BMJ Open Quality

Utilisation of a cocreation methodology to develop claims-based indicators for feedback on implementation of comparative effectiveness research results into practice

Vera de Weerdt ,[1,2] Hanna Willems,[3] Geeske Hofstra,[4] Sjoerd Repping,[2] Xander Koolman,[1] Eric van der Hijden[1,5]

To cite: de Weerdt V, Willems H, Hofstra G, *et al*. Utilisation of a cocreation methodology to develop claims-based indicators for feedback on implementation of comparative effectiveness research results into practice. *BMJ Open Quality* 2025;**14**:e002542. doi:10.1136/bmjoq-2023-002542

► Additional supplemental material is published online only. To view, please visit the journal online (https://doi.org/10.1136/bmjoq-2023-002542).

Received 7 August 2023
Accepted 30 January 2025

[1]Health Economics, School of Business & Economics, Vrije Universiteit, Amsterdam, the Netherlands
[2]University of Amsterdam, Amsterdam, the Netherlands
[3]Department of Geriatrics, University of Amsterdam, Amsterdam, the Netherlands
[4]Zorginstituut Nederland, Diemen, the Netherlands
[5]Zilveren Kruis Health Insurance, Zeist, the Netherlands

Correspondence to
Vera de Weerdt;
v.de.weerdt@vu.nl

ABSTRACT

Introduction Comparative effectiveness research (CER) often fails to create quality improvement since implementation of CER results in clinical practice is lacking. Claims-based Audit & Feedback (A&F) provides a resource efficient tool to stimulate implementation, but it is unknown whether medical professionals accept claims-based A&F in the context of CER. Therefore, in this study, we developed claims-based indicators using a cocreation approach and evaluated medical professionals' perception regarding the validity and acceptability of these indicators.

Methods Between July 2019 and November 2021, we used a cocreation approach with medical experts to develop claims-based indicators for six CER trials. The aim is to use the indicators for group level feedback on implementation of CER results to medical professionals across all healthcare providers in the Netherlands. To build the indicators, we used the most recent available Dutch national healthcare–related claims data of the year 2017. The cocreation process consisted of the following steps: (1) defining the target indicator, (2) selecting relevant claims codes, (3) testing feasibility of the indicators using Dutch claims data, (4) discussing results of feasibility testing and (5) defining the final indicators and reflecting on the acceptability of the indicators for feedback on implementation of CER results by the experts.

Results Claims-based indicators could not perfectly reflect the CER population for any of the six CER trials. However, the cocreation process did lead to a final indicator that medical experts found acceptable in four of six cases. Recommendations of medical experts for improving claims-based indicators included: select patients with minimal over- or underestimation of the CER population, use proxies to identify patients, determine incidence rather than prevalence for chronic conditions and use data linkage with diagnostic test results.

Conclusion A cocreation approach was a successful way to develop claims-based indicators on implementation of CER results, which were imperfect, but in some cases still acceptable as feedback to medical experts. Thus, for certain topics, claims data may provide a resource efficient data source for A&F interventions aiming to implement CER trials.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Healthcare-related claims data provides a resource efficient data source for Audit & Feedback (A&F) interventions, but claims-based A&F on implementation of comparative effectiveness research (CER) results may not be accepted as valid feedback by medical professionals.

WHAT THIS STUDY ADDS

⇒ This study showed that using a cocreation approach enables the development of claims-based indicators for feedback on implementation of CER results, which are accepted by medical professionals.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ Healthcare managers and policy makers can use cocreation with medical professionals to develop claims-based indicators on CER, to stimulate implementation of CER results in practice.

INTRODUCTION

Quality improvement in healthcare is hampered as clinical studies, such as comparative effectiveness research (CER), fail to be implemented into clinical practice.[1–3] CER trials aim to identify which interventions (treatments or diagnostics) are most effective or cost-effective for specific patient populations.[4] If the most (cost)-effective intervention is identified, this intervention should be implemented into clinical practice, and the other intervention should be de-implemented. However, results of CER trials are slowly, or not at all, implemented into clinical practice, failing to result in meaningful impact on both quality and costs of care.

Audit & Feedback (A&F) is an effective quality-improvement intervention, which could be used to stimulate the implementation of CER.[5–7] By providing medical professionals feedback on the number of patients

receiving the most (cost)-effective versus the other intervention, A&F can stimulate medical professionals to implement CER results.[8–10] Healthcare-related claims data contains information on interventions performed by providers and thus could be used to develop A&F on implementation levels of CER results.[11]

The benefit of reusing claims data for A&F on implementation of CER results is that it offers readily available data.[12–15] Other A&F data sources, such as quality registries or data collected specifically for A&F interventions, can pose additional administrative burden on professionals.[16–18] Using medical record data for A&F is also challenging, since considerable time and costs are associated with extracting this data. Thus, reusing claims data can prevent administrative burden on medical professionals through reducing cost and time related to data collection for A&F.

However, it is unknown whether claims-based A&F provides a suitable source for A&F on implementation of CER results. While claims data have been used for A&F interventions in the past,[19–24] a review suggests claims-based A&F may be less effective in creating quality improvement than A&F based on other data sources.[25] A possible explanation for the lower effectiveness of claims-based A&F is that professionals may not accept claims data as clinically representative.[26 27] Claims data are collected for re-imbursement purposes and thus require less clinical detail than medical records or quality registries, which inform clinical decision making.[28] Due to the high aggregation level of claims codes, patient populations identified in claims data can reflect clinically heterogeneous groups.[13 14 26] A&F interventions aim for quality improvement, by changing providers' clinical behaviour. To effectively change providers' behaviours, it is considered imperative that professionals accept the data on which they receive feedback.[15 29] It is unknown whether medical professionals accept claims-based A&F on implementation of CER results.

While claims data have limitations as a source for A&F, the advantages of using claims data for A&F on CER are important, as the need to implement CER results is high and claims data provide the most resource efficient source for A&F on CER. Therefore, in this study, we developed claims-based indicators using a cocreation approach and evaluated medical professionals' perceptions regarding the validity and acceptability of these indicators.

METHODS

Between July 2019 and November 2021, we conducted a cocreation approach (figure 1) to develop claims-based indicators for six CER trials from varying specialties. This cocreation method was developed for this study, as we found no previous published methods to develop claims-based indicators for CER specifically. Therefore, we developed a systematic cocreation approach, which used elements of established methods for developing guideline indicators, but was adapted for the CER context.[30]

Cocreation was defined as collaboration with stakeholders, in this case medical professionals, during the research process to inform process iterations to the desired outcome, in this case the claims-based indicators.

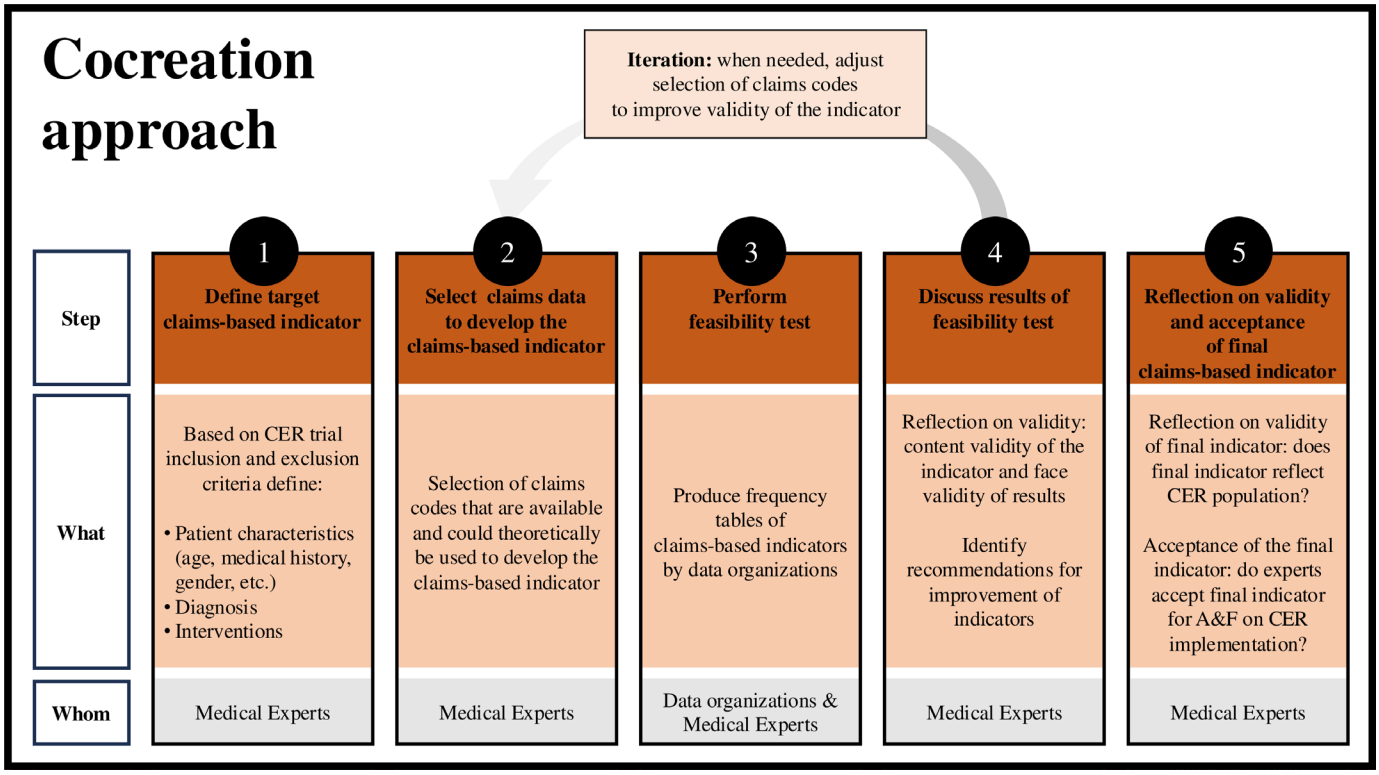

Figure 1 Cocreation approach. CER, comparative effectiveness research

Thus, cocreation allows to use stakeholders' opinions not merely as research data, but to inform the research process and outcome.[31] Methods for developing guideline indicators are not fully applicable for CER trials, as we do not select, assess and prioritise indicators from guidelines. Instead, an indicator which presents the implementation level of a CER outcome is defined by the population under study in the CER trial. Thus, a CER indicator should present the volume of patients with the diagnosis and characteristics receiving the interventions as defined by the CER study (figure 2). The cocreation approach was based on the agreement of study authors.

Aim is to use the indicators for group level feedback on implementation of CER results to medical professionals across all healthcare providers in the Netherlands.

Patient or public involvement

Patients were not involved in this study. We involved medical professionals during the cocreation process, as they are the target audience for the claims-based A&F on CER developed in this study.

The medical professionals were not involved in the development of the research question, but the research question was informed by previous qualitative research among other medical professionals. In the role of principal investigator of the CER trials, the medical experts will be responsible for national dissemination of the CER trials results when the trials are complete. The claims-based A&F developed in this study can be used by the medical experts and their professional associations to stimulate implementation of CER results to healthcare providers.

Cocreation process

The cocreation approach involved five steps, which were followed for each CER trial. To prevent bias, the cocreation process was performed for CER studies which are currently in the patient inclusion phase; thus, results are unknown, and it is yet unclear which treatment is most (cost)-effective. Participants in the cocreation approach were medical experts (n=14), from six different specialties and five different hospitals in the Netherlands, who were each involved in one of the CER trials. For two trials, there were three medical experts involved in the respective cocreation process; for four trials, there were two medical experts involved in the respective cocreation process. The majority of the medical experts (86%, n=12) was a practicing physician. All medical experts were a principal investigator or an executive researcher in one of the CER trials. The principal investigators are acknowledged by their scientific medical association (Dutch: 'Wetenschappelijke Vereniging') as national medical experts of the CER study. All steps with medical experts were conducted in an online or face-to-face meeting in which researchers VW and EvdH or GH were present. All meetings were audio- or videorecorded. Furthermore, two Dutch national organisations specialised in claims-based benchmarking VEKTIS and DHD were involved, to perform feasibility testing.

VEKTIS is commissioned by the Dutch association of Health Insurers (Dutch: Zorgverzekeraars Nederland) and receives claims data from all Dutch Health insurers on healthcare in primary, secondary and tertiary care, including private clinics. VEKTIS has data coverage of

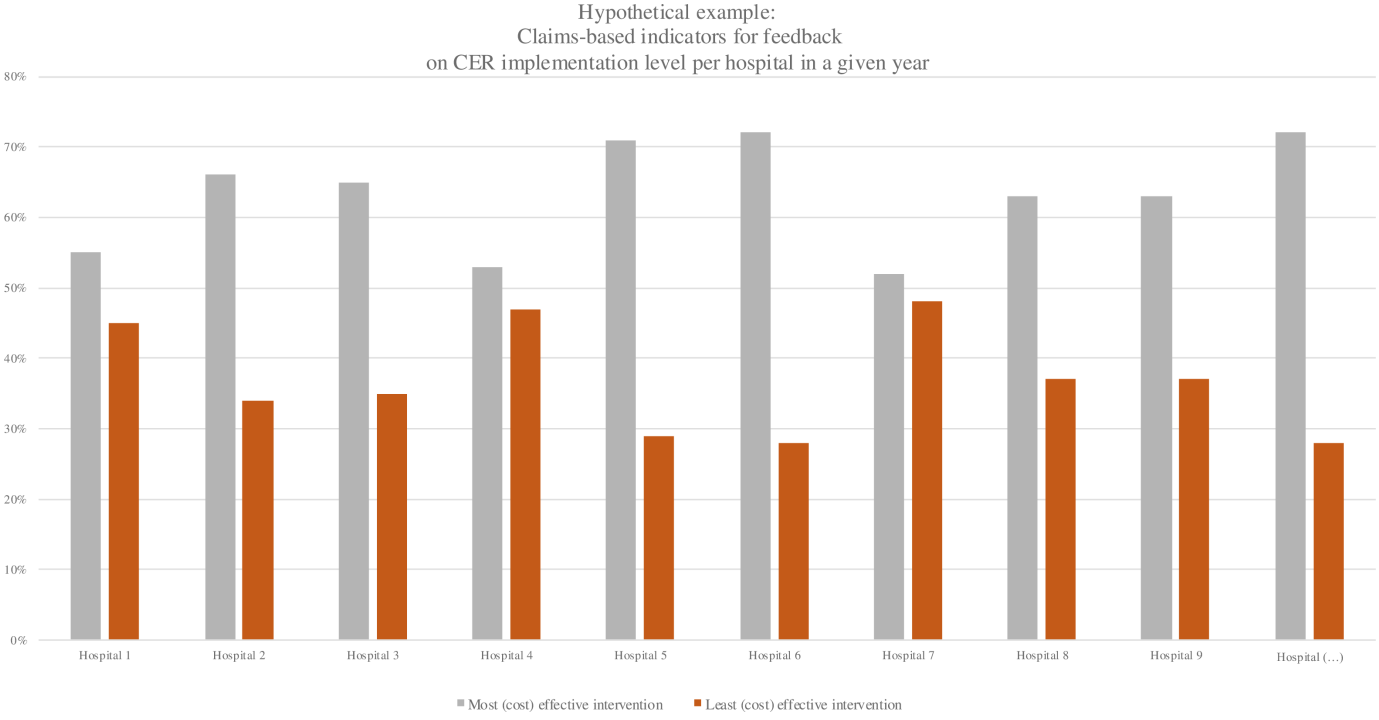

Figure 2 Hypothetical example of claims-based indicator for feedback on implementation of comparative effectiveness research (CER) results per hospital

99% of insured healthcare in the Netherlands). DHD (Dutch Hospital Data) is commissioned by the associations of hospitals (NVZ) and association of academic hospitals (NFU). DHD collects claims data for 99% of hospitals in the Netherlands.

The cocreation process comprised the following five steps: First, we asked the medical experts to define the target indicator based on the CER trial protocol as the volume of patients receiving the hypothesised most (cost)effective intervention (numerator) as a portion of the total volume of patients both studied interventions (denominator) in a given year.

Second, we asked the medical experts to select Dutch claims diagnostic codes and intervention codes which reflect the patient population and interventions of the CER trial. For this, we presented the medical experts with publicly available lists of Dutch claims codes and asked them to select relevant codes.[32] Based on this selection, we defined a concept claims-based indicator for each CER trial.

Third, indicators were tested for feasibility. The organisations who specialise in claims-based benchmarking calculated the indicators as the proportion of patients with the hypothesised most (cost)-effective intervention for each medical specialist care provider (hospital or private clinic) in the Netherlands in the year 2017. Data for the year 2017 were used as this was the first full year for which claims data were complete at the beginning of this study. Healthcare-related claims data in the Netherlands are specifically developed for national use and comprise patient-level data with billing codes including a claims diagnostic code and an intervention code (similar to a Diagnosis-Related-Group type of structure); prescribed drugs; and sociodemographic data such as age and sex.[32]

Fourth, we held a meeting with the medical experts to discuss the results of the feasibility test and to collect recommendations for improvement of the concept claims-based indicator. We presented the selection of claims codes used to develop the indicator and the results of the feasibility test to the medical experts. We asked the experts to reflect on validity in terms of content validity of the indicator and the face validity of the results. Content validity refers to whether an indicator measures what it intends to measure and was in this case defined as: does the claims-based indicator, considering the definitions of the selected claims codes, reflect the CER population or a different patient population?[33 34] Face validity was defined as whether results of feasibility testing (patient volume per treatment on national level and provider level) were considered plausible based on the experts' professional experience and existing literature.[33] The experts were then asked whether they had recommendations for improving the claims-based indicator (e.g., by including or excluding other claims codes). If the experts had no recommendation to improve the claims-based indicator, we concluded the indicator development process. In case the experts did have recommendations to improve the claims-based indicator, we repeated steps (2), (3) and (4) until they could not generate further recommendations.

As the fifth and final step, we asked the experts again to reflect on the validity of the final indicator in terms of content validity and face validity and whether they would accept the indicator for feedback on implementation of CER results. Acceptance was based on verbal informal consensus between the medical experts and was defined as whether they accepted the results of feasibility testing as valid feedback on implementation levels of CER results.

Selection of CER trials

We selected CER trials based on the following criteria: (1) the CER trial is supported by the national professional society of the respective medical specialty (Each medical speciality in the Netherlands is represented by a professional society, called the "academic society" (Dutch: Wetenschappelijke Vereniging)); (2) the CER trial is an ongoing trial; thus, it yet unknown which is the most (cost)-effective treatment; and (3) the CER trials are from diverse clinical specialties and medical topics.

At the time of this study, there were 42 CER trials in the Netherlands which met the first two criteria. We purposively selected eight diverse trials and invited these research teams for participation in our study.

One research team declined participation in the cocreation process before finalising the inclusion process for their CER trial, because they were concerned about possible influence on the inclusion process of the CER trial. Furthermore, we ended collaboration with one research team because the CER trial was discontinued for lack of inclusions. Thus, the final selection included six CER trials (see table 1).

RESULTS

Summary of results

Using the cocreation approach, we developed claims-based indicators for all six CER trials (table 2) (resulting indicators are presented in online supplemental file 1). For none of the six CER trials, claims data contained all information needed to directly identify the CER population. The medical experts gave several recommendations to improve the indicators: select patients with minimal over- or underestimation of the CER population, use proxies to identify patients, determine incidence rather than prevalence for chronic conditions and use data linkage with diagnostic test results. Despite improving the indicators through multiple iterations, none of the final claims-based indicators perfectly reflected the CER population. In four cases (Proclion, CAPP, MIRA2, STONE), the inaccuracy between the final claims-based indicator and the CER population was minimal; thus, the indicator was accepted by the experts for feedback on implementation level of CER results.

Table 1 Research topics of the CER trials

CER trial	Trial registration number*	Research question	Patient population	Intervention
CAPP	NL9371	What is the most (cost)-effective treatment for children with complex appendicitis: operative or conservative treatment?	Children of 0–18 years with complex appendicitis with/without infiltrate/abscess	► Operative – Laparoscopic – Open procedure ► Conservative
Proclion	NA†	What is the most (cost)-effective treatment for patients with critical ischaemia of the legs: operative or conservative treatment?	Patients of ≥18 years with critical ischemia of the legs	► Operative – Laparoscopic – Open procedure ► Conservative
DART	NL6201	What is the most (cost)-effective treatment for elderly with an intra-articular distal radius fracture type C: operative or conservative treatment ?	Patients of ≥65 years with intra-articular distal radial fracture type C	► Operative ► Conservative
MIRA2	NL7817	What is the most effective treatment for woman with heavy menstrual bleeding: endometrial ablation or combined treatment of endometrial ablation with hormonal IUD?	Patients of 25–60 years with heavy menstrual bleeding	► Endometrial ablation ► Endometrial ablation combined with hormonal IUD
STONE	NL8128	What is the most (cost)-effective treatment for patients with obstructive kidney stones: double J or nephrostomy catheter?	Patients of ≥18 years with obstructive kidney stones	► Double J catheter ► Nephrostomy catheter
Growth hormone	NL6440	What is the most (cost)-effective treatment for children with idiopathic growth hormone deficiency that have normal growth hormone levels mid-puberty: continued growth hormone treatment or discontinued growth hormone treatment?	Patients of 10–18 years with idiopathic growth hormone deficiency	► Continued growth hormone treatment ► Discontinued growth hormone treatment

*From the Netherlands Trial Register.
†Not registered due to observational study character.
CER, comparative effectiveness research; IUD, intrauterine device.

Recommendations to develop claims-based indicators

Select patients with minimal over- or underestimation of the CER population

Claims data did not provide sufficient detail to perfectly identify the CER populations. However, the medical experts were able to select patient populations based on claims diagnoses codes that closely resembled the CER populations. Inclusion of certain patient populations could lead to an overestimation of the true CER population, while excluding patient populations could lead to an underestimation of the true CER population. The medical experts chose the combination of claims codes that led to minimal over- or underestimation of the CER population. In other words, if exclusion of a population led to a small underestimation of the CER population, while inclusion of that population would lead to a large overestimation of the CER population, the population was excluded. For example, for the Proclion trial, some patients are coded under claims diagnostic code

Table 2 Results of cocreation process per CER trial

CER trial	Step 1 CER population and intervention	Step 2 Claims code available in dutch claims data version 2018 | Y/N/P*	Step 4 Alternative dutch claims codes selected to develop final indicator	Step 5.1 Final indicator accurate representation of CER population?	Step 5.2 Final indicator accepted by medical experts for A&F on implementation of CER results?
Proclion	PP	PP	PP	No, underestimation of patients registered under claims code diabetic foot	Yes
	► Patients≥18 years	► Yes	► N.A.		
	► Critical ischemia of the legs	► Partly	► Peripheral Arterial Occlusive Disease stage 3 and 4, exclusion of patients with diabetic foot		
	INT	INT	INT		
	► Operative open procedure	► Yes	► N.A.		
	► Operative endoscopic procedure	► Yes	► N.A.		
	► Conservative	► Partly	► Selection of newly diagnosed patients		
STONE	PP	PP	PP	No, overestimation of patients with ureter stone that do not suffer from obstructive kidney disease	Yes
	► Patients≥18 years	► Yes	► N.A.		
	► Obstructive kidney disease	► Yes	► Ureter stone and/or Kidney stone with exclusion of same day surgeries for other diagnoses		
	INT	INT	INT		
	► Double J catheter	► Yes	► N.A.		
	► Nephrostomy catheter	► Yes	► N.A.		
MIRA2	PP	PP	PP	No, overestimation of patients who suffer from menstrual disorder, but not heavy menstrual bleeding; and overestimation of patients who had therapeutic hysteroscopy, but not endometrial ablation	Yes
	► Patients 25–60 years	► Yes	► N.A.		
	► Heavy menstrual bleeding	► No	► Menstrual disorder		
	INT	INT	INT		
	► Endometrial ablation (EA)	► No	► Therapeutic hysteroscopy (TH)		
	► EA and hormonal IUD	► No	► TH and same day IUD placement		
CAPP	PP	PP	PP	No, overestimation of patients who do not have complex appendicitis, but simple appendicitis with complications	Yes
	► Patients 0–18 years	► Yes	► N.A.		
	► Complex appendicitis	► Partly	► Appendicitis, proxy length of stay for complex appendicitis		
	► with or without infiltrate/abscess	► Partly	► No alternative code available		
	INT	INT	INT		
	► Conservative	► Yes	► N.A.		
	► Operative open procedure	► Yes	► N.A.		
	► Operative endoscopic procedure	► Yes	► N.A.		

Continued

Table 2 Continued

CER trial	Step 1 CER population and intervention	Step 2 Claims code available in dutch claims data version 2018 | Y/N/P*	Step 4 Alternative dutch claims codes selected to develop final indicator	Step 5.1 Final indicator accurate representation of CER population?	Step 5.2 Final indicator accepted by medical experts for A&F on implementation of CER results?
DART	PP	PP	PP	No, overestimation of patients with patients who have other wrist fracture other than distal radial fracture Type C	No, large gap with target population
	► Patients≥65 years	► Yes	► N.A.		
	► Distal radial fracture	► No	► Wrist fracture		
	► Intra-articular	► No	► No alternative code available		
	► Type C	► No	► No alternative code available		
	INT	INT	INT		
	► Conservative	► Yes	► N.A.		
	► Operative	► Yes	► N.A.		
Growth-hormone (GH)	PP	PP	PP	No, impossible to detect mid-puberty characteristic and not fully possible to distinguish patients with known causes from GH deficiency from idiopathic GH deficiency	No, results do not compare to existing quality registration and hard difficult to interpret
	► Patients in mid-puberty	► No	► Patients between 10–18 years		
	► with Idiopathic GH deficiency (GHD)	► No	► Growth hormone deficiency with exclusion of a number of specific causes of growth hormone deficiency		
	► and normal GH stimulation test result	► No	► GH stimulation test executed during mid-puberty age range		
	INT	INT	INT		
	► GH suppletion	► Yes	► N.A.		
	► Discontinuation of GH suppletion	► Yes	► N.A.		

*Yes=claims code/ICD-10 code exists, no=claims code/ICD-10 code does not exist, partly=claims code/ICD-10 code similar
CER, comparative effectiveness research; ICD-10, International Classification of Diseases, Tenth Revision; INT, Intervention; IUD, intrauterine device; PP, patient population; T, treatment.

'diabetic foot', while the majority of patients coded under claims diagnostic code 'diabetic foot' are not the Proclion population. Thus, exclusion of patients with claims diagnostic code 'diabetic foot' resulted in a smaller inaccuracy of the Proclion population.

Use proxies to identify patients

In our study, we used three proxies to identify CER populations: (1) length of admission as a proxy for disease severity, (2) combination of diagnostic and intervention codes as a proxy for the intervention and (3) exclusion of patients based on medical history/same day interventions. Using these proxies enabled us to give a more valid representation of the CER populations: length of admission ≥3 days was used for the CAPP trial to distinguish patients with complex appendicitis (the CER population) from patients with simple appendicitis (<3 days of admission). For one CER trial (MIRA2), the intervention studied (endometrial ablation) was not captured within claims data. Only a more generic intervention code (therapeutic hysteroscopy) was available in claims data, which includes irrelevant interventions. The experts proposed using the combination of this generic intervention code, with the diagnostic code menstrual disorder as a proxy for the studied intervention endometrial ablation. We used exclusion of medical history/same day interventions to identify the CER populations of the Growth Hormone and STONE trial, respectively. For the Growth Hormone trial, claims data did not capture whether patients suffer from an idiopathic growth hormone deficiency or have a known cause for the deficiency. Thus, we excluded known causes of growth hormone deficiency based on medical history, to create a more valid representation of the CER population. For the STONE trial, claims data did not capture whether patients suffered from an obstructive ureter/kidney stone. Thus, we excluded patients who underwent same day interventions for other obstructive disorders, such as tumours, to create a more valid representation of the CER population.

Determine incidence rather than prevalence for patients with chronic conditions

To give a valid representation of implementation levels of CER results, it is necessary to correctly classify patients into the numerator versus the denominator of the indicator; thus, it is important to distinguish which patients have received which intervention. For one CER trial (Proclion), the experts raised that there were such classification issues, as patients often belonged to both intervention groups. First, the chronic character of the condition (critical ischemia) hampers the ability to classify patients as conservative versus operative, as patients who received operative treatment remain under follow-up, which is registered under conservative claims codes. Thus, patients who have received operative treatment in 1 year, but remain under follow-up, could be misclassified as having received only conservative treatment. Furthermore, as patients can receive multiple follow-up appointments in a year, this may create an overestimation of the conservatively treated proportion of patients. To resolve these issues, the experts suggested to determine incidence, rather than prevalence for these patients. Newly diagnosed patients in a given year were identified and subsequently were classified as operatively treated if they received operative treatment within 12 months of first diagnosis, while patients who did not receive operative treatment within 12 months of first diagnosis were classified as conservatively treated. This process excludes patients who remain in follow-up after operative treatment and thereby gives a more accurate reflection of the portion operatively treated patients versus conservatively treated patients.

Second, the experts raised that the condition occurs bilateral. Thus, patients may receive conservative treatment for one limb, while they receive operative treatment for the other limb. Dutch claims data currently do not contain information on the side of the condition. However, the experts accepted the incidence of patients as indicative of implementation of CER results, under the assumption that treatment patterns are not significantly different for the first leg and the second leg.

Use data linkage with diagnostic test results

Diagnostic test results, including blood test results and X-ray results, would be necessary to accurately identify the patient population of several CER trials (Growth Hormone, DART, CAPP, STONE). For example, we were unable to identify patients for the Growth Hormone trial as being 'mid-puberty', as this is not a diagnosis, but a clinical state which is not described in claims data.However, claims data only encompass information on whether a diagnostic test is performed, but not the results of the test. Thus, in our study, we were unable to accurately identify these CER populations. The experts suggested to use data linkage of claims data and diagnostic test results to identify patients.

Reasons for non-acceptance of claims-based indicators for CER

For two cases (DART and Growth Hormone), the experts did not find the final indicator acceptable for A&F on implementation of CER results.

For the DART trial, the claims-based indicator presented a population (>12 000 patients with wrist fractures) which was three times larger than the actual CER population (~4500 patients with distal radial fracture type C). Also, the indicator ratio (numerator-denominator) was considered inaccurate. Thus, the experts assessed this indicator as not acceptable for A&F on implementation of CER results. For the Growth Hormone trial, the experts also did not accept the indicator as the indicator volume and the indicator ratio were inaccurate, due to the inability to identify whether patients were 'mid-puberty'.

DISCUSSION

The aim of this study was to develop claims-based indicators using a cocreation approach and to evaluate medical professionals' perceptions regarding the validity and acceptability of these indicators.

For none of the CER trials, claims-based indicators accurately reflected the CER population. However, by using the cocreation approach, we were able to improve the validity of the indicators by selecting the patients with minimal over- or underestimation of the CER population, using proxies to identify patients and determining incidence rather than prevalence for patients with chronic conditions. As a result, the medical experts accepted four of six claims-based indicators for feedback on implementation of CER results. The experts only accepted claims-based indicators for CER trials if they considered the difference between the claims-based indicator, and the CER population was minimal and if the indicator ratio (numerator–denominator) seemed accurate.

This study primarily focused on the feasibility of using claims data to develop quality indicators and whether the resulting claims-based indicators are acceptable to physicians, the intended users. Various methods are used for quality indicator development in healthcare, with various types of end results.[30 35 36] Evidence-based approaches are employed when scientific evidence or guidelines on quality of care are available. Consensus-based approaches, such as Delphi- or RAND studies, are employed for clinical topics lacking scientific evidence or when patient perspectives are desired. Furthermore, a combination between evidence-based and consensus-based approaches is often used.[36–41] Many studies use these approaches solely focus on the identification, selection and prioritisation of indicators, often resulting in a merely theoretical description of an indicator.[36 42–44] However, quality indicators are only relevant when they are used in practice to measure performance. Therefore, the feasibility

of building indicators with new data or existing data is a crucial attribute in indicator development.[35 38 45]

For the development of indicators, we followed a methodology similar to that of Blozik *et al*,[30] which employs an evidence-based approach to indicator development, including feasibility testing using claims data. Unlike Blozik *et al*, we did not focus on the identification, selection and prioritisation of indicators. Instead, we developed single process indicators directly defined by the CER studies. The definition of the indicators in this study can therefore be categorised as an evidence-based approach to indicator development.[35] This study employed a cocreation method for feasibility testing.

Other methods for feasibility testing include formal validity testing in which sensitivity and specificity of indicators are tested against data like medical records or administrative data.[35 45 46] In this type of validity testing, claims data are easily discarded as a source for quality indicators, since claims data lacks detail which is often needed to identify patient populations.[45 47] Other studies employ purely qualitative feasibility testing in which participants are asked to describe whether they find indicators feasible to implement.[48] Similar to Blozik *et al*,[30] this study used real-world claims data to develop indicators. In contrast to Blozik and other studies, we did not employ Likert scales to assess the feasibility of indicators.[49 50] Instead, we used face-to-face meetings in which validity of results were discussed. This qualitative approach allowed to create an in-depth understanding of physicians' reasoning when assessing validity of the indicators.

In our study, the medical experts accepted A&F with a minimal level of over- or underestimation of the CER population. As in other studies, it was not fully clear what level of data accuracy is sufficient for medical professionals to accept A&F.[51] Also, acceptance of A&F is not merely determined by the objective quality of the data, but also by recipients' perceptions of the data accuracy. Other studies found that medical professionals were more likely to accept data as accurate when the feedback showed high performance, when recipients have an internal locus of control and when the feedback is not used in a punitive manner.[10 51–55] Thus, whether claims-based A&F will be accepted by medical professionals presumably remains recipient- and context dependent.

Furthermore, if claims-based A&F is accepted by medical professionals, this does not guarantee effectiveness of A&F in implementation of CER results. The effectiveness of A&F depends on many factors, such as frequency and timeliness of the feedback, co-interventions alongside A&F and the context in which A&F is given. Furthermore, using claims-based A&F may pose the risk of incentivising medical professionals to change their claims registration, rather than implementing CER results.[51] Future research should examine whether claims-based A&F can indeed stimulate implementation of CER results and whether additional measures such as co-interventions are needed.

Strengths and limitations

This study builds on a previous approach regarding indicator development[30] and similarly incorporates a strong focus on the feasibility attribute of indicators, thereby taking indicator development beyond a mere theoretical description.[35 38 41] In contrast to other studies,[35 42 45] we take an iterative, qualitative approach to feasibility testing, which allowed to gain an in-depth understanding of medical experts' consideration for acceptance of indicators and allowed us to make continuous improvements to the indicators. Third, during development of the indicators we focused on the physicians' perspective on the validity of indicators, which ensured physicians were unable to discard indicators out of conflict of interests, while increasing the acceptability of the indicators, by focusing on the perspective of the physicians as intended users of the indicators.[10 35] Fourth, by defining the indicators based on the CER trials, the indicators are evidence based, which is seen as the gold standard in indicator development.[35] Lastly, we developed claims-based indicators for a wide range of clinical topics from varying specialties, thereby increasing the general and broad utility of our results.

Both a strength and a limitation of this study are our approach to collaborate with national experts on the respective CER evaluations in development and evaluation of the indicators. It is a strength because these experts are national authorities on this topic and possess in-depth knowledge of the indicators' validity. Furthermore, their perspective on the acceptability of these indicators may be more indicative than the general professional perception as they are opinion leaders in the field, who take the role of creating support or resistance towards indicators. However, this dual role could also introduce bias, as the experts were involved in the development process. On one hand, the medical experts may be positively biased towards the indicators, which is in line with a previous study that found acceptance of feedback is increased by knowledge of the audit process.[54] On the other hand, the medical experts may be negatively biased as they may be more critical of the indicators due to their future responsibility of advocating for the accuracy of the indicators to their peers. Therefore, future research should test the generalisability of our results by examining whether medical professionals who are not involved in CER studies also accept the claims-based indicators for feedback on implementation of CER results.

Other limitations include that this was an exploratory study in which we created a limited number of claims-based indicators and collaborated with a limited number of medical experts per study for feasibility reasons. The number of professionals and the composition of a panel determines how indicators are rated;[41] thus, the limited number of medical experts involved in this study may decrease the external validity of our results. Furthermore, we did not formally test the validity of the claims-based indicators; thus, reporting

issues in the claims data which are unknown to the medical experts cannot be ruled out. However, our cocreation approach was designed to increase validity of the indicators by minimising issues with data quality and reporting. Our experts were selected based on their quantitative methodological knowledge in combination with expert content knowledge and relevant physician experience. Before their assessment of indicator validity, they were made familiar with the aim and context of this study. We explicitly discussed numerators, denominators and the exclusion criteria for each indicator. In the Dutch setting, this should enable the experts to assess the content validity, information bias, confounding and selection bias of the proposed indicators. The two Dutch national organisation specialised in claims-based benchmarking also provided expert knowledge regarding the validity and reliability of the claims data, which increased the validity of our results. A selected number of studies in the Netherlands did formally test other claims-based indicators against medical charts and found those claims-based indicators to be valid.[56–59] Finally, while we focused on a wide range of clinical conditions, the CER trials mainly involved surgical treatments, thus whether our results are generalisable to other clinical fields is unclear.

Conclusion

A cocreation approach was a successful way to develop claims-based indicators on implementation of CER results, which were imperfect, but in some cases still acceptable as feedback to medical experts. Thus, for certain topics, claims data may provide a readily available and cost-efficient data source for A&F interventions aiming to implement CER trials. Future research should examine whether claims-based A&F indicators from this approach are effective in increasing implementation of CER results.

Contributors Conception/design: VW and EH. Acquisition/analysis: VW, GH, XK and EH. Interpretation: VW, GH, SR, XK and EH. Drafted or contributed substantive revisions: VW, HW, SR, XK and EH. VW is the guarantor. All authors read and approved the final manuscript. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

Funding This study was commissioned by 'Leading the Change'[61]. 'Leading the Change' is a collaboration between the Dutch Federation of Medical Specialists, the Dutch Association of Health Insurers and the Dutch Patient Federation and is funded by the Dutch Association of Health Insurers. 'Leading the Change' has no financial aims. Author HW is chair of Leading the Change. 'Leading the Change' and author HW had no role in the design of the study, data collection, data analysis, interpretation of data. HW contributed to writing of the manuscripin her academic role as supervisor of PhD candidate VW. HW did not contribute to this manuscript on behalf of 'Leading the Change'.

Patient and public involvement Patients and/or the public were involved in the design, conduct, reporting or dissemination plans of this research. Refer to the Methods section for further details.

Patient consent for publication Not applicable.

Ethics approval This study involved human participants, but the Research Ethics Review Committee (BETHCIE) of the Faculty of Sciences, Vrije Universiteit Amsterdam, committee declared, based on the submitted information, that the research proposal complied with the ethical guidelines of the faculty. 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 upon reasonable request. The data generated and analysed during the current study are not publicly available since these include recordings of cocreation sessions with individual persons, but recordings/transcripts are available from the corresponding author on reasonable request.

ORCID iD

Vera de Weerdt http://orcid.org/0000-0003-3644-620X

REFERENCES

1 Avorn J, Fischer M. 'Bench To Behavior': Translating Comparative Effectiveness Research Into Improved Clinical Practice. *Health Aff (Millwood)* 2010;29:1891–900.
2 Gibson TB, Ehrlich ED, Graff J, *et al*. Real-world impact of comparative effectiveness research findings on clinical practice. *Am J Manag Care* 2014;20:e208–20.
3 Kristensen N, Nymann C, Konradsen H. Implementing research results in clinical practice- the experiences of healthcare professionals. *BMC Health Serv Res* 2016;16:48.
4 Sox HC, Recovery A. Annals of Internal Medicine Medicine and Public Issues Comparative Effectiveness Research: A Report From the Institute. *Ann Intern Med* 2009;151:203–5.
5 Ivers N, Jamtvedt G, Flottorp S, *et al*. Audit and feedback: effects on professional practice and healthcare outcomes. *Cochrane Database of Systematic Reviews* 2012;2012:1–217.
6 Hysong SJ, Best RG, Pugh JA. Audit and feedback and clinical practice guideline adherence: making feedback actionable. *Implement Sci* 2006;1:9.
7 Ivers NM, Desveaux L. De-implementation of low-value care: Use audit and feedback wisely. *Healthc Pap* 2019;18:41–7.
8 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.
9 Gude WT, van Engen-Verheul MM, van der Veer SN, *et al*. How does audit and feedback influence intentions of health professionals to improve practice? A laboratory experiment and field study in cardiac rehabilitation. *BMJ Qual Saf* 2017;26:279–87.
10 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. *Implement Sci* 2019;14:40.
11 Ulyte A, Bähler C, Schwenkglenks M, *et al*. Measuring diabetes guideline adherence with claims data: systematic construction of indicators and related challenges. *BMJ Open* 2019;9:e027138.
12 Weiner JP, Powe NR, Steinwachs DM, *et al*. Applying insurance claims data to assess quality of care: a compilation of potential indicators. *QRB Qual Rev Bull* 1990;16:424–38.
13 McKee M. Routine data: a resource for clinical audit? *Quality and Safety in Health Care* 1993;2:104–11.
14 Iezzoni LI. Annals of Internal Medicine Assessing Quality Using Administrative Data Content of Administrative Databases. *Ann Intern Med* 1997;127:666–74.
15 Bonis PAL. Quality incentive payment systems: promise and problems. *J Clin Gastroenterol* 2005;39:S176–82.

16 Wilcox N, McNeil JJ. Clinical quality registries have the potential to drive improvements in the appropriateness of care. *Med J Aust* 2016;205:S27–9.
17 Levay C. Policies to foster quality improvement registries: lessons from the Swedish case. *J Intern Med* 2016;279:160–72.
18 Hoeijmakers F, Beck N, Wouters MWJM, *et al*. National quality registries: how to improve the quality of data? *J Thorac Dis* 2018;10:S3490–9.
19 Schectman JM, Kanwal NK, Schroth WS, *et al*. The effect of an education and feedback intervention on group-model and network-model health maintenance organization physician prescribing behavior. *Med Care* 1995;33:139–44.
20 Hux JE, Melady MP, DeBoer D. Confidential prescriber feedback and education to improve antibiotic use in primary care: a controlled trial. *CMAJ* 1999;161:388–92.
21 Søndergaard J, Andersen M, Vach K, *et al*. Detailed postal feedback about prescribing to asthma patients combined with a guideline statement showed no impact: a randomised controlled trial. *Eur J Clin Pharmacol* 2002;58:127–32.
22 Pimlott NJG, Hux JE, Wilson LM, *et al*. Educating physicians to reduce benzodiazepine use by elderly patients: a randomized controlled trial. *CMAJ* 2003;168:835–9.
23 Beck CA, Richard H, Tu JV, *et al*. Administrative Data Feedback for Effective Cardiac Treatment: AFFECT, a cluster randomized trial. *JAMA* 2005;294:309–17.
24 Brunette MF, Dzebisashvili N, Xie H, *et al*. Expanding Cessation Pharmacotherapy Via Videoconference Educational Outreach to Prescribers. *Nicotine Tob Res* 2015;17:960–7.
25 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.
26 Powell AE, Davies HTO, Thomson RG. Using routine comparative data to assess the quality of health care: understanding and avoiding common pitfalls. *Qual Saf Health Care* 2003;12:122–8.
27 Freitas JA, Silva-Costa T, Marques B, *et al*. Implications of data quality problems within hospital administrative databases. *IFMBE Proc* 2010;29:823–6.
28 Peek N, Holmes JH, Sun J. Technical challenges for big data in biomedicine and health: data sources, infrastructure, and analytics. *Yearb Med Inform* 2014;9:42–7.
29 Van HoofTJ, Pearson DA, Giannotti TE, *et al*. Lessons Learned from Performance Feedback. *J Healthc Qual* 2001;28:20–31.
30 Blozik E, Farcher R, Graber SM, *et al*. Evidence-based indicators for the measurement of quality of primary care using health insurance claims data in Switzerland: update of the SQUIPRICA working group. *BMC Health Serv Res* 2022;22:1.
31 Labib K, Pizzolato D, Stappers PJ, *et al*. Using co-creation methods for research integrity guideline development – how, what, why and when? *Account Res* 2024;31:531–56.
32 Nederlandse Zorgautoriteit, Documenten Dbc-Pakket. De-implementation of low-value care: use audit and feedback wiselydutch healthcare authority (dutch: nederlandse zorgautoriteit (nza),dutch claims codes 2017. 2017. Available: https://puc.overheid.nl/nza/doc/PUC_259810_22/1/
33 Leemans K, Deliens L, Francke AL, *et al*. Quality indicators for palliative care services: mixed-method study testing for face validity, feasibility, discriminative power and usefulness. *Palliat Med* 2015;29:71–82.
34 Koolman X, Zuidgeest M, Visser J, *et al*. Indicatorstandaard 2.0: methodologische criteria voor de ontwikkeling van betrouwbare kwaliteitsindicatoren in de zorg. 2012. Available: https://www.zorginzicht.nl/binaries/content/assets/zorginzicht/ontwikkeltools-ontwikkelen/Indicatorstandaard+2.0.pdf
35 Campbell SM, Braspenning J, Hutchinson A, *et al*. Research methods used in developing and applying quality indicators in primary care. *Qual Saf Health Care* 2002;11:358–64.
36 Kötter T, Blozik E, Scherer M, *et al*. *MethodsCitationKotter.Pdf*. 7.2012.Available: http://www.implementationscience.com/content/7/1/21%0AImplementation
37 McArthur A, Munn Z, Lizarondo L. Methods for developing quality indicators for evidence implementation: a scoping review protocol. *JBI Evid Synth* 2023;21:601–8.
38 Fekri O, Leeb K, Gurevich Y. Systematic approach to evaluating and confirming the utility of a suite of national health system performance (HSP) indicators in Canada: a modified Delphi study. *BMJ Open* 2017;7:e014772.
39 Parmelli E, Langendam M, Piggott T, *et al*. Guideline-based quality assurance: a conceptual framework for the definition of key elements. *BMC Health Serv Res* 2021;21:173.
40 Langendam MW, Piggott T, Nothacker M, *et al*. Approaches of integrating the development of guidelines and quality indicators: a systematic review. *BMC Health Serv Res* 2020;20:875.
41 Bobrovitz N, Parrilla JS, Santana M, *et al*. A qualitative analysis of a consensus process to develop quality indicators of injury care. *Implement Sci* 2013;8:45.
42 Wollersheim H, Hermens R, Hulscher M, *et al*. Clinical indicators: development and applications. *Neth J Med* 2007;65:15–22.
43 Ohtera S, Kanazawa N, Ozasa N, *et al*. Proposal of quality indicators for cardiac rehabilitation after acute coronary syndrome in Japan: a modified Delphi method and practice test. *BMJ Open* 2017;7:e013036.
44 Rendell N, Rosewell A, Lokuge K, *et al*. Common Features of Selection Processes of Health System Performance Indicators in Primary Healthcare: A Systematic Review. *Int J Health Policy Manag* 2022;11:2805–15.
45 Chalmers K, Badgery-Parker T, Pearson S-A, *et al*. Developing indicators for measuring low-value care: mapping Choosing Wisely recommendations to hospital data. *BMC Res Notes* 2018;11:163.
46 Spertus JA, Eagle KA, Krumholz HM, *et al*. American College of Cardiology and American Heart Association methodology for the selection and creation of performance measures for quantifying the quality of cardiovascular care. *Circulation* 2005;111:1703–12.
47 Klazinga N, Fischer C, ten Asbroek A. Health services research related to performance indicators and benchmarking in Europe. *J Health Serv Res Policy* 2011;16 Suppl 2:38–47.
48 Hyatt A, Gough K, Chung H, *et al*. Development of consensus quality indicators for cancer supportive care: a Delphi study and pilot testing. *BMC Health Serv Res* 2024;24:377.
49 Helsloot K, Walraevens M, Besauw SV, *et al*. A systematic approach towards the development of quality indicators for postnatal care after discharge in Flanders, Belgium. *Midwifery* 2017;48:60–8.
50 Boulkedid R, Abdoul H, Loustau M, *et al*. Using and reporting the Delphi method for selecting healthcare quality indicators: a systematic review. *PLoS ONE* 2011;6:e20476.
51 de Bekker PJGM, de Weerdt V, Vink MDH, *et al*. 'Give me something meaningful': GPs perspectives on how to improve an audit and feedback report provided by health insurers - an exploratory qualitative study. *BMJ Open Qual* 2022;11:e002006:1–9:.
52 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.
53 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.
54 Payne VL, Hysong SJ. Model depicting aspects of audit and feedback that impact physicians' acceptance of clinical performance feedback. *BMC Health Serv Res* 2016;16:260.
55 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;22:113.
56 Maier B, Wagner K, Behrens S, *et al*. Comparing routine administrative data with registry data for assessing quality of hospital care in patients with myocardial infarction using deterministic record linkage. *BMC Health Serv Res* 2016;16:605.
57 Avoundjian T, Gidwani R, Yao D, *et al*. Evaluating Two Measures of Lumbar Spine MRI Overuse: Administrative Data Versus Chart Review. *J Am Coll Radiol* 2016;13:1057–66.
58 Eindhoven DC, van Staveren LN, van Erkelens JA, *et al*. Nationwide claims data validated for quality assessments in acute myocardial infarction in the Netherlands. *Neth Heart J* 2018;26:13–20.
59 Foundation healthcare evaluation netherlands (dutch: stichting zorgevaluatie nederland), "leading the change," leading the change: about us. n.d. Available: https://zorgevaluatienederland.nl/programs/8f056ae2-45f7-4dcc-a57d-1906ef1b02f0