthesise evidence from behaviour change interventions: the example of audit and feedback. *Soc Sci Med* 2010;70:1618–25.
4 Brown B, Gude WT, Blakeman T, *et al*. Clinical performance feedback intervention theory (CP-FIT): a new theory for designing, implementing, and evaluating feedback in health care based on a systematic review and meta-synthesis of qualitative research. *Implementation Science* 2019;14:1–25.
5 Foy R, Eccles MP, Jamtvedt G, *et al*. What do we know about how to do audit and feedback? pitfalls in applying evidence from a systematic review. *BMC Health Serv Res* 2005;5:1–7.
6 Ivers NM, Sales A, Colquhoun H, *et al*. No more 'business as usual' with audit and feedback interventions: towards an agenda for a reinvigorated intervention. *Implementation Science* 2014;9:1–8.
7 Ivers NM, Grimshaw JM, Jamtvedt G, *et al*. Growing literature, stagnant science? systematic review, meta-regression and cumulative analysis of audit and feedback interventions in health care. *J Gen Intern Med* 2014;29:1534–41.
8 Brown B, Balatsoukas P, Williams R, *et al*. Multi-method laboratory user evaluation of an actionable clinical performance information system: implications for usability and patient safety. *J Biomed Inform* 2018;77:62–80.
9 Brehaut JC, Colquhoun HL, Eva KW, *et al*. Practice feedback interventions: 15 suggestions for optimizing effectiveness. *Ann Intern Med* 2016;164:435–41.
10 Hysong SJ. Meta-Analysis: audit and feedback features impact effectiveness on care quality. *Med Care* 2009;47:356–63.
11 Hysong SJ, Best RG, Pugh JA. Audit and feedback and clinical practice guideline adherence: making feedback actionable. *Implement Sci* 2006;1:1–10.
12 Larson EL, Patel SJ, Evans D, *et al*. Feedback as a strategy to change behaviour: the devil is in the details. *J Eval Clin Pract* 2013;19:230–4.
13 Payne VL, Hysong SJ. Model depicting aspects of audit and feedback that impact physicians' acceptance of clinical performance feedback. *BMC Health Serv Res* 2016;16:1–12.
14 Gude WT, Brown B, van der Veer SN, *et al*. Clinical performance comparators in audit and feedback: a review of theory and evidence. *Implementation Science* 2019;14:1–14.
15 Mafi JN, Parchman M. Low-Value care: an intractable global problem with no quick fix. *BMJ Qual Saf* 2018;27:333–6.
16 Sypes EE, de Grood C, Clement FM, *et al*. Understanding the public's role in reducing low-value care: a scoping review. *Implementation Science* 2020;15:1–13.
17 Müskens JL, van Dulmen SA, Wiersma T, *et al*. Low-Value pharmaceutical care among Dutch GPs: a retrospective cohort study. *Br J Gen Pract* 2022;72:e369–77. Feb 10;BJGP.2021.0625.
18 Kool RB, Verkerk EW, Meijs J, *et al*. Assessing volume and variation of low-value care practices in the Netherlands. *Eur J Public Health* 2020;30:236–40.
19 Pendrith C, Bhatia M, Ivers NM, *et al*. Frequency of and variation in low-value care in primary care: a retrospective cohort study. *CMAJ Open* 2017;5:E45–51.
20 Kerry Influence of postal distribution of The Royal College of Radiologists' guidelines, together with feedback on radiological referral rates, on X-ray referrals from general practice- a randomized controlled trial.
21 Gerber JS, Prasad PA, Fiks AG, *et al*. Effect of an Outpatient Antimicrobial Stewardship Intervention on Broad-Spectrum Antibiotic Prescribing by Primary Care Pediatricians A Randomized Trial [Internet]. Available: www.jama.com
22 WHJM V, van der Weijden T, Sijbrandij J, *et al*. Effect of a Practice-Based Strategy on Test Ordering Performance of Primary Care Physicians A Randomized Trial [Internet]. Available: http://jama.jamanetwork.com/
23 Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. *Int J Qual Health Care* 2007;19:349–57.
24 van Someren MaartenW. The think aloud method. Vol. 31, Information Processing & Management 1994.
25 Lewis C. *Using the "thinking-aloud" method in cognitive interface design*. Yorktown Heights: IBM TJ Watson Research Center, 1982.
26 de Lusignan S, Liyanage H, Sherlock J, *et al*. Atrial fibrillation dashboard evaluation using the think aloud protocol. *BMJ Health Care Inform* 2020;27:e100191–7.
27 Sullivan GM, Sargeant J. Qualities of qualitative research: Part I. *J Grad Med Educ* 2011;3:449–52.
28 Colquhoun HL, Sattler D, Chan C, *et al*. Applying User-Centered design to develop an audit and feedback intervention for the home care sector. *Home Health Care Manag Pract* 2017;29:148–60.

29 Landis-Lewis Z, Kononowech J, Scott WJ, *et al*. Designing clinical practice feedback reports: three steps illustrated in veterans health affairs long-term care facilities and programs. *Implementation Science* 2020;15:1–13.

30 Colquhoun H, Michie S, Sales A. Reporting and design elements of audit and feedback interventions: a secondary review. vol. 26, BMJ quality and safety. *BMJ Publishing Group* 2017:54–60.

31 Foy R, Skrypak M, Alderson S, *et al*. Revitalising audit and feedback to improve patient care. *BMJ* 2020;368:m213.

32 Colquhoun HL, Carroll K, Eva KW, *et al*. Informing the research agenda for optimizing audit and feedback interventions: results of a prioritization exercise. *BMC Med Res Methodol* 2021;21:20.

33 van der Veer SN, de Keizer NF, Ravelli ACJ, *et al*. Improving quality of care. A systematic review on how medical registries provide information feedback to health care providers. *Int J Med Inform* 2010;79:305–23.

34 Cooke LJ, Duncan D, Rivera L, *et al*. How do physicians behave when they participate in audit and feedback activities in a group with their Peers? *Implementation Science* 2018;13.

35 Desveaux L, Ivers NM, Devotta K, *et al*. Unpacking the intention to action gap: a qualitative study understanding how physicians engage with audit and feedback. *Implementation Science* 2021;16:1–9.

36 Verstappen WHJM, ter Riet G, Dubois WI, *et al*. Variation in test ordering behaviour of GPs: professional or context-related factors? *Fam Pract* 2004;21:387–95.

37 Trietsch J, van Steenkiste B, Grol R, *et al*. Effect of audit and feedback with peer review on general practitioners' prescribing and test ordering performance: a cluster-randomized controlled trial. *BMC Fam Pract* 2017;18:1–13.

38 Rohrbasser A, Harris J, Mickan S, *et al*. Quality circles for quality improvement in primary health care: Their effectiveness, gaps of knowledge, origins and significance - a scoping review. *bioRxiv.*2018:1–19.

39 Ivers N, Jamtvedt G, Flottorp S, *et al*. Audit and feedback: effects on professional practice and healthcare outcomes. *Cochrane Database Syst Rev* 2012;2012:CD000259.

40 van Braak M, Visser M, Holtrop M, *et al*. What motivates general practitioners to change practice behaviour? A qualitative study of audit and feedback group sessions in Dutch general practice. *BMJ Open* 2019;9:e025286–8.

41 Rouleau G, Reis C, Ivers NM, *et al*. Lipstick on a pig : Understanding efforts to redesign audit and feedback reports for primary care 2022:1–16.

42 Ivers N, Barnsley J, Upshur R, *et al*. "My approach to this job is one person at a time": Perceived discordance between population-level quality targets and patient-centred care. *Can Fam Physician* 2014;60:258–66.

43 Arvidsson E, Dahlin S, Anell A. Conditions and barriers for quality improvement work: a qualitative study of how professionals and health centre managers experience audit and feedback practices in Swedish primary care. *BMC Fam Pract* 2021;22:1–13.

44 Ivers NM, Desveaux L. De-implementation of low-value care: use audit and feedback wisely. *Healthcare Papers* 2019;18:41–7.

COREQ (COnsolidated criteria for REporting Qualitative research) Checklist

A checklist of items that should be included in reports of qualitative research. You must report the page number in your manuscript where you consider each of the items listed in this checklist. If you have not included this information, either revise your manuscript accordingly before submitting or note N/A.

Topic	Item No.	Guide Questions/Description	Reported on Page No.
Domain 1: Research team and reflexivity			7
Personal characteristics			7
Interviewer/facilitator	1	Which author/s conducted the interview or focus group?	7
Credentials	2	What were the researcher's credentials? E.g. PhD, MD	7
Occupation	3	What was their occupation at the time of the study?	7
Gender	4	Was the researcher male or female?	7
Experience and training	5	What experience or training did the researcher have?	
Relationship with participants			7
Relationship established	6	Was a relationship established prior to study commencement?	7
Participant knowledge of the interviewer	7	What did the participants know about the researcher? e.g. personal goals, reasons for doing the research	
Interviewer characteristics	8	What characteristics were reported about the inter viewer/facilitator? e.g. Bias, assumptions, reasons and interests in the research topic	17
Domain 2: Study design			
Theoretical framework			
Methodological orientation and Theory	9	What methodological orientation was stated to underpin the study? e.g. grounded theory, discourse analysis, ethnography, phenomenology, content analysis	8
Participant selection			
Sampling	10	How were participants selected? e.g. purposive, convenience, consecutive, snowball	6
Method of approach	11	How were participants approached? e.g. face-to-face, telephone, mail, email	6
Sample size	12	How many participants were in the study?	
Non-participation	13	How many people refused to participate or dropped out? Reasons?	9
Setting			7
Setting of data collection	14	Where was the data collected? e.g. home, clinic, workplace	
Presence of non-participants	15	Was anyone else present besides the participants and researchers?	7
Description of sample	16	What are the important characteristics of the sample? e.g. demographic data, date	7
Data collection			9-10
Interview guide	17	Were questions, prompts, guides provided by the authors? Was it pilot tested?	
Repeat interviews	18	Were repeat inter views carried out? If yes, how many?	7-8
Audio/visual recording	19	Did the research use audio or visual recording to collect the data?	
Field notes	20	Were field notes made during and/or after the inter view or focus group?	8
Duration	21	What was the duration of the inter views or focus group?	7
Data saturation	22	Was data saturation discussed?	8
Transcripts returned	23	Were transcripts returned to participants for comment and/or	8
			7
			8

Supplemental material

BMJ Publishing Group Limited (BMJ) disclaims all liability and responsibility arising from any reliance placed on this supplemental material which has been supplied by the author(s)

BMJ Open Qual

Topic	Item No.	Guide Questions/Description	Reported on Page No.
		correction?	
Domain 3: analysis and findings			8
Data analysis			yes
Number of data coders	24	How many data coders coded the data?	
Description of the coding tree	25	Did authors provide a description of the coding tree?	8-9
			9
Derivation of themes	26	Were themes identified in advance or derived from the data?	8
Software	27	What software, if applicable, was used to manage the data?	
Participant checking	28	Did participants provide feedback on the findings?	
Reporting			8-13
Quotations presented	29	Were participant quotations presented to illustrate the themes/findings? Was each quotation identified? e.g. participant number	8-13
Data and findings consistent	30	Was there consistency between the data presented and the findings?	8-13
Clarity of major themes	31	Were major themes clearly presented in the findings?	8-13
Clarity of minor themes	32	Is there a description of diverse cases or discussion of minor themes?	

Developed from: Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. *International Journal for Quality in Health Care*. 2007. Volume 19, Number 6: pp. 349 – 357

Once you have completed this checklist, please save a copy and upload it as part of your submission. DO NOT include this checklist as part of the main manuscript document. It must be uploaded as a separate file.

Supplemental File B – Interview Topic Guide

1. Descriptive Information
- Information used to gain insight in clinical practice
- Feedback used to gain insight in quality of care

2. 'Thinking aloud' method
- What do you see and do you understand what you see?
- What are your thoughts on what you see?
- How does what you see influence you and/or your practice?

3. Semi-structured interview
- Opinion on PCPR
- Recommendations to improve the usability of the PCPR

PCPR 2.3

3. Patient population

3.1 Number of patients registered in practice
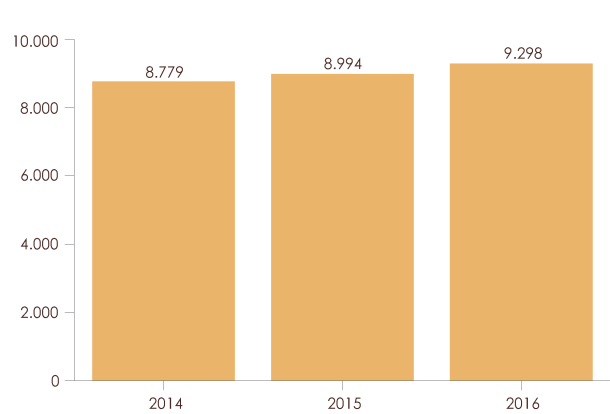
3.2 Percentage of registered patients by income level based on zip-code (2016)
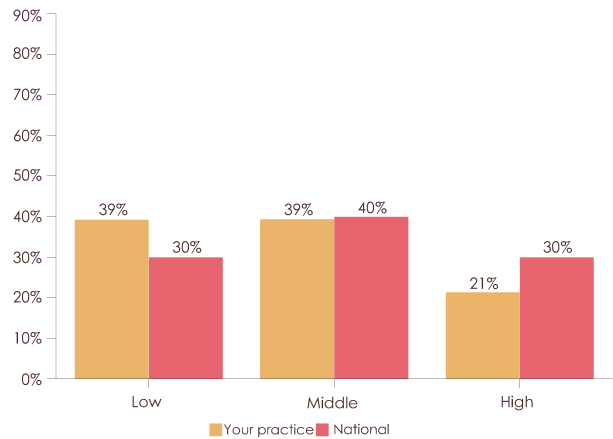
3.3 Percentage of registered patients by age and gender (2016)
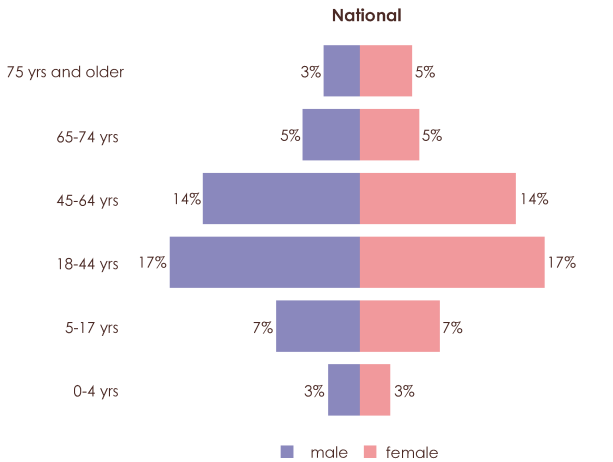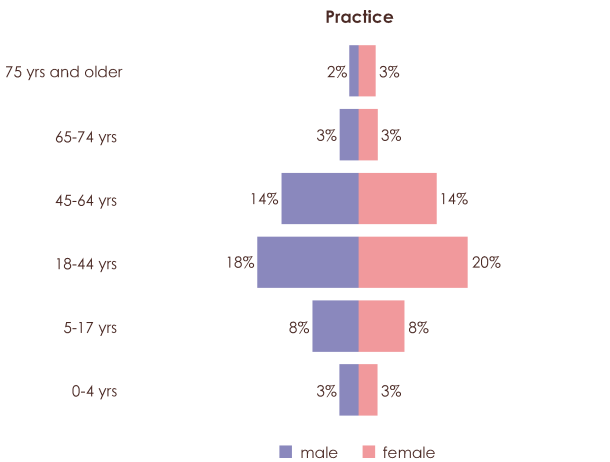
Practice

National

Number of patients	0-4 yrs	5-17 yrs	18-44 yrs	45-64 yrs	65-74 yrs	75 yrs and older
male	317	763	1.684	1.280	311	152
female	312	763	1.814	1.299	316	284

Distribution by gender	% Gender practice	% Gender national
male	48%	50%
female	52%	50%

* Source: "CBS: Key figures zipcode areas 2008-2010". See appendix for more information.

PCPR2.3

5.3 Breakdown of registered patients by type of healthcare user (2016)

		0-4 yrs	5-17 yrs	18-44 yrs	45-64 yrs	65-74 yrs	75 yrs and older

□ None (0 contacts)   □ Incidental (1-2 contacts)   ■ Structural (3-8 contacts)   ■ Intensive (9 or more contacts)

PCPR2.3

Reference date: 01-07-2016
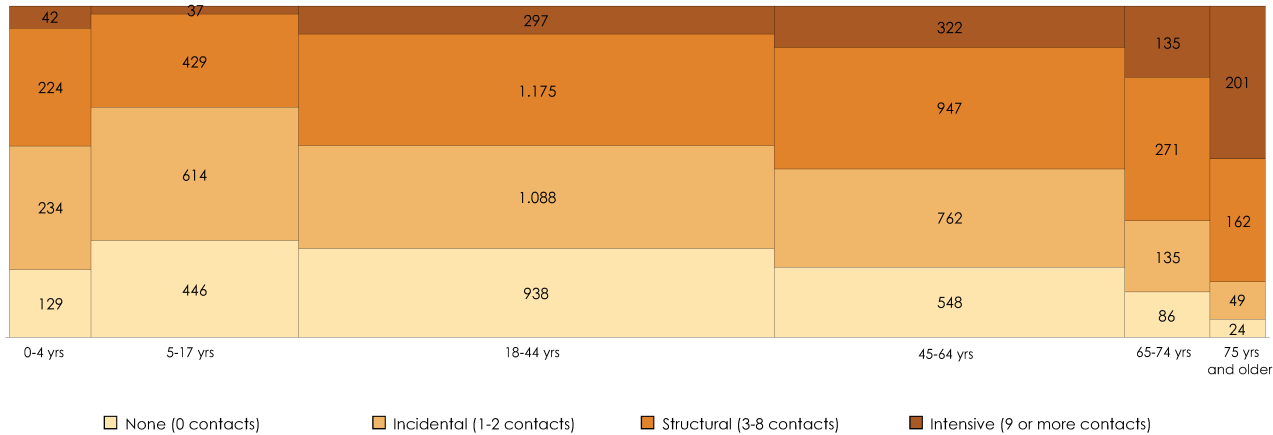
5.4 Average number of consultations and visits per registered patient (2016)

Consultations

Deviation	0-4 yrs	5-17 yrs	18-44 yrs	45-64 yrs	65-74 yrs	75 yrs and older
%	-3%	-3%	2%	-3%	-8%	22%

consultations > 20 minutes

Deviation	0-4 yrs	5-17 yrs	18-44 yrs	45-64 yrs	65-74 yrs	75 yrs and older
%	-40%	-30%	-7%	-11%	-9%	15%

visit

Deviation	0-4 yrs	5-17 yrs	18-44 yrs	45-64 yrs	65-74 yrs	75 yrs and older
%	-61%	36%	-64%	-38%	-4%	-21%

visit > 20 minutes

Deviation	0-4 j yrs	5-17 yrs	18-44 yrs	45-64 yrs	65-74 yrs	75 yrs and older
%	-100%	-100%	-43%	-48%	-49%	-48%

Phone consultations

Deviations	0-4 yrs	5-17 yrs	18-44 yrs	45-64 yrs	65-74 yrs	75 yrs and older
%	-16%	-19%	-10%	-15%	3%	-8%

E-mailconsultations

Deviations	0-4 yrs	5-17 yrs	18-44 yrs	45-64 yrs	65-74 yrs	75 yrs and older
%	-100%	-85%	-60%	-92%	-28%	6%

PCPR 2.3

Reference date: 01-07-2016

5.5 Overview percentage deviation number of claimed M&I interventions in your pratice vs. expected (2016)

For the 10 nationally most common claimed interventions

Intervention	Deviation
13042 - IUD placement or removal implanon	13%
13001 - Diagnostics with doppler	-8%
13008 - 24-hrs blood pressure measurement	-19%
13012 - Small surgery	-25%
13004 - Longfunction measurement	-32%
13023 - Therapeutic injection	-38%
13005 - ECG-diagnostics	-71%
13034 - Intensive care, day	-84%
13043 - Structured GPcare, nursing home	-100%
13036 - Intensive care, extended acces, evenings and weekends	-100%

Reference date: 01-07-2016

11. Primary care diagnostics

11.1 Average costs primary care diagnostics per registered patient

Deviation	2014	2015	2016
%	-47%	-25%	-11%
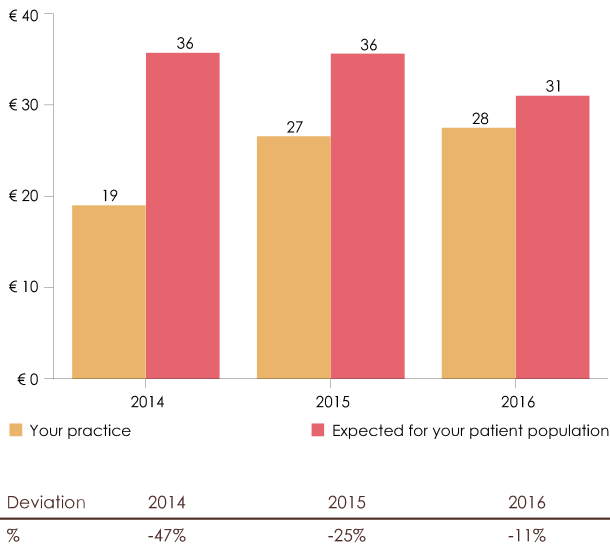
11.2 Overview primary care diagnostics (2016)

	% Care using patients	Per Care using patient	Per registered patient
Your practice	36%	€ 77	€ 28
Expected	37%	€ 85	€ 31
% deviation	-2%	-9%	-11%

Reference date: 01-07-2016

12. Primary care diagnostics division of requests

12.1 Number of requests primary care diagnostics per registered patient
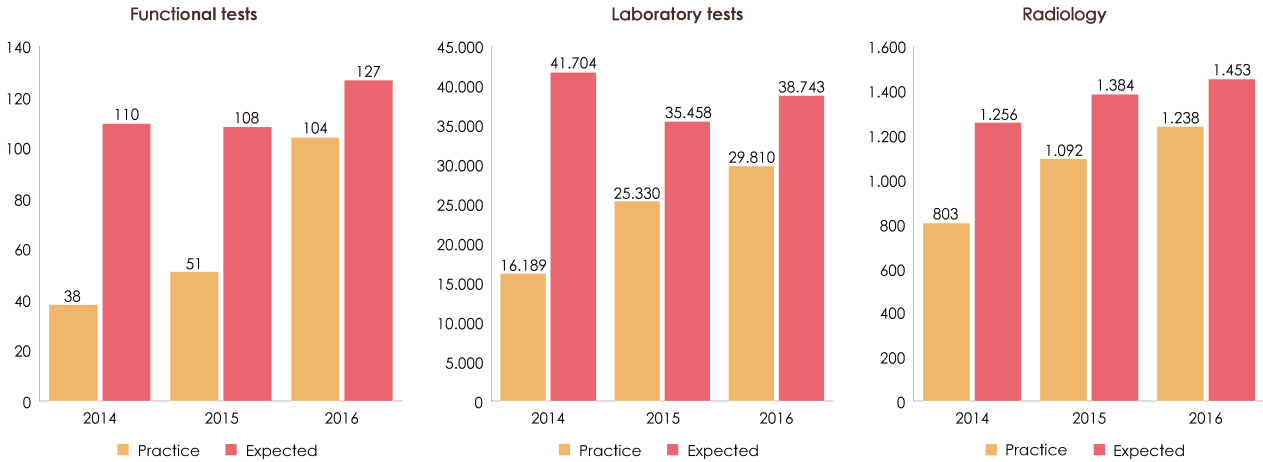
12.2 Percentage requests primary care diagnostics per type of provider

12.3 Percentage requests primary care diagnostics per GP in your practice

Themes | Subthemes | Codes Think-aloud | Codes Interview - recommendations

1) Poor usability of the PCPR

1.1 Desire for clinically meaningful feedback
- Lack of feedback on quality
- Lack of clinical detail of feedback

Feedback on quality
Feedback on common disorders

1.2 Limitations of claims data as source for A&F
- Limitations in reliability and validity of claims data
  - Lack of justification
  - Comparability over time
  - Source health insurer mistrusted
- Lack of individual level feedback
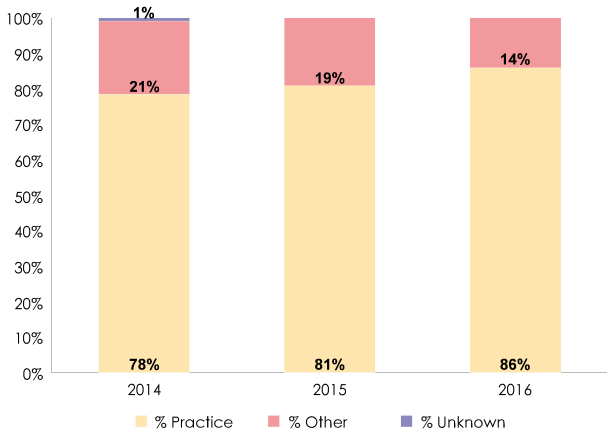- Feedback is outdated

Reliable and valid data
Individual feedback
Recent data

1.3 Performance comparators offer insufficient guidance to assess clinical performance
- Lack of targets
- Unstable trends
- National comparison
- Unclear case-mix correction
- Lack of advice/action plan

Include targets
Compare with similar peers
- City vs. rural
- Type of care
- Integrated care
Include action plan

1.4 Desire to discuss feedback with peers

1.5 Unclear Definitions & figures
- Unclear definitions
- Unclear figures
  - Varying units
  - Varying scales

Clarify definitions
Simplify figures

2) Minimal motivation to change based on the PCPR
- Find usability of the PCPR poor
- Motivated to change claim behavior
- Motivated to change based on well-designed A&F
- A&F in general does not motivate to change behavior