

RESEARCH ARTICLE

Effectiveness of clinical audit and feedback in improving healthcare professionals' practice in primary healthcare settings: A systematic review

Mrs. Badriya AL Musharfi

Master of Administration in Nursing, Health Cluster; Directorate of General Health Services, South Sharqiya Governorate, Ministry of Health, Sultanate of Oman

Corresponding Author: Mrs. Badriya AL Musharfi, **E-mail:** badriyanasser1@gmail.com

ABSTRACT

Clinical audit and feedback (A&F) are one of the most widely applied quality-improvement strategies for changing healthcare professional behaviour. It entails methodical appraisal of clinical practice versus a clear standard, benchmark or peer comparator, and then feedback that is aimed to provoke reflection, objective and corrective action. It is particularly relevant in primary healthcare where the bulk of prescribing, investigation, preventive and chronic disease management is practised and where relatively small changes in professional behaviour have major population impacts. This systematic review critically assesses the modern evidence of effectiveness of clinical audit and feedback on the practice of healthcare professionals in the primary healthcare setting with special attention to the articles published in the period between 2022 and 2025 considered in the corpus of references used in this review. The review was designed following PRISMA 2020 guidelines. Bibliographic records were initially availed (n=40); ten of them were replicated creating 30 reports. The primary narrative synthesis included sixteen key effectiveness studies and six systematic reviews or meta-analyses and eight qualitative, implementation, design or consensus studies were left as the contextual evidence. Another pooled meta-analysis was not methodologically justified because of the large differences between the results, in terms of clinical domain, unit of analysis, intensity of intervention, and statistical measurement. The data confirm that A&F can most likely lead to small-to-moderate alterations in professional practice but is not consistent or automatic. In a primary care-based trial involving 5,046 primary care physicians in Ontario, the feedback to the physicians decreased overall antibiotic prescribing by about 5 percent compared to control, unnecessary antibiotic prescribing by 11 percent, prescriptions greater than seven days by 15 percent and broad-spectrum prescribing by about 6 percent (Schwartz et al., 2024). Personalised peer-comparison feedback with therapeutic recommendations in British Columbia with appropriate first-line nitrofurantoin treatment of uncomplicated lower urinary tract infection resulted in an odds ratio of 1.28 (95% CI 1.07-1.52) in 5,073 family physicians (Carney et al., 2025). A trial with high-imaging general practitioners in Australia reported an adjusted decrease in musculoskeletal imaging requests of 2.66 requests per 1,000 visits at 12 months (O'Connor et al., 2022). Whereas in South African community primary-care centres, peer audit and feedback boosted adherence to antimicrobial prescribing quality indicators (19% to 47%), it also resulted in a 19.3% reduction in antibiotic use (de Vries et al., 2022). On the other hand, a Swiss multi-centre randomised trial of 3,426 physicians reported no essential change in total antibiotic prescribing with quarterly feedback on personalised rates and a British Columbia trial reported no significant impact of feedback on initial opioid prescribing (Aghlmandi et al., 2023; McCracken et al., 2023). Similarly, an A&F trial of patient-safety in Spanish, 59 primary healthcare centres and 1,053 professionals did not demonstrate any significant change in safety culture in which the implementation was mostly disrupted by the COVID-19 pandemic (Fiol-deRoque et al., 2025). These results demonstrate that the effectiveness is determined by the design of interventions, baseline performance, actionability, clinician engagement, implementation fidelity, peer comparison, contextual barriers and the amount of control clinicians have over the target behaviour. These findings are in line with the wider evidence. In Cochrane review by Ivers et al., the median of the desired professional practice was observed to improve by 2.7 percent and the model-based mean of the desired improvement by 6.2 percent but in the primary-care antibiotic meta-analysis by Xu et al., the percentage change in total antibiotic prescribing, unnecessary initiation, excessive duration and broad-spectrum antibiotic choice were all reduced by 11, 23, 13 and 17 percent respectively. However, the heterogeneity was high and that implies that there is no average effect that can be applied to explain performance in various settings. Clinical A&F is not in general to be regarded as a stand-alone information-delivery endeavour but as a behaviour-change system. Its most powerful impacts are those in which feedback is centred around a significant performance gap, where it contains plausible individualised data, provides an actionable comparator, fosters interpretation,

allows action planning and is situated in an environment that has the capacity to make change. The evidence appears to support the idea of A&F used regularly and thoughtfully in primary care and caution against the assumption that dashboards, mailed reports or repeated measurement will only improve care.

KEYWORDS

Clinical Audit, Feedback and Improving health care

ARTICLE INFORMATION

ACCEPTED: 7 September 2026

PUBLISHED: 09 October 2026

DOI: 10.32996/jmhs.2026.7.10.6

1. Introduction

Audit and feedback (A&F) has become a quantitatively validated implementation strategy to alter professional behaviour in primary healthcare, but its effect size differs markedly in response to the design of interventions, baseline performance and clinical background. The most rigorous state-of-the-art synthesis was able to identify 292 randomised studies and reported a median absolute improvement of 2.7 percent in wanted professional practice, a model based mean absolute improvement of 6.2 percent and odds ratio of 1.47 suggesting a statistically significant but heterogeneous overall effect (Ivers et al., 2025). In primary-care antibiotic stewardship, a meta-analysis of 56 randomised trials identified an 11% relative decrease in total antibiotic prescribing, a 23% decrease in inappropriate initiation, a 13% decrease in excess duration of treatment and a 17% decrease in broad-spectrum antibiotic choice although there was a large degree of heterogeneity, with I^2 values between 72 and 97 (Xu et al., 2025). Large pragmatic trials also have variation in their effectiveness. Schwartz et al. (2024) demonstrated a relative reduction of antibiotic prescribing by 5% among 5,046 physicians and Carney et al. (2025) had odds ratio of 1.28 between first-line nitrofurantoin prescribing and increased first-line nitrofurantoin prescribing. Likewise, O'Connor et al. (2022) found 2.66 reduced musculoskeletal imaging requests per 1,000 consultations after individualised feedback. These findings justify a critical systematic review of the magnitude, frequency and predictors of A&F effectiveness in contemporary primary care.

1.1 Background

The point of healthcare decision making where a large percentage of the population-level decisions are taken is in primary healthcare. Repeated decisions of general practitioners, family physicians, pharmacists, nurses and multidisciplinary primary-care teams are concerned with prescribing, diagnostic testing, screening, chronic disease management, preventive measures, referrals and treatment follow-up. Since these choices are made at a high rate, seemingly small deviations of the evidence-based practice may add up to significant clinical, economic and public-health effects. On the other hand, even a minor increase in the adherence to recommended practice will result in significant benefit with implementation in thousands of clinicians and millions of consultations. The audit and feedback has thus emerged as a significant implementation and improvement on quality strategy to minimize unwarranted clinical variation.

A&F is operationally complex yet conceptually simple. The audit element entails valid measurement of professional practice over a given period, which is usually done through the retrieval of electronic health record data, claims data, prescription records, chart-review data or patient-reported data. The performance is then compared to a clear standard like a guideline standard, past personal performance, population standard, average peer performance or high-performing peer group. This comparison is then conveyed to the concerned clinician or team of clinicians by the feedback component with a view to inducing behaviour change. The contemporary interventions can be provided by means of letters, dashboards, electronic messages, pharmacist consultation or by means of face-to-face discussion or a combination of these means. Educational resources, suggestions, patient-specific lists, action steps, behavioural prompts or information about peer-comparison can also be included in the feedback (Ivers et al., 2025; Tsang et al., 2022; Schwartz et al., 2023).

Clinical audit should not however be equated to data collection. The lack of interpretation as well as the lack of opportunity to take corrective action makes measurement without interpretation a complete improvement cycle. Similarly, inaccurate, untimely, or incomprehensible feedback or feedback that does not relate to actions that a clinician can control might have little behavioural impact. The contemporary theory concentrates on A and F as a process that has measurement, disparity perception, acceptance of the information, intention to act, prioritisation of the disparity and taking the corrective action. Any malfunction in each of these stages can water down performance. This interpretation is supported by qualitative studies. Irish GPs receiving feedback on postal antibiotic-prescribing feedback were generally satisfied with the intervention, but they doubted the accuracy of the information, the perceived effect, and data processing using reports (Roche et al., 2022). Canadian family physicians also claimed that technically better reports were also challenging to utilize where measures were viewed as invalid, irrelevant or beyond the control of the physician as well as when time, workflow and quality-improvement capacity were constrained (Rouleau et al., 2023).

Contemporary evidence base is particularly pertinent as A&F has transformed significantly in the field of digitalisation. Administrative datasets and electronic medical records and individualised reports can be generated in a short period of time

and thousands of clinicians can be automatically measured. Andrade et al. (2022) have provided an example of how Australian general practitioners could be given a national-scale digital intervention in A&F, which achieved an equivalent 11.2% reduction in the dispensing of gabapentinoids over 12 months, as with postal delivery, and a higher rate of psychologist initiation was achieved with digital feedback. It is worth noting that the primary finding of the study was the relative effectiveness of delivery mode and not the causal effectiveness of A&F and no feedback since both trial arms were subjected to a certain level of A&F. That is why it is necessary to interpret the results critically: the statistically significant pre-post change does not imply that the change was due to A&F itself.

There is evidence that A&F is likely to have a positive impact on practice, obtained using systematic-reviews, but the mean effects conceal significant quantities of variability. Ivers et al. (2025) summarised 292 trials in the healthcare settings. Median change of 2.7% was found in 177 studies that used dichotomous results of professional-practice with an interquartile range of 0.0% to 8.6% and meta-analysis demonstrated a desired change of 6.2% in practice and odds ratio of 1.47. The variability of these effects was enormous and the focus was that effectiveness is primarily context and intervention design-dependent. Xu et al. (2025) identified 56 randomised trials in the specific area of antibiotic prescribing and discovered reductions in the total amount of antibiotic prescribing (11%), unnecessary initiation (23%), longer prescriptions and non-selective selection (13 and 17% respectively). However, heterogeneity was between $I^2=72$ and 97, once again proving that pooled effectiveness is not synonymous with guaranteed effectiveness of a single primary-care programme.

1.2 Contemporary Importance of Audit and Feedback in Primary Care

One of the most exemplary examples of the significance of primary-care behaviour is the prescribing of antimicrobials, where de Vries et al. (2022) found that in the antimicrobial stewardship context, approximately 80% of antimicrobial use was in primary care. A study of 2,077 prescriptions of their ten-centre Cape Town intervention found that only 19% of their prescriptions were able to baseline adherence to indicators of antibiotic quality. This capacity to enhance such prescribing is not only significant to the quality of treatment provided to individuals but also since inappropriate exposure to antibiotics is a cause of antimicrobial resistance. Equally, Schwartz et al. (2024) assessed over 5,000 Ontario primary-care physicians, demonstrating that feedback grounded on regularly gathered administrative data can be utilized at health-system scale and not just in small demonstrations.

A&F can be used in a non-antibiotic-stewardship setting as well. O'Connor et al. (2022) examined unnecessary musculoskeletal imaging, Vicens et al. (2022) examined benzodiazepine prescribing, McCracken et al. (2023) and Lacroix et al. (2025) examined opioid prescribing, Gillespie et al. (2023) examined recognition and response to the social needs of patients, Hocking et al. (2022) examined maintenance of chlamydia testing, Vanstone et al. (2022) examined documentation of indications of antimicrobial prescription, and Fiol-deRoque et al. (2025) tested a multicomponent A&F strategy in maternal and child health care in 154 primary-health facilities in Mozambique. Reduction of low value care, more recommended care, safer prescribing, better documentation, preventive testing, patient safety and health-system preparedness are, therefore, present in current A&F studies.

1.3 Rationale for the Review

The possibility or impossibility of feedback to affect professional behaviour or not is no longer the key question. The question of greater importance is in what circumstances clinically meaningful improvement takes place. There are positive trials and null trials, even where both trials seem to be using apparently sophisticated personalised feedback. Indicatively, Schwartz et al. (2024) demonstrated that the rate of antibiotic prescribing in Canada was greatly decreased, whereas no significant decrease in personalised quarterly antibiotic-prescribing feedback in Switzerland was found by Aghlmandi et al. (2023). Similarly, O'Connor et al. (2022) observed that imaging decreases were evident in high-requesting general practitioners and McCracken et al. (2023) observed that personalised opioid-prescribing feedback did not significantly influence opioid initiation. These differences can only be understood through looking at the baseline performance, selecting recipients, specificity of behaviour, feedback comparator, credibility, intensity, co-interventions and implementation conditions.

The second reason why an exclusive review was necessary in the current situation is the flourishing advancement of electronic feedback and dashboards. Tsang et al. (2022) identified 88 studies on 65 electronic A&F systems but found out that existing systems only applied eight out of 18 recommended feedback practices in the middle. Action planning and near-real time information-providing were only present in 20 and 32 systems respectively. Xie et al. (2022) also found similar results, as they reported inconsistent results of standalone clinical dashboards but more consistent positive changes with multicomponent dashboard interventions. The review of 2025 primary-care dashboard by Meier et al. identified six studies and found that four multifaceted dashboard interventions showed better performance than at least one primary outcome, and standalone dashboard interventions showed mixed effects. These lessons suggest that digitalisation helps to provide feedback, but does not eradicate behavioural and organisational conditions to change.

1.4 Review Aim

The aim of the systematic review is to critically examine the effectiveness of clinical audit and feedback in improving the practice of healthcare professionals in primary healthcare facilities on the basis of the existing evidence published within the past two to five years in the recommended evidence corpus.

1.5 Research Question

The principal research question is: among healthcare professionals working in primary healthcare, how effective are clinical audit and feedback interventions, compared with usual care, no feedback, alternative feedback strategies or baseline practice, in improving objectively measured professional practice?

The size and consistency of effects across prescribing, investigation, preventive care, documentation, patient safety and service-delivery outcomes; the characteristics of successful and unsuccessful interventions; the quality and limitations of the evidence and implications of the review to future primary-care quality-improvement programmes are also discussed in the review.

2. Systematic Review

2.1 Overall Effectiveness of Audit and Feedback

Recent evidence indicates that audit and feedback (A&F) can improve the clinical practice of medical practitioners, albeit to a varying degree and dependability, depending upon the situation, behaviour and intervention packages. The Ivers et al. (2025) Cochrane review involved 292 randomised studies (678 randomised intervention arms) and is the most powerful modern synthesis of A&F efficacy. Among 177 studies that reported 558 dichotomous outcomes in professional practice, the median improvement (absolute) with A&F was 2.7 (IQR 0.0 to 8.6). The mean absolute change is estimated by model-based analysis to be 6.2 (95% CI 4.1–8.2) which is an odds ratio of 1.47 (95% CI 1.31–1.64) of desired professional practice. However, the quality of methodology was uneven, approximately 46% of the studies with low risk of bias and 39% with high risk proved that the effect estimates must be considered along with a high level of methodological and clinical heterogeneity (Ivers et al., 2025).

2.2 Effectiveness in Antibiotic Prescribing

The best evidence of primary-healthcare is evidence on antimicrobial prescribing. Synthesising 56 randomised controlled trials which were carried out in 16 countries, Xu et al. (2025) statistically showed significant improvements in various prescribing dimensions. A&F reduced overall antibiotic prescribing by 11% (RR 0.89, 95% CI 0.84–0.95; $I^2=97\%$), unnecessary antibiotic initiation by 23% (RR 0.77, 95% CI 0.68–0.87; $I^2=72\%$), excessive treatment duration by 13% (RR 0.87, 95% CI 0.81–0.94; $I^2=86\%$), and broad-spectrum antibiotic selection by 17% (RR 0.83, 95% CI 0.75–0.93; $I^2=96\%$). Even though the combined effects are always in favour of A&F, the values of heterogeneity of 72 to 97% are quite high and cannot be explained by sampling error only. The variations in the baseline prescribing, feedback frequency, choice of comparators, healthcare systems, and co-interventions are thus likely to affect the effectiveness (Xu et al., 2025).

2.3 Evidence from Large Pragmatic Trials

Big pragmatic trials are useful evidence of the potential of statistically significant impacts on a healthcare-system level. Schwartz et al. (2024) evaluated 5,046 primary-care physicians and discovered the rate of antibiotic prescribing 56.0 per 1,000 patient visits in physicians receiving feedback compared to 59.4 per 1,000 patient visits in controls (relative rate 0.95 (95% CI 0.94–0.96)). The reduction in the number of unnecessary antibiotics prescribed by 11% (RR 0.89), the duration of prescribed antibiotics over seven days by 15% (RR 0.85), and broad-spectrum antibiotics by 6% (RR 0.94) was also related to the intervention (Schwartz et al., 2024). The importance of these findings is that even relatively small impacts on individuals at the individual level can result in substantial changes in the population level when interventions are implemented in thousands of clinicians.

Carney et al. (2025) randomised 5,073 family physicians the same way and assessed 21,307 episodes of uncomplicated lower urinary tract infection. Peer-comparison feedback and therapeutic advice based on individualised feedback were significantly found to positively influence the correct first-line prescription of nitrofurantoin, odds ratio, 1.28 (95% CI 1.07–1.52). The findings suggest that A&F will be able not only to calculate the amount of prescriptions but also to define the suitability of its therapy and a congruity of the guidelines.

2.4 Effects on Diagnostic Testing and Clinical Decision-Making

Its efficacy may be demonstrated not only in the prescription of medications. O'Connor et al. (2022) randomised 3,819 Australian general practitioners who had a relatively high musculoskeletal imaging-request rate. Targeted imaging was done at 12 months at a rate of 27.7 requests per 1,000 consultations and 30.4 per 1,000 consultations in the intervention group and the controls respectively. This represented an adjusted mean difference of -2.66 requests per 1,000 consultations (95% CI -3.24 to -2.07 ; $p<0.001$). Even though the relative difference is rather small, having it implemented in a high number of consultations can lead to significant cuts in the potentially unnecessary investigation (O'Connor et al., 2022).

2.5 Evidence from Resource-Constrained Primary Care

A&F has demonstrated potential in primary-care environments with limited resources as well. de Vries et al. (2022) compared a peer-based antimicrobial stewardship intervention in Cape Town community healthcare centres. The proportion of compliance with quality measures of antimicrobial prescribing changed with an absolute change of 19 to 47 percent ($p=.001$). There was also a decrease in the use of antibiotics by 19.3%. However, since this is not a typical individually randomised controlled trial, secular trends and quality-improvement efforts concurrently cannot be completely excluded to account for some of the observed effect (de Vries et al., 2022).

2.6 Evidence of Limited and Null Effects

Although there are positive aggregate results A&F is not always effective. The McCracken et al. (2023) studies are not statistically significant in the effects of personalised prescribing feedback on opioid analgesics initiation in opioid-naïve patients. The difference-in-percent-differences was -2.1% (95% CI -4.4% to 0.3%), while the odds ratio for initiating any opioid was 0.98 (95% CI 0.96–1.01). The results fail to provide strong evidence on the efficacy of the intervention on this outcome, as both the null value fell within the confidence intervals (McCracken et al., 2023).

Similarly, Aghlmandi et al. (2023) randomised 3,426 Swiss primary-care physicians who were given personalised feedback in quarterly doses of antibiotic prescribing or usual care. The median annual prescribing of the intervention group ($n=8.2$) was marginally less than the controls ($n=8.4$) and the difference in prescribing change between the two groups was only -0.1% (95% CI -1.2 to 1.0). The results indicate that repeated and individualised feedback does not necessarily result in clinically significant behavioural change (Aghlmandi et al., 2023).

2.7 Heterogeneity and Determinants of Effectiveness

The degree of statistical heterogeneity in systematic reviews is high indicating that the effectiveness of A&F differs according to the methods and sites of its use. In situations where there is much room to improve performance relative to the baseline, where feedback is provided relative to behaviour due to direct influence of clinicians, where performance is compared with meaningful peers or benchmarks, where recommendations can be implemented, and where feedback is facilitated by education or organisation, larger effects are generally observed (Ivers et al., 2025; Schwartz et al., 2023). Passive data provision, on the other hand, can prove to be inadequate in cases where clinicians doubt the integrity of the data, have limited time/resources, or are unable to easily convert the identified gaps into specific interventions (Rouleau et al., 2023).

2.8 Critical Synthesis

Overall, it is possible to suggest A&F as statistically effective, but extremely context-specific intervention to improve the practice of primary healthcare. Antibiotic stewardship and the chosen diagnostic behaviours have the best evidence, and the opioid-prescribing studies show that none of the clinical outcomes can be assumed to have benefits. There is high between-study heterogeneity, a risk of bias is variable, and opposite trial outcomes suggest that A&F cannot be viewed as a homogeneous intervention with a constant effect size. Rather, its effectiveness is a result of the interaction of baseline performance, intervention design, data credibility, peer comparison, behavioural specificity, implementation fidelity and organisational capacity. The most suitable way of conceptualising A&F, therefore, is as a structured behaviour-change strategy, as opposed to the mere transference of performance statistics.

Table 2. Systematic Reviews and Meta-Analyses Supporting the Review

Systematic Review	Evidence Base	Main Statistical Finding	Critical Interpretation
Ivers et al. (2025)	292 randomised studies	Median $+2.7\%$; modelled $+6.2\%$; OR 1.47	A&F improves practice, but effects vary substantially
Xu et al. (2025)	56 RCTs; 16 countries	Antibiotic prescribing reduced 11%; unnecessary initiation 23%	Strong benefit, but $I^2=72-97\%$
Carter et al. (2023)	11 cluster RCTs; nine countries	Six-study pooled RR 0.78 (95% CI 0.64–0.94)	Pharmacist-supported feedback improves prescribing
Meier et al. (2025)	Six primary-care dashboard studies	Four multifaceted interventions improved primary outcomes	Multifaceted dashboards outperform standalone approaches

Systematic Review	Evidence Base	Main Statistical Finding	Critical Interpretation
Tsang et al. (2022)	88 studies; 65 electronic systems	Median eight of 18 best-practice features	Digital availability alone does not ensure effectiveness
Xie et al. (2022)	11 randomised dashboard trials	Standalone dashboards mixed; multicomponent interventions stronger	Actionable supporting components appear important

The strongest aggregate evidence comes from Ivers et al. (2025), where the median absolute improvement was 2.7% and the model-based improvement was 6.2%, and Xu et al. (2025), where antibiotic outcomes improved by 11–23%, despite substantial heterogeneity. Pasted text Carter et al. (2023) additionally reported a pooled RR 0.78 (95% CI 0.64–0.94) for pharmacist-involved prescribing interventions.

3. Methods

3.1 Review Design and PRISMA 2020 Approach

This review was organized in the principles of PRISMA 2020 to report on evidence identification, screening, eligibility assessment, synthesis and interpretation transparently. One significant methodological flaw is that the review is based on a pre-determined set of references, which are included in the assignment, and does not claim to have searched a de novo bibliographic database, which was not searched. This difference is needed by methodological integrity. The review can therefore be described as a selectively screened, critically reviewed and synthesised narrative review of a pre-defined body of existing evidence in comparison to an uninhibited database search of all the literature that could be eligible.

The provided corpus had 40 bibliographic entries. Ten of them were duplicates of known references and 30 single publications were kept. The 30 reports were narrowed down according to preset eligibility criteria. There were six major quantitative analyses that could be included in the major effectiveness synthesis. Higher-level corroborative evidence kept six systematic reviews or meta-analyses: Carter et al. (2023), Ivers et al. (2025), Meier et al. (2025), Tsang et al. (2022), Xie et al. (2022) and Xu et al. (2025). Eight were mostly qualitative, intervention-development, process-evaluation or consensus/guidance and were retained to describe mechanisms and implementation: Arsenault-Lapierre et al. (2022), McClatchey et al. (2024), Roche et al. (2022), Rouleau et al. (2023), Schwartz et al. (2023), Shuldiner et al. (2023), Shuldiner et al. (2024) and van der Winden et al. (2023).

3.2 Eligibility Criteria

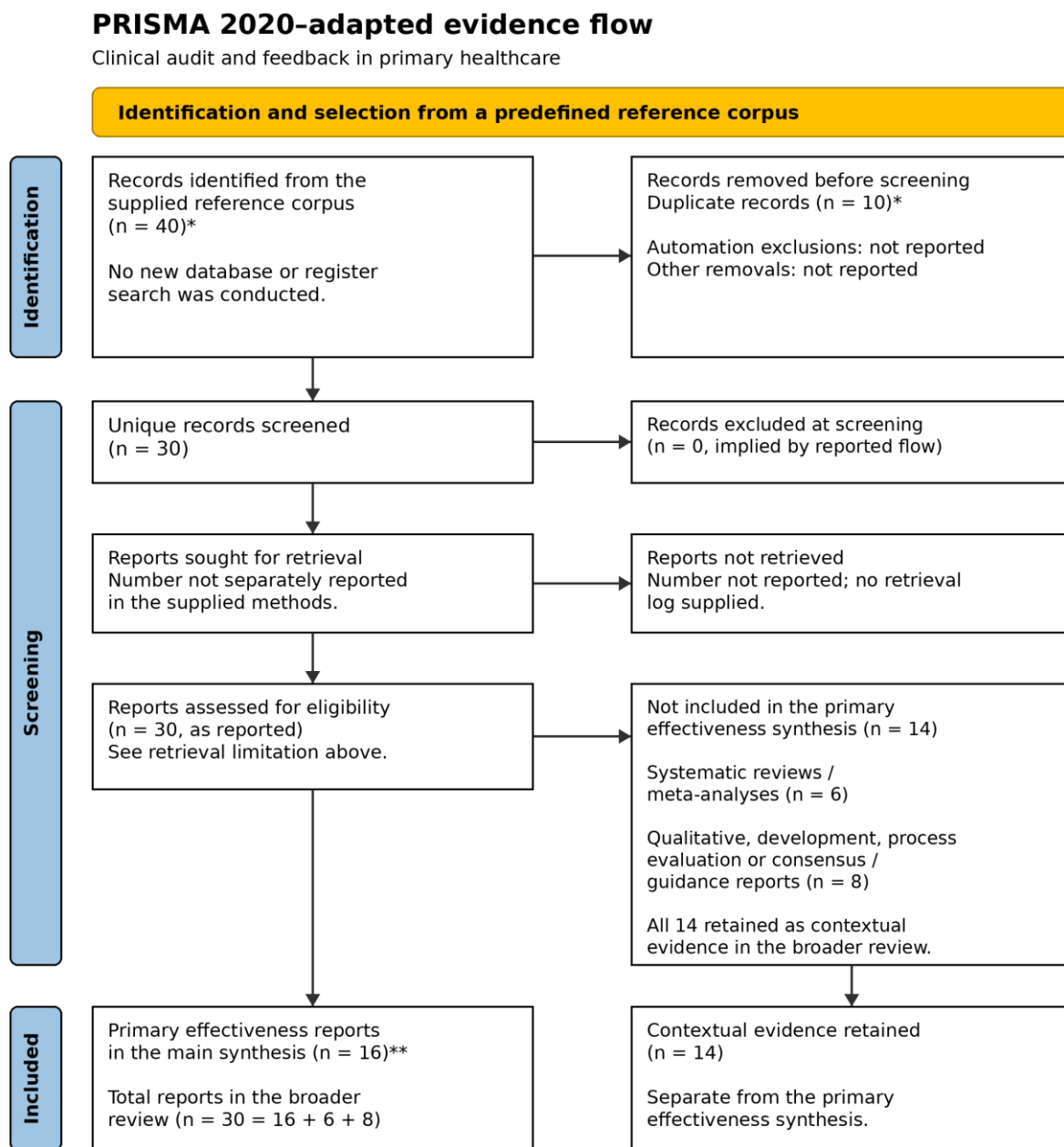
Population, intervention, comparator, outcomes and study design were used as a basis to design eligibility. Professionals working in healthcare and general practice or family medicine and community primary-care clinics or other first-contact primary-healthcare locations were included in the sample. The major scope did not include inpatient hospital-based studies unless it had a complementary systematic review that provided contextual information that could be applied in primary care. Interventions which were eligible needed to have a recognisable A&F component. Audit was defined as planned measurement of clinician, team or facility performance, whereas feedback needed to convey the performance outcomes to the professional or team. The use of other elements like education, peer comparison, academic detailing, reminders, facilitated discussion or action planning could also be considered as interventions since contemporary A&F is often a multicomponent implementation strategy.

Ordinary care, no feedback, delayed feedback, other types of feedback, feedback withdrawal/baseline practice of interrupted time-series and controlled/uncontrolled quality-improvement experiments were comparators. Primary outcomes were connected with the changes in objective measures of professional behaviour or practice, including the volume or suitability of prescribing, ordering diagnostic tests, following guidelines, screening or testing, clinical documentation, managing social needs and providing evidence-based care. The patient-safety, service preparedness and patient-level outcomes culture were seen as a secondary one because professional practice was the main focus. Randomised controlled trials, cluster randomised trials, factorial trials, pragmatic trials, quasi-experimental studies, and interrupted time-series evaluations, controlled cohort studies and quality-improvement studies that are appropriately analysed could be included in the primary effectiveness synthesis. Causal effectiveness estimation did not involve qualitative research, intervention-development studies or consensus statements, but was a part of explanatory implementation synthesis.

3.3 Study Selection and PRISMA 2020 Flow

Forty provided bibliographic records were included in the PRISMA 2020 identification stage. Ten duplicate entries were filtered out before screening and 30 unique reports were received. All the 30 different reports were subject to relevancy assessment. The key synthesis involved sixteen key effectiveness studies that passed the test. The main synthesis of effectiveness did not include fourteen reports since they were systematic reviews, qualitative studies, process evaluations or intervention-development studies or consensus/guidance documents, but they were not omitted in the review as a whole but were included to be interpreted in context and to be mechanistic. Through this, the PRISMA flow in relation to this limited corpus can be summarized as 40 records identified, 10 duplicate records removed, 30 unique reports screened, 30 reports in the broader review, 16 primary effectiveness studies included in the main effectiveness synthesis and 14 reports as contextual or secondary evidence. This approach prevents the presence of two studies that count as primary studies and are also present in the given systematic reviews.

Figure 1:



Accuracy and reporting notes

* The draft reports 40 initial records and 10 duplicates. Only the final 30 unique references were supplied; the earlier counts cannot be independently verified.

** The 16 / 6 / 8 allocation follows the supplied review. Section 3.1 says "six" primary analyses; this conflicts with Section 3.3 and should read "sixteen". Counts refer to reports, not independently verified unique studies.

Adapted from PRISMA 2020: Page et al. (2021), BMJ 372:n71. doi:10.1136/bmj.n71 (CC BY 4.0).

This diagram documents the supplied corpus; it does not establish a comprehensive systematic search.

3.4 Data Extraction

The data were restricted based on the year of publication, the country, setting, professional population, sample size, unit of randomisation or analysis, nature of the intervention, the delivery of feedback, comparator, follow-up, primary and secondary outcome, effect estimates, confidence, statistical significance, implementation fidelity and limitations in the causal interpretation. Special consideration was paid to whether the outcomes were a measure of absolute change or relative change since the confusion of the two can overstate the clinical impact.

Where appropriate, rate ratios, odds ratios, absolute differences of the means, percentage point differences and confidence intervals were left in their published form instead of being converted into a different form without reason. This was important,

as the underlying behaviours which were measured in the original studies varied. An odds ratio that describes the choice of nitrofurantoin to treat urinary tract infection, e.g., cannot be directly replaced by the change in the number of imaging requests per 1,000 consultations or a change in the number of defined daily doses of benzodiazepines.

3.5 Risk of Bias and Critical Appraisal

Randomised and cluster-randomised trials were conceptually appraised when compared to domains that are consistent with Cochrane Risk of Bias 2 framework such as randomisation, non-use of intended intervention, missing outcome data, outcome measurement and selective reporting. Special consideration was paid to cluster allocation and possible imbalance of the baselines. Most of the trials of prescribing and test-ordering, which were routinely done, used routinely collected administrative or electronic data that reduced the likelihood of observer-based outcome measurement bias even though participating clinicians could not be blinded.

The evaluation of non-randomised studies was conducted in accordance with the principles that are similar to ROBINS-I, especially confounding, selection of participants, the classification of interventions, deviation of interventions, missing data and outcome measurement. Prior to quality-improvement assessment before and after without concurrent controls was taken with caution since secular trends, regression to the mean, policy modifications and current interventions can result in remote enhanced results without A&F.

Such a difference is so remarkable within the opioid evidence. Lacroix et al. (2025) noted high-risk opioid prescribing pre-existing decreases in intervention and comparison groups. These secular trends may not have been modeled and this could have erroneously attributed the decline to A&F. They discovered that there was no significant overall incremental effect post-implementation but high-volume prescribers who received A&F had a significant incremental change in slope. This paper explains the reason behind not making causal assertions based on improvement following feedback.

3.6 Synthesis Strategy

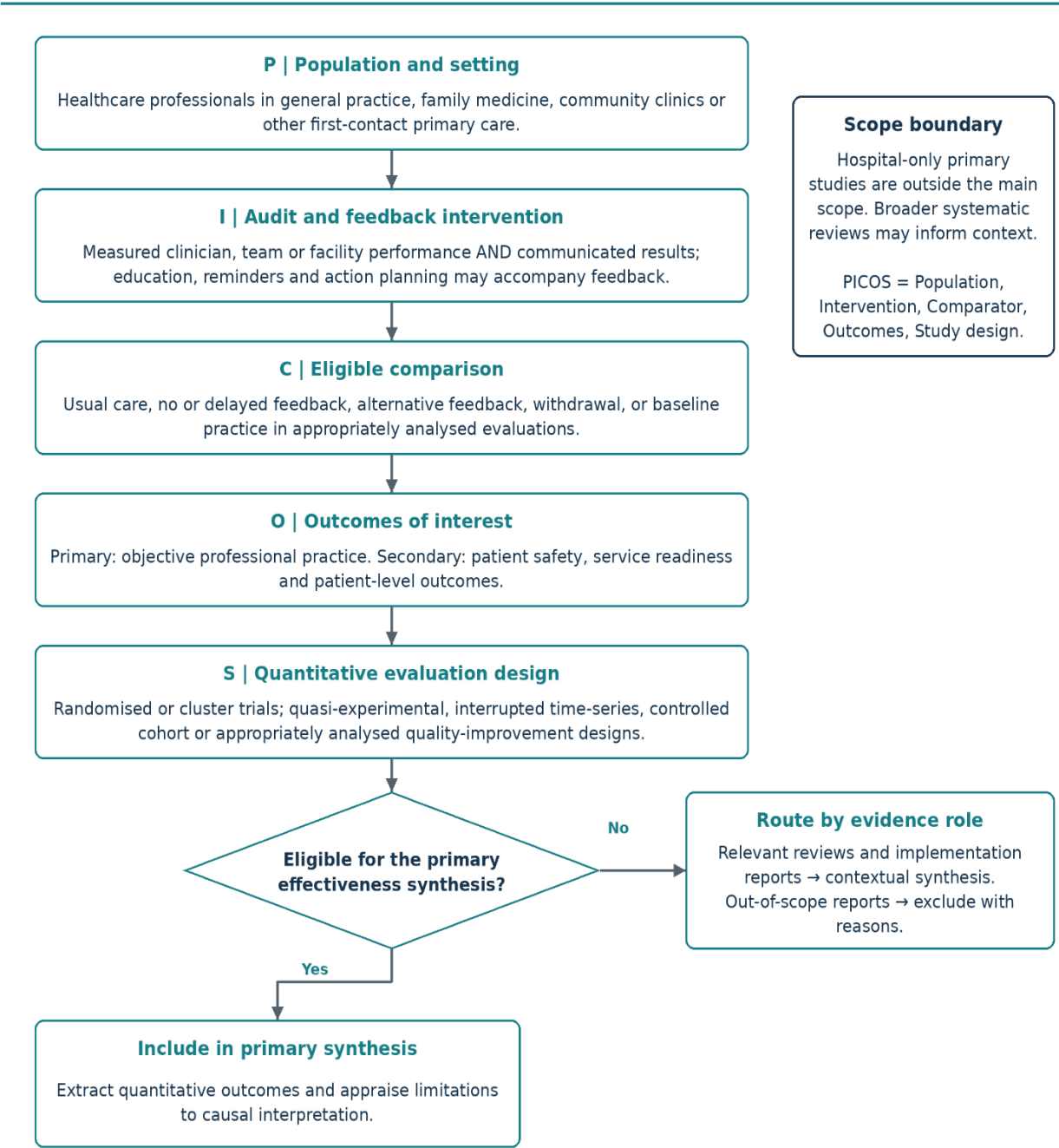
The reason to use a narrative synthesis was that a new pooled estimate was not calculated. There was a high level of clinical and methodological heterogeneity involved in this decision. The primary studies dealt with various behaviours and employed various intervention structures, comparators, populations and outcome scales. Interventions were made to attempt to decrease inappropriate prescribing and some attempted to increase recommended testing and there were also interventions aimed at attempting to better documentation or to better identify social needs. The interventions included only A&F, education, academic detailing, peer discussion, dashboards or action planning. The overall impact of the sum of these effects in a single new numerical estimate would then be a statistically correct but clinically ambiguous result.

In cases where authors had already arrived at the conclusion that there was enough similarity to combine particular outcome categories, meta-analytic outcomes were based on systematic reviews provided. In particular, the review will use Xu et al. (2025) to present the up-to-date antibiotic-prescribing pooled estimates, Carter et al. (2023) to present the pharmacist-involved A&F, and Ivers et al. (2025) to present the broad professional-practice effects.

Figure 2: PICOS Eligibility:

Eligibility Assessment Using PICOS

Operational representation of the eligibility criteria stated in Section 3.2; this is a decision framework, not a record-level screening log.



Source: (Ivers et al., 2025); (Xu et al., 2025)

4. Results

4.1 Overview of the Evidence Base

The 16 main effectiveness studies exhibit a good geographical and clinical diversity. The studies were carried out in Canada, South Africa, Mozambique, Pakistan, Spain, Switzerland, United States and Australia. The study sizes ranged between retrospective evaluation of 61 primary healthcare physicians in Karachi to pragmatic trials that comprised of several thousand

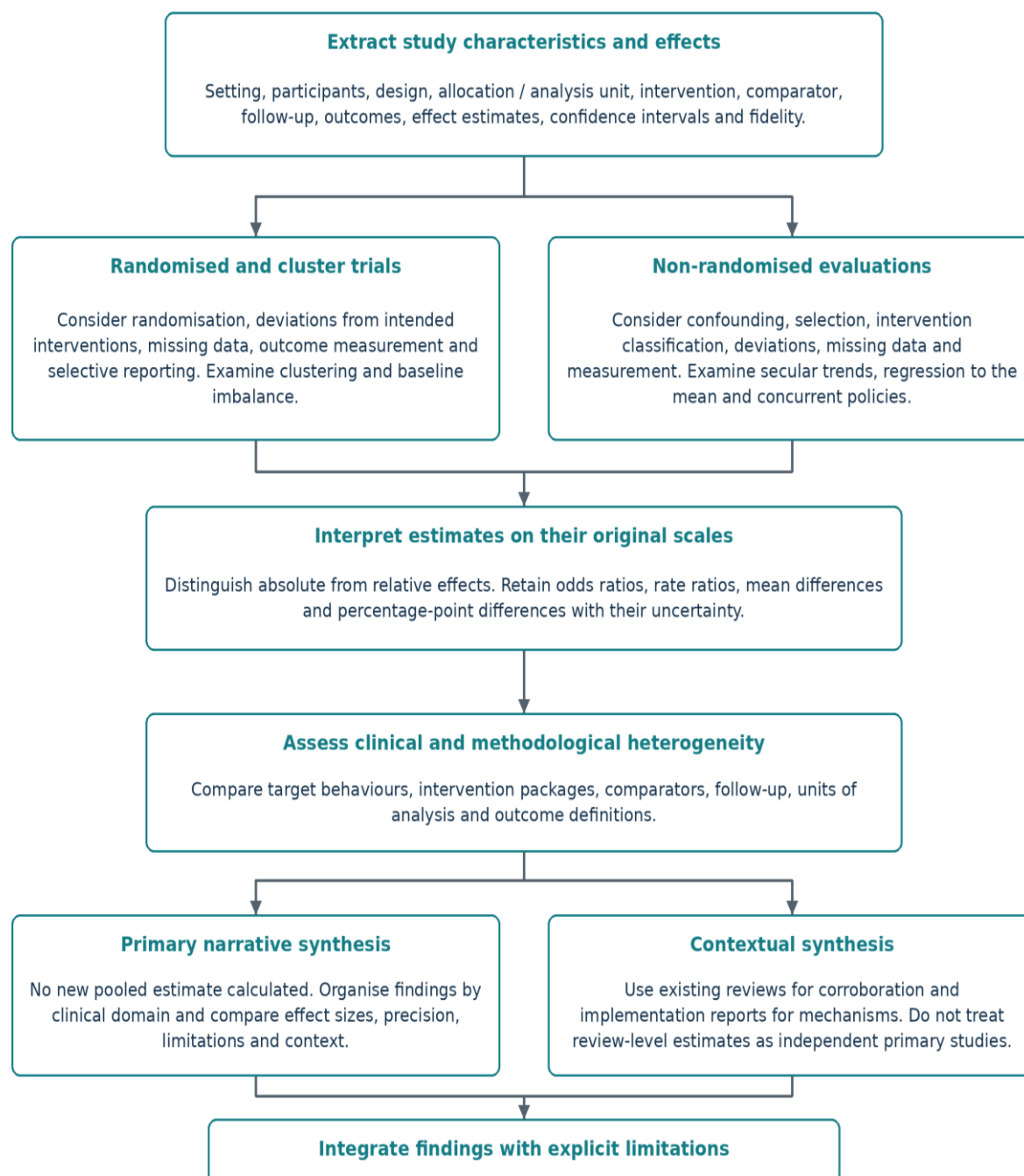
family physicians. Carney et al. (2025) randomised 5,073 family physicians, Schwartz et al. (2024) analysed 5,046 physicians, O'Connor et al. (2022) randomised 3,819 Australian general practitioners, and Aghlmandi et al. (2023) randomised 3,426 Swiss primary-care physicians. Dinis et al. (2025) evaluated 154 primary health facilities within 12 districts in Mozambique in their study.

The primary behavioural objectives were; medication prescribing, ordering diagnostic tests, preventive testing, clinical documentation, social-need assessment, patient safety and delivery of maternal-child health services. Prescribing was the most dominant area, as it is congruent with the greater body of literature. Ivers et al. (2025) found that prescribing was approximately 47 percent of the professional practices appraised in the broader A&F evidence base and test ordering was approximately 35 percent.

Figure 3: Appraisal and Evidence:

Critical Appraisal and Evidence Synthesis

Analytical pathway described in Sections 3.4–3.6; conceptual appraisal is distinguished from a completed formal risk-of-bias assessment.



Source: (Aghlmandi et al., 2023); (Carney et al., 2025)

4.2 Antibiotic Prescribing

Some of the most homogeneous, and some of the most heterogeneous, modern evidence is found in antibiotic stewardship. Schwartz et al. (2024) carried out a large pragmatic factorial randomised trial in Ontario primary-care doctors who prescribe to patients aged 65 years and older. Five thousand four hundred six physicians participated in the study, 4,041 of which received mailed peer-comparison feedback intervention and 1,005 served as controls. The mean prescription rate of antibiotics at six

months was 56.0/1000 patient visits in the intervention group and 59.4 in the controls with a relative rate of 0.95 (95% CI 0.94-0.96). More to the point, the impacts were not limited to total prescribing. The unnecessary prescribing rate was 0.89, long-lasting prescription (over seven days) was 0.85 and broad-spectrum prescribing was 0.94. Effects were still observed at 12 months. Feedback effectiveness and messages focusing on harms did not become any better with case-mix adjustment (Schwartz et al., 2024). These results suggest that a similarly low-intensity mailed intervention can be employed to change behaviour when the feedback is tailored and involves peer comparison and targets specific measurable prescribing behaviours.

Carney et al. (2025) offer complimentary information on the selection of antibiotics as opposed to the amount of antibiotics prescribed. A pragmatic study in the form of a cluster-randomised approach was applied to 5,073 family physicians in British Columbia who were put into audit-and-feedback and educational summary, educational summary only and control groups. Personalised A&F information and educational material increased the probability of taking nitrofurantoin as the first-line cure as advised in 21,307 cases of lower urinary tract infection without complications. The odds ratio was reported as 1.28 (95% CI 1.07-1.52) which is about 8.7 cases per 100 family physicians first-line nitrofurantoin-treated over a period of six months. To maintain the effect, repetition seemed to be used. This is clinically important in that it proves that A&F can not only enhance the presence of an antibiotic being prescribed but also the drug selection quality.

de Vries et al. (2022) show how peer A&F can be effective in a less resource-intensive primary-care setting. It was also used in 10 Cape Town community healthcare centres and it included multidisciplinary learning and collaboratives and facility level audit discussions. Of 2,077 prescriptions, the percentage of compliance with a set of antibiotic-quality measures increased 19 percent at baseline to 47 percent post-intervention and antibiotic use decreased 12.9 defined daily doses between the pre- and post-intervention period, a difference of 19.3 percent. Lack of randomised concurrent control implies that causality is not as strong as in Schwartz et al. or Carney et al.; however, the size, consistency across quality indicators and integration into existing clinical meetings confirm the practicality of peer-led A&F in resource-constrained settings (de Vries et al., 2022).

Megatrial antibiotic trials were not necessarily successful. Aghlmandi et al. (2023) chose randomly 3,426 Swiss primary-care physicians with a median-to-high pattern of antibiotic prescribing to quarterly automated individualised feedback including peer benchmarking or control. During the second year of intervention, the median annual prescribing rates in the two groups were 8.2 and 8.4 in the intervention group and the control group, respectively. The difference in the rate of antibiotic prescribing between the intervention and the overall increase in prescribing over time was only -1.1 percent with a 95 percent confidence interval between -1.2 percent and 1.0. The majority of the antibiotic categories did not produce any clinically significant reductions, but quinolone prescribing had a slight reduction. The study is, therefore, a good sign that periodic personalised performance reports in isolation do not result in behavioural change (Aghlmandi et al., 2023).

Singer et al. (2022) assessed A&F in the context of Choosing Wisely Canada. The number of practices randomised was 46 that were controlled to receive no recommendation, recommendation summaries and recommendation summaries with personalised feedback. In the recommendation-only condition, 91.6% of clinicians compared to 95.9% in the recommendation-plus-feedback condition decreased the relevant antibiotic or antipsychotic prescribing outcome between baseline and immediate post-intervention period, and the effect was still noticeable in the longer follow-up. What is more important, however, the recommendation-only group also achieved a lot. The research thus endorses the efficacy of an implementation package, yet offers less conclusive data to the incremental impact of personalised A&F as compared to recommendations only (Singer et al., 2022).

The evidence of antibiotics can therefore be interpreted in the context of heterogeneity and not of positive studies selection. Xu et al. (2025) merged 56 randomised primary-care trials. A relative decrease of 11% in A&F was found in 21 studies that reported total volume of antibiotics (RR 0.89, 95% CI 0.84-0.95) with a very high heterogeneity of $I^2=97\%$. Across 16 studies, unnecessary initiation decreased by 23% (RR 0.77, 95% CI 0.68-0.87; $I^2=72\%$). Prolonged treatment decreased by 13% across four studies (RR 0.87, 95% CI 0.81-0.94; $I^2=86\%$), while broad-spectrum selection decreased by 17% across 17 studies (RR 0.83, 95% CI 0.75-0.93; $I^2=96\%$). These meta-estimates do not only ensure effectiveness on a population level, but also demonstrate that contextual and design variables are in control of the results of a single programmes (Xu et al., 2025).

4.3 Opioid and Other Medication Prescribing

The opioid literature demonstrates that the effectiveness of interventions can be dependent on the baseline prescribing and target choice. In British Columbia, McCracken et al. (2023) randomised family physicians to early or delayed feedback of individualised opioid-prescribing feedback. The number of physicians in the early and late groups was 2,260 and 2,156 respectively. Although there was a slight decrease in opioid initiation in both groups (around 2.10 and 1.94/month in early-feedback and delayed groups respectively), there was no significant difference between the two groups. The percent difference between groups was 2.1 with the 95% CI of -4.4 to 0.3 and odds ratio of initiating any opioid was 0.98 (0.96-1.01). No significant differences were observed in the prescription quantity, morphine milligram equivalents or duration. The trial failed to provide evidence, therefore, of any change in the initiation of opioid therapy in opioid-naïve patients due to the intervention (McCracken et al., 2023).

Lacroix et al. (2025) compared voluntary A&F in the presence or absence of academic detailing among Ontario family physicians using matched administrative data. These researchers have utilized 1,469 physicians who received A&F, 245 physicians who received A&F and academic detailing and 4,211 matched controls. High risk opioid prescribing pre-intervention declines were found in all groups and no significant overall post-intervention difference was found in either intervention. However, higher-volume prescribers A&F were associated with an incremental reduction in prescribing slope of -0.37% (95% CI -0.65 to -0.09; $p < 0.01$), but not with academic detailing. This is a good lesson to implement: A&F can be implemented with a higher marginal effect when applied to clinicians with a high area of improvement rather than applied to everyone (Lacroix et al., 2025).

The benzodiazepines use is better supported. Vicens et al. (2022) randomised 81 Spanish primary health centres (749 general practitioners) to multicomponent programme or usual care. The intervention was a combination of a two-hour educational workshop, monthly A&F on benzodiazepine prescribing and an additional supporting webpage. At 12 months, the intervention group had a much lower level of total benzodiazepine use with the adjusted mean difference of -3.24 being defined daily doses per 1,000 inhabitants per day (95% CI -4.96 -1.53; $p = 0.001$). The use of benzodiazepines in the long term also decreased. Nevertheless, since education, feedback and web support were provided in combination, the independent contribution of A&F can not be separated. The intervention and control practices also have baseline imbalances, which are also a reason to be cautious in interpretation but the data on the randomised design and objective prescription point to a real impact of intervention (Vicens et al., 2022).

In one of the studies by Andrade et al. (2022), the authors compared digital delivery and postal delivery of an A&F intervention to minimize the possibility of inappropriate use of gabapentinoids among 2,552 Australian general practitioners. The decrease in the dispensing of gabapentinoids in the two groups was 11.2%. Basically, there was no difference in the outcome of medicine-use between digital and postal delivery and digital delivery with increased psychologist initiations, 3.8% and 2.0% respectively. This decrease cannot be safely explained by A&F alone as there was no no-A&F control. Nevertheless, the argument of the digital and postal modalities equivalence suggests that a viable alternative could be provided by electronic delivery with an apparent absence of an apparent efficiency loss (Andrade et al., 2022).

This article by Carter et al. (2023), a pharmacist-centered review, adds to the evidence that the individual administering or providing feedback can be impactful. Eleven cluster-randomised trials in nine countries were included and six of the sufficiently similar studies underwent meta-analysis to provide a pooled risk ratio of 0.78 (95% CI 0.64 0.94) of prescribing outcomes. Verbal and written feedback, computerised decision support, training of pharmacists, continuing support or patient review were more likely to be included as more effective interventions. These results are aligned with the interpretation that feedback is more intense when the recipients are assisted to make tangible clinical decisions based on performance data instead of being asked to make an interpretation of a report on their own (Carter et al., 2023).

4.4 Diagnostic Imaging and Testing

The evidence presented by O'Connor et al. (2022) is especially good as far as diagnostic imaging is concerned. Cluster-randomised trial involved 3,819 high-requesting Australian general practitioners in 2,271 practices. In the primary study that comprised 3,660 practitioners, the overall percentage of the 11 specific musculoskeletal imaging tests done at 12 months after A&F was 27.7 per 1000 consultation after A&F as compared to 30.4 in control. The adjusted mean difference was -2.66 requests per 1,000 consultations (95% CI -3.24 to -2.07; $p < 0.001$). The researchers have approximated that around 47,318 fewer imaging tests were ordered due to the intervention during the period of the study. The time-lagged feedback also provided a small incremental reduction at 18 months compared to feedback at a single time (O'Connor et al., 2022).

There are several attributes that can be credited to this success. Participants were targeted high-requesting practitioners, in particular, which gave the intervention a considerable room of improvement. The target behaviours were measurable objectively and to a great extent they were under the direct control of the clinicians. It was founded on individualised feedback and involved peer comparison. Also, the outcome was measured using the administrative data but not the self-report. These characteristics are similar to attributes of increased A&F effectiveness in the broader evidence base (Ivers et al., 2025).

Hocking et al. (2022) present quite an odd but informative withdrawal design. Instead of merely introducing A&F, the ACCEPt-able cluster trial analysed implications of withdrawing financial incentives and/or audit and feedback to practices that have been previously involved in a chlamydia-testing programme. The independent effect of elimination of financial incentives was small, whereas elimination of A&F resulted in lowering the annual testing of chlamydia. Practices in which A&F were dropped were nearly 30 percent less likely to test compared to practices in which feedback was maintained at year two with an adjusted odds ratio of approximately 0.7. The finding confirms the likelihood that a behaviour can be maintained with the help of an improvement programme set up by a continuous feedback. However, the overall testing declined considerably in all groups and the result should be considered as an evidence regarding maintenance and not the fact that A&F prevents contextual deterioration completely (Hocking et al., 2022).

4.5 Clinical Documentation and Antimicrobial Stewardship Processes

Vanstone et al. (2022) evaluated monthly A&F and the intention was to impact primary-care prescribers to record clinical indications to prescribe antimicrobials. The average prescription rate with a documented indication was 21.4% before the implementation. The average went up to 37.8% after A&F started. Mixed-effects logistic regression showed that odds of an indication being recorded increased about three times following intervention in comparison with the pre-intervention phase (OR 3.0, 95% CI 2.73.4). The effect is large upon a documentation outcome, yet, as the outcome is not controlled at the same time, then other time-related factors cannot be excluded. However, the intervention focused on a specific and actionable behaviour, and provided repeated information on individual performance, which could be the reason behind the size of change (Vanstone et al., 2022).

Roche et al. (2022) confirm the significance of data quality. Data validity and interpretation were found to be a significant variable in general practitioners who were provided with national postal antibiotic-prescribing feedback as to whether they would utilize the feedback or not. When the recipient does not believe that the metric is suitable to reflect their clinical practice, then a report cannot effectively develop discrepancy. As such, technical quality of the audit itself is an intervention and not just a pre-data-management activity.

4.6 Social Determinants of Health

Gillespie et al. (2023) constructed the A&F concept by prescribing and experimenting with unannounced standardised patients to determine how primary-care physicians diagnosed and responded to the social needs. The number of standardised-patient visits was 108 pre-intervention and 183 intervention. Recognition of housing needs increased in groups receiving A&F by 46 to 60 percent ($p=0.045$) and active response to housing needs increased by 15 to 41 percent ($p=0.004$). More frequently identified (39/53) ($p=0.041$) was social isolation. Similar improvements were not observed in the comparison group. Though the sample sizes were less than large prescribing trials, the research is critical since it proves that feedback can impact communication and assessment behaviour and not just transactional prescribing choices (Gillespie et al., 2023).

Simultaneously, social isolation has not been firmly explored, which is why the understanding of a performance gap could be simpler to modify as opposed to more complicated interpersonal and organisational behaviour. This difference leads to one of the main theoretical hypotheses, namely that A&F is going to be stronger when the recipients know what action to take after receiving the feedback and have the opportunity and possibility to take it.

4.7 Patient Safety and Organisational Feedback

The example provided by Fiol-deRoque et al. (2025) is of good quality as it is theoretically plausible and the null result is caused by the action being constrained by the implementation conditions. It was a SinergiAPS cluster-randomised trial, which involved 59 primary healthcare centres and 1,053 healthcare professionals in Spain. Educational material and a structured improvement-planning template containing patient-reported safety information were availed to centres and audited and reported to the centres. At 12 months the mean patient-safety culture score was approximately 3.60 in intervention centres and 3.64 in control centres and there was no significant overall intervention effect. There were also no significant differences in secondary patient-reported safety outcomes and preventable hospital admission outcomes.

There was low implementation fidelity. Out of 30 intervention centres, just ten of them conducted action-plan team meetings and 8 centres reported 57 improvement actions. Immediately after implementation, the COVID-19 pandemic started, which significantly decreased the organisational capacity to work on improvement. The null result does not in any sense mean that patient-centred A&F is necessarily ineffective. Rather, it shows that feedback must have adequate organisational strength to create and implement improvement actions. Feedback report that is technologically advanced will not improve results when the feedback-to-action pathway is interrupted (Fiol-deRoque et al., 2025).

4.8 Evidence from Resource-Limited Primary Healthcare Systems

Traditionally, evidence base has been geared towards high-income health systems. Ivers et al. (2025) state that the wide A&F evidence base only included approximately 9 percent of the trials in low- and middle-income countries and limited the confidence in the transportability. Recent research in South Africa, Mozambique and Pakistan is thus of specific significance.

Dinis et al. (2025) explored the Integrated District Evidence-to-Action strategy in 154 primary healthcare facilities, in 12 Mozambican districts. The intervention included A&F meetings, readiness assessment and facility support. Controlled interrupted time-series and propensity-score methods were used to evaluate ten service-delivery and five readiness outcomes. Results were mixed. The initial appointments of at-risk kids were improved with an incidence rate ratio of 1.06 (95% CI 1.04-1.07) and the initial polymerase-chain-reaction testing of HIV-exposed children was improved with a ratio of 1.02 (95% CI 1.01-1.03). Odds ratio of infrastructure preparedness was significantly higher with odds ratio of 5.84 (95% CI 1.3225.88). Conversely, the initiation of new contraceptive use and long-acting contraceptive use had poor relationships, and the preparedness to essential care improved, and six out of ten service-delivery outcomes did not have statistically significant effects. The data subsequently challenges a naive belief that A&F will equally increase all indicators simultaneously, especially where there are structural resources limitations to response capacity (Dinis et al., 2025).

In a retrospective assessment of A&F among 61 primary healthcare physicians in Karachi, Omair et al. (2025) assessed the A&F during the 2019-2023. On-site and electronic-medical-record audits followed by structured feedback involved in repeated audits, statistically significant improvements in consultation skills, history taking, provisional diagnosis, protocol adherence and management of common conditions were associated with statistically significant improvements in consultation skills, history taking, provisional diagnosis, protocol adherence and management of common conditions. The reason is important because it suggests that the evidence can be practically applied in the setting of Pakistani primary healthcare, but the lack of parallel control, retrospective design and potential maturation effects suggests that the study has less causal evidence compared to randomised studies (Omair et al., 2025).

Together with de Vries et al. (2022), these studies indicate that A&F is applicable in resource-limited environments and the effectiveness of interventions is mediated by the health system in general. Feedback can reveal a performance gap, and no mechanism of feedback can generate unavailable medicines, staffing, infrastructure or clinical time. Findings during audits in such settings should therefore be linked to managerial and clinician level action.

4.9 Digital Feedback, Dashboards and Delivery Mode

Digitalisation is often thought to enhance A&F since information can be updated quickly, personalised automatically and provided at low marginal cost. The contemporary realities show that the digital platforms address major logistical problems compared to behaviour-change problems that cannot be addressed by the digital platforms alone. Tsang et al. (2022) have reviewed 88 articles and evaluated 65 electronic A&F systems. Eight out of 18 best-practice features that were on the median were on these systems. Only 32 of the systems had near-real time feedback and only 20 included action planning. The authors came up with the conclusion that effective electronic A&F necessitates actionable, role-specific and timely data, which is connected to action-supporting mechanisms. It is relevant since dashboards with big amounts of passive indicators can amplify information without enhancing it (Tsang et al., 2022).

Xie et al. (2022) searched 11 randomised trials of clinical dashboards in money of feedback or decision support. Eight standalone dashboards were reviewed and generated incongruent evidence of antibiotic prescribing and little evidence of benefit of some other medication outcomes. Three multicomponent interventions had more persistently positive outcomes, such as decreases in opioid use of low back pain, improvement in cardiovascular risks screening and decreased antibiotic prescription in upper respiratory infections. The authors thus concluded that the evidence of dashboards was scarce and context-specific (Xie et al., 2022).

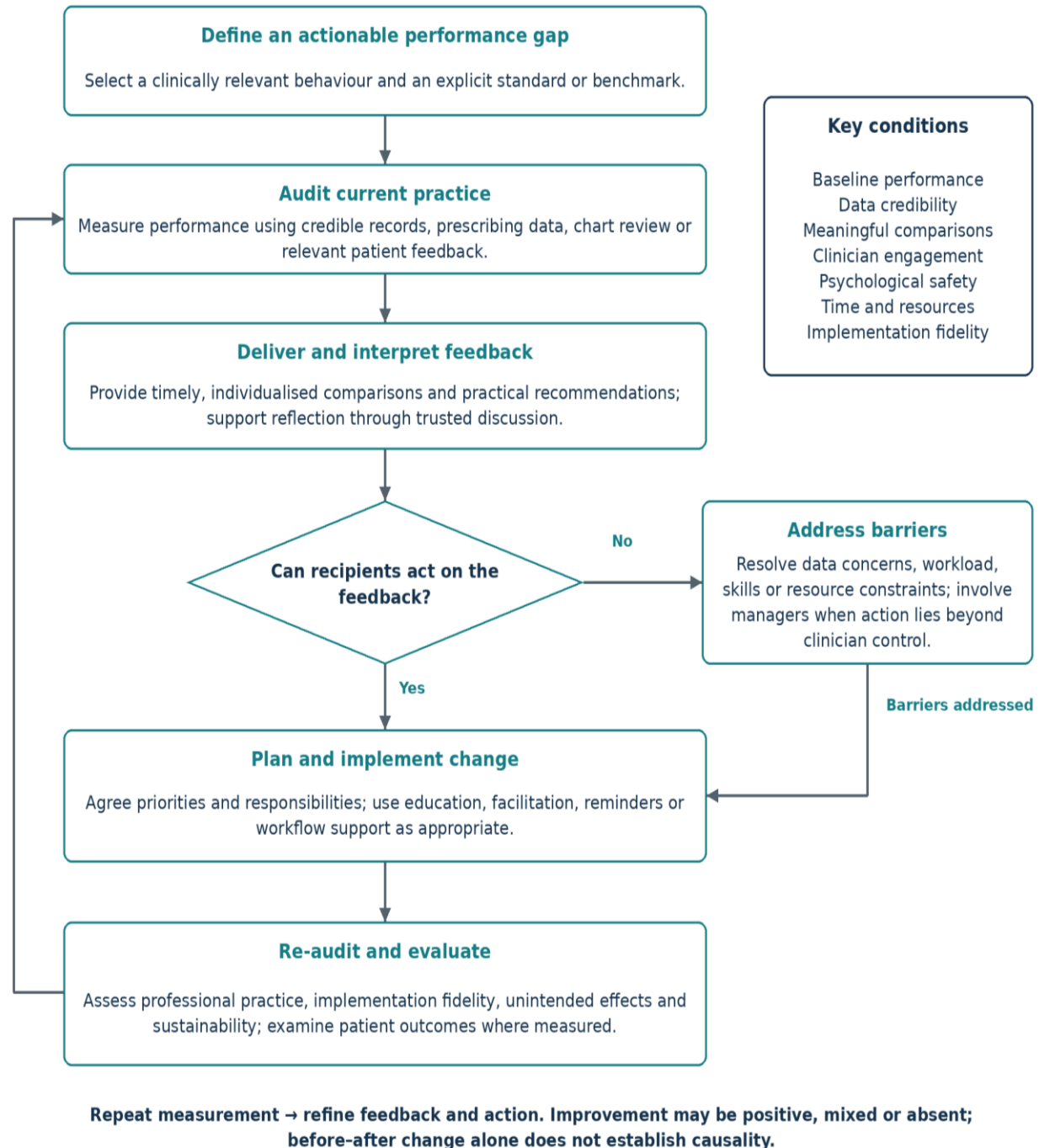
Meier et al. (2025) were interested in interactive dashboards in primary care. They found 9,354 records and six studies were identified. There were five randomised designs. Four interventions involving dashboards as part of more comprehensive interventions demonstrated improvement at least in one key quality indicator, and the two that evaluated more standalone dashboard approaches had an uneven result. Dashboard functionality, settings, diseases and outcomes heterogeneity restricted generalisability. These results are consistent with the primary-study findings: feedback systems seem to be the most efficient when they are a part of an active process of improvement and not stores of performance data (Meier et al., 2025).

Shuldiner et al. (2023) back this up with user-centred development of dashboard. To make sure that shown metrics corresponded to clinical workflow and decision-making, A&F designers used theoretical evidence, personas and interviews with physicians. Nevertheless, Rouleau et al. (2023) showed that even reformed reports may not be sufficient because the physicians lack time, trust, and relevant actions or ability to convert the information into action. The implication is that the interface usability is required but not enough.

Figure 4: Audit Feedback Action Improvement Cycle

Audit-Feedback-Action Improvement Cycle

Conceptual implementation pathway synthesised from the review; arrows represent proposed process links, not proven mediation or guaranteed benefit.



Source: (Meier et al., 2025); (Omair et al., 2025).

4.10 Context, Engagement and Mechanisms of Action

Qualitative and implementation evidence can be used to describe why quantitatively similar interventions can have varied results. Arseneault-Lapierre et al. (2022) provided a single round of A&F cycle in eight Canadian primary-care (interdisciplinary)

dementia-care centres. Sixty-three participants discussed the outcomes of the audit and emphasised organisational problems, such as human-resource constraints, relationships with community services, role definition and external support of specialists, and clinician problems, such as confidence and attitudes. Solutions to the problems that were suggested by the participants on a local level were pathways, documentation tools, training, benchmarking and financial incentives. The research is ineffective in the causal sense, but it proves the way feedback can act as an organised diagnostic process in the improvement of the organisation.

Roche et al. (2022) discovered that Irish general practitioners desired an antibiotic feedback to be believable, digestible and provably meaningful. Rouleau et al. (2023) also discovered that two general determinants of report usability existed, such as the alignment with clinician expectations and the possibility to engage in quality improvement. The less probable to do so were the physicians in whom the indicators were believed to be inaccurate, in whom the improvement was beyond their power or in whom the workload and the existing systems prevented action.

van der Winden et al. (2023) studied the peer discussion of A&F in Dutch general practice. The deeper reflection was also identified by participants, placing numbers and professional ownership as the possible advantages of peer discussion, yet they were also aware of the risks related to psychological safety, unwilling participants and inappropriate transparency. The importance of these results is that feedback may either be a type of developmental information or perceived surveillance. Defensive reactions may destroy learning in case recipients perceive feedback as punishment or imposed by an external authority.

Schwartz et al. (2023) create an evidence-based best-practice advice on the use of antibiotic A&F in primary care via modified Delphi consensus. They concentrate on their work on reliable data, meaningful comparators, clinical relevant measures and designs to help clinicians decide on what to do next. Shuldiner et al. (2024) also validated the process evaluation that the clinician interest in antibiotic A&F reports is determined by how the reports are perceived, understood, and put into practice. McClatchey et al. (2024) demonstrated the same study the explicit design of A&F according to the behaviour-change theory rather than the selection of report elements according to aesthetic or technical requirements can be beneficial.

Table 1. Summary of Findings from Primary Effectiveness Studies:

Study	Sample / Design	A&F Intervention	Main Finding
Schwartz et al. (2024)	5,046 physicians; pragmatic factorial RCT	Mailed peer-comparison antibiotic feedback	Prescribing RR 0.95; unnecessary prescribing RR 0.89
Carney et al. (2025)	5,073 physicians; cluster-randomised trial	Peer comparison plus therapeutic recommendations	Nitrofurantoin prescribing improved: OR 1.28 (95% CI 1.07–1.52)
de Vries et al. (2022)	2,077 prescriptions; ten centres	Peer antimicrobial audit and feedback	Quality adherence 19%→47%; antibiotic use decreased 19.3%
Aghlmandi et al. (2023)	3,426 physicians; randomised trial	Quarterly personalised prescribing feedback	No meaningful reduction: –0.1% (95% CI –1.2–1.0)
Singer et al. (2022)	46 practices; cluster-randomised trial	Recommendations plus personalised feedback	95.9% improved; incremental feedback benefit remained uncertain
McCracken et al. (2023)	4,416 physicians; randomized-delay trial	Personalised opioid-prescribing feedback	No significant effect: OR 0.98 (95% CI 0.96–1.01)
Lacroix et al. (2025)	5,925 physicians; matched evaluation	Opioid A&F ± academic detailing	High-volume prescribers improved: slope –0.37%; overall effect nonsignificant
Vicens et al. (2022)	749 GPs; 81 centres; cluster RCT	Education, monthly A&F, web support	Benzodiazepine use decreased –3.24 DDD/1,000/day
Andrade et al. (2022)	2,552 GPs; cluster-randomised comparison	Digital versus postal feedback	Both groups reduced gabapentinoid dispensing by 11.2%
O'Connor et al. (2022)	3,819 GPs; cluster-randomised trial	Individualised imaging feedback	Imaging decreased –2.66 requests/1,000 consultations

Study	Sample / Design	A&F Intervention	Main Finding
Hocking et al. (2022)	General practices; cluster-randomised trial	Withdrawal of ongoing A&F	Removing feedback reduced chlamydia testing; adjusted OR ≈ 0.70
Vanstone et al. (2022)	Primary-care quality-improvement study	Monthly antimicrobial documentation feedback	Documented indications increased 21.4%→37.8%; OR 3.0
Gillespie et al. (2023)	291 standardised-patient visits; quasi-experimental	Social-needs audit and feedback	Housing identification 46%→60%; response 15%→41%
Fiol-deRoque et al. (2025)	59 centres; 1,053 professionals; cluster RCT	Patient-safety feedback and action planning	Safety culture 3.60 versus 3.64; nonsignificant effect
Dinis et al. (2025)	154 facilities; interrupted time-series	Facility A&F and readiness support	At-risk visits IRR 1.06; infrastructure readiness OR 5.84

5. Critical Discussion

5.1 Principal Finding

The main finding of this systematic review is that the clinical audit and feedback is effective in the average but highly varied in particular implementation in primary-care. It has to be then conceptualised as a conditional and not a deterministic intervention. When A&F is shown to work at an aggregate level, the more technically true conclusion is that in the event that the behavioural target, the data, the comparator, the delivery system and the implementation environment are conducive to action, A&F can lead to significant changes in professional practice.

This goes in line with Ivers et al. (2025). They conducted a systematic review that was broad with 292 randomised studies. Out of 177 trials that provided 558 dichotomous professional-practice outcomes, the median absolute improvement was 2.7% with an IQR of 0.0 to 8.6. They had a model-based absolute improvement of 6.2% (95% CI 4.18-8.2) and odds ratio of 1.47 (95% CI 1.31-1.64) which is of moderate-certainty evidence. However, the variance is so great that the average cannot be reported in a manner that would imply that all the programmes will be enhanced by 6.2 percentage points. The impact of other modern interventions in primary care was much greater than this, but others had almost no incremental impact (Ivers et al., 2025).

5.2 Why Some Interventions Produce Larger Effects

Room for improvement is one of the most uniform effect modifiers. O'Connor et al. (2022) have specifically chosen high-requesting GPs and achieved a significant imaging reduction. The overall additional opioid-prescribing effect was found to be insignificant as per the study by Lacroix et al. (2025), but the high-volume prescriber subgroup significantly improved. The delivery of a feedback report to a clinician with a high performance already has a limited benefit that can be achieved because of ceiling effect. Developed to address clinicians or practices whose performance can be quantified with gaps, thereby, can increase efficiency and effect size.

The other significant mechanism is feedback specificity. Schwartz et al. (2024) did not simply tell clinicians that their use of antibiotics might be better; the reports given the performance of prescribing as compared to others. Carney et al. (2025) integrated personalised prescribing information with concrete therapeutic recommendations as far as the initial treatment of UTI is concerned. In comparison, passive reports where the clinicians need to recognize the appropriate issue and come up with a response on their own place a heavier cognitive and organisational load.

Discrepancy is also accepted on the basis of credibility. An administrative data can be scaled to create feedback and clinicians might find that case mix, diagnostic coding or unusual patients create unfair comparisons. Interestingly, Schwartz et al. (2024) concluded no extra advantage of case-mix-adjusted feedback, implying that a more and more complicated statistical adjustment is not necessarily associated with the enhancement of behaviour performance. The question of interest here can be whether the comparison is perceived by the recipients as fair and actionable, as opposed to being statistically maximally sophisticated. The origin of feedback and the correlation of the source of feedback can also play a role. Carter et al. (2023) found positive effects of pharmacist-involved A&F with certain benefits in situations when pharmacists were provided with the appropriate training and provided with the constant support. Ownership and contextual interpretation can be enhanced with peer discussion, which is proposed by van der Winden et al. (2023). Similar results were reported by Ivers et al. (2025) who observed greater effect sizes related to local champions, interactive methods and top-peer or benchmark comparators.

5.3 Baseline Performance and Regression to the Mean

Regression to the mean also has to be faced by studies which particularly recruit poor performers. The natural extreme baseline performance can be improved even without treatment in a second measurement. Randomised controlled groups are thus fundamental in the estimation of causal effects among the selected high performers or high prescribers. O'Connor et al. (2022)

mitigated it by randomisation, which enabled it to compare itself to the current changes in control physicians. This is particularly an issue with uncontrolled before-after studies. Vanstone et al. (2022), de Vries et al. (2022) and Omair et al. (2025) noted large post-feedback gains, but in the absence of untreated groups, some percentage of change can be due to secular trends, policy influences, or regression to the mean. These studies provide good implementation evidence of a real-world implementation but cannot be given equal consideration to large randomised trials.

Lacroix et al. (2025) give the example of this principle. There was already a decrease in high-risk opioid prescribing prior to intervention. Because the investigators modeled the pre-intervention slopes specifically, they could determine that most of the decline was not necessarily as a result of feedback. Less rigorous before-after evaluation could have reached a falsely positive conclusion.

5.4 The Importance of Co-Interventions

Among the largest interpretive problems is that A&F is not typically administered as a unit. Education, reminders, peer comparison, recommendations, academic detailing, web resources, facilitated discussions and action plans are common accompaniments of feedback. The broader evidence base reveals education and reminders, which were found by Ivers et al. (2025). This makes it more practical but less able to isolate the particular causal contribution of feedback.

The example of Vicens et al. (2022) incorporated educational training, monthly prescribing feedback and online support. The programme reduced the prescribing of benzodiazepines and it is not known to what extent the effect was caused by feedback alone. Singer et al. (2022) is especially informative as the recommendation-only group also showed significant improvement, which indicates that educational content had something to do with the overall behavioural impact. A combination of A&F and an education summary was also used by Carney et al. (2025). These multicomponent strategies cannot be criticised on the basis that individual components cannot be isolated but, in many cases,, combinations in healthcare implementation are needed. However, systematic reviews have to differentiate between the question: What is the effectiveness of an A&F-centred programme and the question: Isolated causal effect of feedback.

5.5 Frequency of Feedback

Frequent feedback is usually suggested as it is possible that an intervention has a relapse when it is stopped. O'Connor et al. (2022) found another small effect of repeated feedback at 18 months and Carney et al. (2025) found that repetition seemed to maintain improvement. Hocking et al. (2022) concluded about the importance of the A&F continuation as withdrawal was associated with the further worsening of chlamydia tests. Nevertheless, it is not just about frequency. Repeated presentation of the same non-actionable information can result in the creation of alert fatigue or report fatigue. The increased evidence base which Ivers et al. (2025) consider could not be used to make a linear assumption that more feedbacks would inevitably lead to more effects. The right frequency will thus tend to be based on how fast the underlying behaviour changes, the frequency of new data becoming available and whether each of the feedback cycles justifies a new corrective action.

5.6 Digital Dashboards: Scalability Versus Actionability

Digital dashboards provide enticing scaling. The process of extracting electronic records can be automated to make the audit less expensive, enable regular updates and give feedback to thousands of clinicians. Andrade et al. (2022) showed that postal delivery is potentially as effective as electronic delivery in terms of medication-use outcomes at national levels. However, Meier et al. (2025), Tsang et al. (2022) and Xie et al. (2022) all caution against the assumption that the dashboard availability is concomitant with the quality improvement.

The difference between behavioural action and data visibility is crucial. A dashboard can be a true indicator that a clinician is overprescribing antibiotics but it needs the clinician to pay attention to the number, have faith in the calculation, know why performance is not aligned with the desired standard and then put in place priority, a different behaviour and apply it regularly. Dashboards must thus not add to but lessen the cognitive efforts involved in transforming measurement to action. This adheres to the following design principles listed in the evidence provided: prioritize a few meaningful indicators; prioritize areas of concern over showing all possible data; present clinically interpretable comparators; show trends; make action-oriented lists of patients visible where needed; integrate feedback with normal workflow; and provide specific recommendations (Tsang et al., 2022; Shuldiner et al., 2023; Rouleau et al., 2023).

5.7 A&F in Low-Resource Settings

The Cape Town, Mozambique, and Karachi studies enhance external validity but also demonstrate why the health-system capacity needs to be taken into consideration. de Vries et al. (2022) obtained significant improvement with multidisciplinary peer feedback incorporated into regular clinical meetings. This is promising as the intervention was based heavily on existing systems, and not costly proprietary infrastructure. But a multicomponent intervention by Dinis et al. (2025) with many components had mixed results. This makes sense. Where the failure to perform is as a result of the decision making of clinicians, then there is a possibility that the behaviour may be altered by the feedback. In situations where there are shortages in terms of inadequate personnel, drugs, transport, equipment or infrastructure, clinician feedback cannot be leveraged adequately. The audit can be able to come up with the problem right, but the locus of corrective action is at managerial or system level.

The policy implications on the primary-care are important. A&F ought to make a difference between structural-performance gaps and professional-performance gaps. Offering feedback to clinicians towards outcomes they are not in control of jeopardizes trust and engagement. Facility-level audits, on the other hand, can also be useful in the case when the managers access the information and can take the power and resources to overcome structural constraints.

5.8 Null Effects Are Scientifically Important

Another term that cannot be used to describe a well-done systematic review is the null studies as failure to provide support to a desirable conclusion. The null effects will give important evidence on the boundary conditions. Aghlmandi et al. (2023) illustrate that the failure of repeated automated benchmarking can happen even with a large methodologically robust trial. McCracken et al. (2023) show that personalised feedback might not lead to a decrease in already relatively rare opioid initiation. Lacroix et al. (2025) show that secular prescribing trends may take over marginal effect of intervention. As shown by Fiol-deRoque et al. (2025), a derailment in implementation can prevent action in a scenario where the clinicians believe an intervention to be acceptable. Dinis et al. (2025) demonstrate that one programme can have a positive impact on certain indicators, remain unchanged and exacerbate others.

Such studies help to reinforce and not undermine the overall evidence base as they do away with inappropriate generalisation. A&F is not to be applied in case there are performance data. It must be used when there is a significant variance, the recipient is capable of making a difference, data must be credible, the intended behaviour is well-specified and organisational circumstances allow responding.

5.9 Statistical Versus Clinical Significance

A&F often has numerically small impacts. Scales are important in determining whether they are clinically important. The 5 percent relative reduction in total antibiotic prescribing that Schwartz et al. (2024) found may not be a huge difference on a per-physician basis, but when multiplied by thousands of physicians and multiple consultations, could lead to a large number of prevented prescriptions. O'Connor et al. (2022) found a decrease of 2.66 target imaging requests per 1,000 consultations but determined tens of thousands of fewer investigations in the study population.

As a result, the evaluation must consider absolute effect, population exposed, baseline event rate, cost of intervention, sustainability and clinical impacts of the targeted behaviour. Even in cases where the effect of a person is small, low-cost automated A&F can be cost-effective. On the other hand, the facilitated interventions that have large effects need to be resource intensive to warrant implementation.

5.10 Sustainability

Long term sustainability evidence is not as conclusive as the evidence of short term behaviour change. Schwartz et al. (2024) demonstrated persistence at 12 months and Singer et al. (2022) demonstrated maintenance at a longer period. According to Hocking et al. (2022), behaviour can decrease in the event of withdrawal of feedback.

Sustainability may be informed by the necessity to proceed with recurring external feedbacks or the intervention has permanent impacts on habits, workflows and local norms. The future research should therefore provide the demarcation between interim response to a report and institutionalisation of better practice.

6. Overall Certainty and Quality of Evidence

The overall confidence in A&F to enhance professional practice is moderate to high at the general conceptual level but less at prediction of the effect magnitude in any one particular primary-care setting. This difference resolves the seemingly paradoxical results. Randomised evidence is vast, and contains a number of very large pragmatic primary-care trials. The subjective measurement bias is normally limited by the routinely collected prescribing or test-ordering data, which determines the results. According to recent systematic reviews, there are statistically significant average benefits. These aspects enhance belief in the overall effectiveness proposal.

Certainty regarding precise magnitude is reduced by heterogeneity. Interventions differ with regards to baseline performance, type of feedback, comparator, co-interventions, intensity and target behaviour. Ivers et al. (2025) observed a tremendous variation even in a very large randomised evidence base. The I2 was stated to be 72-97% in the results of antibiotics (Xu et al., 2025). This kind of heterogeneity suggests that a pooled effect can be understood as an average of various settings as opposed to a forecast of a new programme. The fact that some design characteristics are present is also not as decisive because there are numerous characteristics that are not identified with factorial head-to-head tests but with meta-regression or comparisons across studies. As an example, interventions that are effective might involve peer comparison, local champions, action plans and multimodal feedback, yet interventions that have those elements may vary in other aspects. Randomisation of specific components where possible should be carried out in future trials.

7. Implications for Primary Healthcare Practice

This evidence supports the inclusion of A&F in regular primary-care quality-improvement systems, but should be applied by first selecting an appropriate target and not by first selecting a reporting technology. The targeted behaviour should represent an important quality gap, be measurable accurately and be meaningfully modifiable by the recipient. Feedback should be

individualised whereby personal decisions yield the outcome. Selecting peer comparators carefully must be made. Comparison with high-performing peers or evidence-based benchmarks could establish a more specific target of improvement as compared to comparison with an average, which can itself be suboptimal care (Ivers et al., 2025; Schwartz et al., 2023).

Actionability should be emphasized in reports. As a reaction to the feedback, a clinician should be able to answer the following three questions in a short period of time: what do I have that is not in line with the desired standard, how prominent and significant is the difference, and how do I do it differently. An implementation perspective is incomplete with a report that only answers the first question. Clinical workflow should also be incorporated with feedback. The results of Rouleau et al. (2023) showed that the improvements in visual presentation can be overridden by the lack of time and workflow fit. A&F ought to be provided accordingly at a time, frequency and place where the recipient can reasonably read and act on it.

Passive feedback may not be desirable in complex problems as compared to facilitated feedback. Clinicians may need help to interpret the data and translate it into practical plans, and local clinical champions, pharmacist support, or peer discussion or quality-improvement coaching can assist in doing so, particularly when there are several barriers to the performance gap being observed (Carter et al., 2023; van der Winden et al., 2023).

8. Implications for Policy and Healthcare Organisations

At system level, A&F is appealing since electronic clinical and administrative data are more and more making audit affordable without costly chart review. British Columbia and Australian trials show that it is possible to provide feedback to thousands of clinicians, which is done by Big Ontario. This comes in handy particularly when small individual effects are accumulated in a large population. However, A&F cannot be applied in organisations as a major type of punitive accountability. The feedback that will be utilized to make improvements must trust the information and psychological safety to investigate the performance gaps. van der Winden et al. (2023) discovered a conflict between transparency and a safe learning environment. Oversight or retributive utilization of performance data could thus disintegrate the contemplative procedures where enhancement relies.

Audits also demand organisations to resolve structural problems in organisations. The performance indicators should not be blamed on the individual clinicians in case the performance indicators are not able to improve due to insufficient staffing, medicines, equipments, or referral capacity. Dinis et al. (2025) shows the necessity to make the facility-level feedback connected to the organisational decision-making.

9. Conclusion

Clinical audit and feedback is a promising but situation-specific method of improving the practice of healthcare professionals in primary healthcare. Modern evidence shows major gains in the prescribing of antibiotics, medication safety, diagnostic imaging, documentation, preventive testing and chosen clinical behaviours, but with impacts ranging between substantial and null. Even the strongest recent studies suggest that even low intensity interventions can result in substantial population-level change in the case of feedback that is individualised, credible, benchmarked and actionable. Schwartz et al. (2024), Carney et al. (2025) and O'Connor et al. (2022) are particularly good examples of scalable behaviour change in thousands of primary-care physicians. In contrast, Aghlmandi et al. (2023), McCracken et al. (2023), Lacroix et al. (2025) and Fiol-deRoque et al. (2025) demonstrate that feedback may have minimal contribution to behaviour when the baseline behaviour has limited improvement opportunities, when the behavioural target is non-receptive to the intervention, when the results are primarily determined by secular trends, or when implementation capacity is lacking.

The same conclusion is supported by the systematic-review evidence. A large evidence base of A&F reported a 6.2% model-based improvement and a 2.7% median improvement of desired professional practice in a large evidence base of A&F (Ivers et al., 2025). Xu et al. (2025) demonstrated significant improvements in numerous dimensions of primary-care antibiotic prescribing but extremely high heterogeneity indicates that the mean effect cannot be automatically imposed on the particular environments. Carter et al. (2023), Meier et al. (2025), Tsang et al. (2022) and Xie et al. (2022) further prove that facilitated, multicomponent and actionable feedback is likely to be more effective than passive presentation of the performance data.

On this, A&F cannot be conceptualised as the passing of a report. Effective A&F is an organisational and behaviour intervention that fills the gap between proper measurement and recognition of a meaningful performance gap and the following mapping of the gap to a realistic remedial action. Clinically important gaps, individualised data where needed, credible comparators, clear targets, actionable recommendations, workflow integration and adequate organisational resources to respond should be prioritized in programmes. This can be scaled with electronic dashboards and data collected periodically but technology is not going to resolve the irrelevancy of indicators, lack of credibility, inability to control or improve clinicians. The conclusion suggested by the modern evidence is thus most defensible as the clinical audit and feedback must be an essential part of the primary-care quality improvement, but its application has to be carefully planned and not expected to be self-sufficient. Even a small clinician-level effect of communicating a believable performance difference between actual performance and the target performance via the intervention can lead to huge population-level changes. Repeat reports without such conditions may result in information but not in better

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers.

References

- Aghlmandi, S., Halbeisen, F. S., Saccolotto, R., Godet, P., Signorelli, A., Sigrist, S., Glinz, D., Moffa, G., Zeller, A., Widmer, A. F., Kronenberg, A., Bielicki, J., & Bucher, H. C. (2023). Effect of antibiotic prescription audit and feedback on antibiotic prescribing in primary care: A randomized clinical trial. *JAMA Internal Medicine*, 183(3), 213–220. <https://doi.org/10.1001/jamainternmed.2022.6529>
- Andrade, A. Q., Calabretto, J.-P., Pratt, N. L., Kalisch-Ellett, L. M., Kassie, G. M., LeBlanc, V. T., Ramsay, E., & Roughead, E. E. (2022). Implementation and evaluation of a digitally enabled precision public health intervention to reduce inappropriate gabapentinoid prescription: Cluster randomized controlled trial. *Journal of Medical Internet Research*, 24(1), Article e33873. <https://doi.org/10.2196/33873>
- Arsenault-Lapierre, G., Le Berre, M., Rojas-Rozo, L., McAiney, C., Ingram, J., Lee, L., & Vedel, I. (2022). Improving dementia care: Insights from audit and feedback in interdisciplinary primary care sites. *BMC Health Services Research*, 22, Article 353. <https://doi.org/10.1186/s12913-022-07672-5>
- Carney, G., Maclure, M., Patrick, D. M., Otte, J., Ambasta, A., Thompson, W., & Dormuth, C. (2025). Pragmatic randomised trial assessing the impact of peer comparison and therapeutic recommendations, including repetition, on antibiotic prescribing patterns of family physicians across British Columbia for uncomplicated lower urinary tract infections. *BMJ Quality & Safety*, 34(5), 295–304. <https://doi.org/10.1136/bmjqs-2024-017296>
- Carter, M., Abutheraa, N., Ivers, N., Grimshaw, J., Chapman, S., Rogers, P., Simeoni, M., Antony, J., & Watson, M. C. (2023). Audit and feedback interventions involving pharmacists to influence prescribing behaviour in general practice: A systematic review and meta-analysis. *Family Practice*, 40(5–6), 615–628. <https://doi.org/10.1093/fampra/cmz150>
- de Vries, E., Johnson, Y., Willems, B., Bedeker, W., Ras, T., Coetzee, R., Tembo, Y., & Brink, A. (2022). Improving primary care antimicrobial stewardship by implementing a peer audit and feedback intervention in Cape Town community healthcare centres. *South African Medical Journal*, 112(10), 812–818. <https://doi.org/10.7196/SAMJ.2022.v112i10.16397>
- Disis, A., Augusto, O., Fernandes, Q., Birru, E., Etzioni, R., Gimbel, S., Gloyd, S., Ramiro, I., Gremu, A., John-Stewart, G., Wagenaar, B. H., Weiner, B. J., Chicumbe, S., & Sherr, K. (2025). Can audit and feedback improve health service readiness and delivery outcomes in a low-resource setting? Effectiveness results of the IDEAs strategy from central Mozambique. *PLOS Global Public Health*, 5(5), Article e0004216. <https://doi.org/10.1371/journal.pgph.0004216>
- Fiol-deRoque, M. A., Valderas, J. M., Serrano-Ripoll, M. J., Gens-Barbarà, M., Martín-Luján, F., Sánchez-Freire, E., Montaña, J. J., Mira-Martínez, S., Pastor-Moreno, G., Zamanillo-Campos, R., Riera-Serra, P., & Ricci-Cabello, I. (2025). Evaluating SinergiAPS, an intervention based on patient feedback to improve patient safety in primary care: A cluster randomized trial. *BMC Medicine*, 23, Article 202. <https://doi.org/10.1186/s12916-025-04029-7>
- Gillespie, C., Wilhite, J. A., Hanley, K., Hardowar, K., Altshuler, L., Fisher, H., Porter, B., Wallach, A., & Zabar, S. (2023). Addressing social determinants of health in primary care: A quasi-experimental study using unannounced standardised patients to evaluate the impact of audit/feedback on physicians' rates of identifying and responding to social needs. *BMJ Quality & Safety*, 32(11), 632–643. <https://doi.org/10.1136/bmjqs-2021-013904>
- Hocking, J. S., Wood, A., Temple-Smith, M., Braat, S., Law, M., Bulfone, L., Jones, C., van Driel, M., Fairley, C. K., Donovan, B., Guy, R., Low, N., Kaldor, J., & Gunn, J. (2022). The impact of removing financial incentives and/or audit and feedback on chlamydia testing in general practice: A cluster randomised controlled trial (ACCEPt-able). *PLOS Medicine*, 19(1), Article e1003858. <https://doi.org/10.1371/journal.pmed.1003858>
- Ivers, N., Yogasingam, S., Lacroix, M., Brown, K. A., Antony, J., Soobiah, C., Simeoni, M., Willis, T. A., Crawshaw, J., Antonopoulou, V., Meyer, C., Solbak, N. M., Murray, B. J., Butler, E.-A., Lepage, S., Giltenane, M., Carter, M. D., Fontaine, G., Sykes, M., . . . Grimshaw, J. M. (2025). Audit and feedback: Effects on professional practice. *Cochrane Database of Systematic Reviews*, 2025(3), Article CD000259. <https://doi.org/10.1002/14651858.CD000259.pub4>
- Lacroix, M., Abdelmalek, F., Everett, K., Taljaard, M., Salach, L., Bevan, L., Burton, V., Jia, H., Shuldiner, J., Laur, C., Angl, E. N., Ivers, N. M., & Tadrous, M. (2025). Effectiveness of audit and feedback and academic detailing interventions to support safer opioid prescribing in primary care. *The American Journal of Medicine*, 138(1), 70–78.e12. <https://doi.org/10.1016/j.amjmed.2024.09.017>
- McClatchey, K., Sheldon, A., Steed, L., Sheringham, J., Holmes, S., Preston, M., Appiagyei, F., Price, D., Taylor, S. J. C., Pinnock, H., & IMP²ART Programme Group. (2024). Development of theoretically informed audit and feedback: An exemplar from a complex implementation strategy to improve asthma self-management in UK primary care. *Journal of Evaluation in Clinical Practice*, 30(1), 86–100. <https://doi.org/10.1111/jep.13895>
- McCracken, R. K., Narayan, S., Maclure, M., Cooper, I., Cui, Z., Cullen, W., Dormuth, C., Hamilton, M. A., Nolan, S., Singer, J., Socías, M. E., Wong, S., & Klimas, J. (2023). Evaluation of audit and feedback to family physicians on prescribing of opioid analgesics to opioid-naïve patients: A pragmatic randomized delay trial. *Contemporary Clinical Trials*, 134, Article 107354. <https://doi.org/10.1016/j.cct.2023.107354>
- Meier, F. L., Jäger, L., Senn, O., Markun, S., & Burgstaller, J. M. (2025). Effectiveness of interactive dashboards as audit and feedback tools in primary care: A systematic review. *PLOS ONE*, 20(6), Article e0327350. <https://doi.org/10.1371/journal.pone.0327350>
- O'Connor, D. A., Glasziou, P., Maher, C. G., McCaffery, K. J., Schram, D., Maguire, B., Ma, R., Billot, L., Gorelik, A., Traeger, A. C., Albarqouni, L., Checketts, J., Vyas, P., Clark, B., & Buchbinder, R. (2022). Effect of an individualized audit and feedback intervention on rates of

- musculoskeletal diagnostic imaging requests by Australian general practitioners: A randomized clinical trial. *JAMA*, 328(9), 850–860. <https://doi.org/10.1001/jama.2022.14587>
- Omair, W., Jawaid, M. B., & Imam, A. A. (2025). Exploring the role of audit and feedback cycle in primary healthcare quality improvement. *BMJ Open Quality*, 13(Suppl. 1), Article e003089. <https://doi.org/10.1136/bmj-2024-003089>
- Roche, K. F., Morrissey, E. C., Cunningham, J., & Molloy, G. J. (2022). The use of postal audit and feedback among Irish general practitioners for the self-management of antimicrobial prescribing: A qualitative study. *BMC Primary Care*, 23, Article 86. <https://doi.org/10.1186/s12875-022-01695-x>
- Rouleau, G., Reis, C., Ivers, N., & Desveaux, L. (2023). Characterizing the gaps between best-practice implementation strategies and real-world implementation: Qualitative study among family physicians who engaged with audit and feedback reports. *JMIR Human Factors*, 10, Article e38736. <https://doi.org/10.2196/38736>
- Schwartz, K. L., Shuldiner, J., Langford, B. J., Brown, K. A., Schultz, S. E., Leung, V., Daneman, N., Tadrous, M., Witteman, H. O., Garber, G., Grimshaw, J. M., Leis, J. A., Presseau, J., Silverman, M. S., Taljaard, M., Gomes, T., Lacroix, M., Brehaut, J., Thavorn, K., . . . Ivers, N. (2024). Mailed feedback to primary care physicians on antibiotic prescribing for patients aged 65 years and older: Pragmatic, factorial randomised controlled trial. *BMJ*, 385, Article e079329. <https://doi.org/10.1136/bmj-2024-079329>
- Schwartz, K. L., Xu, A. X. T., Alderson, S., Bjerrum, L., Brehaut, J., Brown, B. C., Bucher, H. C., De Sutter, A., Francis, N., Grimshaw, J., Gunnarsson, R., Hoye, S., Ivers, N., Lecky, D. M., Lindbæk, M., Linder, J. A., Little, P., Michalsen, B. O., O'Connor, D., . . . Verheij, T. J. (2023). Best practice guidance for antibiotic audit and feedback interventions in primary care: A modified Delphi study from the Joint Programming Initiative on Antimicrobial resistance: Primary Care Antibiotic Audit and Feedback Network (JPIAMR-PAAN). *Antimicrobial Resistance & Infection Control*, 12, Article 72. <https://doi.org/10.1186/s13756-023-01279-z>
- Shuldiner, J., Kiran, T., Agarwal, P., Daneshvarfar, M., Eldridge, K., Kim, S., Greiver, M., Jokhio, I., & Ivers, N. (2023). Developing an audit and feedback dashboard for family physicians: User-centered design process. *JMIR Human Factors*, 10, Article e47718. <https://doi.org/10.2196/47718>
- Shuldiner, J., Lacroix, M., Saragosa, M., Reis, C., Schwartz, K. L., Gushue, S., Leung, V., Grimshaw, J., Silverman, M., Thavorn, K., Leis, J. A., Kidd, M., Daneman, N., Tradous, M., Langford, B., Morris, A. M., Lam, J., Garber, G., Brehaut, J., . . . Ivers, N. M. (2024). Process evaluation of two large randomized controlled trials to understand factors influencing family physicians' use of antibiotic audit and feedback reports. *Implementation Science*, 19, Article 65. <https://doi.org/10.1186/s13012-024-01393-5>
- Singer, A., Kosowan, L., Abrams, E. M., Katz, A., Lix, L., Leong, K., & Paige, A. (2022). Implementing an audit and feedback cycle to improve adherence to the Choosing Wisely Canada recommendations: Clustered randomized trial. *BMC Primary Care*, 23, Article 302. <https://doi.org/10.1186/s12875-022-01912-7>
- Tsang, J. Y., Peek, N., Buchan, I., van der Veer, S. N., & Brown, B. (2022). Systematic review and narrative synthesis of computerized audit and feedback systems in healthcare. *Journal of the American Medical Informatics Association*, 29(6), 1106–1119. <https://doi.org/10.1093/jamia/ocac031>
- van der Winden, D., van Dijk, N., Visser, M. R. M., & Bont, J. (2023). Walking the line between assessment, improvement and learning: A qualitative study on opportunities and risks of incorporating peer discussion of audit and feedback within quality improvement in general practice. *BMJ Open*, 13(1), Article e066793. <https://doi.org/10.1136/bmjopen-2022-066793>
- Vanstone, J. R., Patel, S., Berry, W., Degelman, M. L., Hanson, C., Phillips, C., & Parker, R. (2022). Using audit and feedback to encourage primary healthcare prescribers to record indications for antimicrobial prescriptions: A quality improvement initiative. *BMJ Open Quality*, 11(1), Article e001760. <https://doi.org/10.1136/bmj-2021-001760>
- Vicens, C., Leiva, A., Bejarano, F., Sempere-Verdú, E., Rodríguez-Rincón, R. M., Fiol, F., Mengual, M., Ajenjo-Navarro, A., Do Pazo, F., Mateu, C., Folch, S., Alegret, S., Coll, J. M., Martín-Rabadán, M., & Socías, I. (2022). Evaluation of a multicomponent intervention consisting of education and feedback to reduce benzodiazepine prescriptions by general practitioners: The BENZORED hybrid type 1 cluster randomized controlled trial. *PLOS Medicine*, 19(5), Article e1003983. <https://doi.org/10.1371/journal.pmed.1003983>
- Xie, C. X., Chen, Q., Hincapié, C. A., Hofstetter, L., Maher, C. G., & Machado, G. C. (2022). Effectiveness of clinical dashboards as audit and feedback or clinical decision support tools on medication use and test ordering: A systematic review of randomized controlled trials. *Journal of the American Medical Informatics Association*, 29(10), 1773–1785. <https://doi.org/10.1093/jamia/ocac094>
- Xu, A. X. T., Brown, K., Schwartz, K. L., Aghlmandi, S., Alderson, S., Brehaut, J. C., Brown, B. C., Bucher, H. C., Clarkson, J., De Sutter, A., Francis, N. A., Grimshaw, J., Gunnarsson, R., Hallsworth, M., Hemkens, L., Høye, S., Khan, T., Lecky, D. M., Leung, F., . . . Ivers, N. (2025). Audit and feedback interventions for antibiotic prescribing in primary care: A systematic review and meta-analysis. *Clinical Infectious Diseases*, 80(2), 253–262. <https://doi.org/10.1093/cid/ciae604>