

Reporting and design elements of audit and feedback interventions: a secondary review

ABSTRACT

Background: Audit and feedback (A&F) is a frequently used intervention aiming to support implementation of research evidence into clinical practice with positive, yet variable, effects. Our understanding of effective A&F has been limited by poor reporting and intervention heterogeneity. Our objective was to describe the extent of these issues. **Methods:** Using a secondary review of A&F interventions and a consensus-based process to identify modifiable A&F elements, we examined intervention descriptions in 140 trials of A&F to quantify reporting limitations and describe the interventions. **Results:** We identified 17 modifiable A&F intervention elements; 14 were examined to quantify reporting limitations and all 17 were used to describe the interventions. Clear reporting of the elements ranged from 56% to 97% with a median of 89%. There was considerable variation in A&F interventions with 51% for individual providers only, 92% targeting behaviour change and 79% targeting processes of care, 64% performed by the provider group, and 81% reporting aggregate patient data. **Conclusions:** Our process identified 17 A&F design elements, demonstrated gaps in reporting and helped understand the degree of variation in A&F interventions.

BACKGROUND

Audit and feedback (A&F) is a frequently used implementation intervention that has been used in a wide variety of clinical contexts, and evaluated in three Cochrane reviews and updates over the past 30 years [1-3]. While the overall effects of these interventions on clinical practice are positive (median adjusted risk difference of 4.3% absolute increase), they are also highly variable (interquartile range of 0.5% to 16%) with no evidence of increasing effect sizes over time [1]. A cumulative analysis of estimates of the effect of A&F by year and a cumulative meta-regression for common effect modifiers suggested that little new knowledge about the effects of A&F and potential effect modifiers has been generated in the last decade despite publication of 32 new trials [4]. Unless substantial improvements are made in how we design, deliver, and test A&F interventions, they will continue to be sub-optimal, working in some cases but not others, without clear knowledge as to why [5].

A&F has been defined broadly as 'any summary of clinical performance of health care over a specified period of time' [1]. In practice, A&F is a group of interventions with substantial variability in design, content and delivery. Progress in designing and delivering consistently effective A&F, and our understanding of effect modifiers, has been limited in part by lack of consensus about key design elements of A&F interventions and systematic reviews that are hindered by poor reporting of interventions in primary studies [1]. Further evidence for the lack of understanding comes from the fact that hypothesized causal mechanisms of change are rarely stipulated for A&F interventions [6].

The 2012 Cochrane review of A&F showed that feedback format (verbal and written), source (a respected colleague or supervisor), frequency, improvement strategies (goal setting and action planning), and baseline performance explained some variation in the effectiveness of A&F. These critical elements (i.e. the components that comprise the A&F intervention),

however, cannot be considered an exhaustive list of all the design decisions that need to be made when designing an A&F intervention. Five modifiable elements of A&F design have been identified by the Improved Clinical Effectiveness through Behavioural Research Group (ICEBeRG): content, intensity, method of delivery, duration, and context [7]. These elements, although limited in detail, provide a framework around which to consider design of A&F interventions. In this study we aim to elaborate these elements and document how they are reported to inform A&F design decisions, help hypothesize causal mechanisms for change, and facilitate knowledge synthesis on the effectiveness and mechanism of A&F including effect modifiers. We had three objectives:

- 1) Identify modifiable A&F-specific design elements
- 2) Document the quality of reporting these elements
- 3) Describe A&F interventions in current literature using this set of elements.

METHODS

We used a consensus-based process for developing the list of design elements, followed by an examination of existing A&F interventions to describe the quality of reporting and describe interventions in the existing literature.

To develop a list of modifiable A&F design elements, we (the author team) used our collective expertise and knowledge of related literatures, including a coding frame of A&F features developed to select and apply theory to synthesize A&F evidence [8], to create a list of A&F design elements that we believe to be modifiable, and applicable to most A&F interventions. Our nine-member research team has expertise in designing and testing A&F interventions, behaviour change, and implementation science. The team also included both clinical expertise and experience as A&F recipients. Consensus was developed through several meetings including one face-to-face author meeting. The resulting list of 17 elements

was organized into the following 6 categories: to **whom** the A&F was delivered (2 items), **what** audited information was delivered (10 items), **when** it was delivered (i.e. what was the lag time between practice and feedback; 1 item), **why** it was provided (i.e. what was the rationale for using A&F; 1 item), **how** it was delivered (2 items), and **how much** (i.e. the number of feedback instances delivered; 1 item). The specific 17 elements, written as the questions that were used for data extraction in the examination of the literature, are illustrated in Table 1.

Table 1: Modifiable A&F design elements

<i>Who</i>	<ol style="list-style-type: none"> 1. was the feedback given to an individual, a group, or both 2. was it given to the person in whom the practice change was desired (e.g. health care provider vs hospital administrator)
<i>What</i>	<ol style="list-style-type: none"> 3. was there feedback about the processes of care (e.g. rate of antibiotic prescription) 4. was there feedback about patient outcomes 5. was there feedback about something other than processes of care or patient outcomes (if yes, specified) 6. was the feedback about individual provider performance 7. was the feedback about the performance of the provider group 8. was the feedback about individual patient cases 9. was the feedback about an aggregate of patient cases 10. did the feedback identify a specific behavior(s) to be changed 11. what was the comparison provided in the feedback (specified) 12. were graphical elements included in the feedback
<i>When</i>	<ol style="list-style-type: none"> 13. what was the lag between the time of the audit and the delivery of the feedback (days, weeks, months, years, a mix)
<i>Why</i>	<ol style="list-style-type: none"> 14. what rationale was given for using A&F (specified)
<i>How</i>	<ol style="list-style-type: none"> 15. was the feedback given face to face 16. were providers explicitly asked to consider the implications the A&F had for their practice
<i>How much</i>	<ol style="list-style-type: none"> 17. what was the total number of times the feedback was given (specified).

We examined the A&F interventions evaluated in the 140 randomized controlled trials in the 2012 Cochrane review [1]. The Cochrane update spanned 1982 to 2011 and included the following databases: Cochrane Central Register of Controlled Trials,

MEDLINE, EMBASE, and CINAHL. For a full description of the methods used see Ivers et al., 2012[1]. To be included, trials had to examine and objectively measure either health professional practice or patient outcomes. Also, the interventions had to consist of a summary of clinical performance over a specified time period, with A&F a core aspect of the intervention for at least one intervention arm if a multifaceted intervention was used. As our interest was in exploring the broadest range of A&F interventions, the most complex study arm from each study was targeted for data abstraction. For example, if a study included control versus A&F versus enhanced A&F, we chose the latter.

A data extraction sheet and guide was developed in order to extract information on each of the elements in our list. The guide and sheet were piloted by three reviewers (HC, KC, MC) on five initial studies in the sample followed by a second pilot on 10 additional studies in the sample. Clarity improvements to terms and definitions of the elements were made after each of the pilots. KC and MC independently completed the variable extraction for each study with one reviewer reviewing all studies (HC). Disagreements were resolved through discussion and agreement.

Fourteen of the 17 elements had the response options of ‘*yes, no, unclear*’, or ‘*specified, unclear*’. These 14 elements, specifically the proportion of times the *unclear* response option was endorsed, formed the basis for determining the quality of reporting. For the remaining three elements, two did not have an ‘*unclear*’ response option as we considered that these questions were better answered as strictly yes/no questions (i.e. ‘was the A&F about something other than processes of care and patient outcomes’, and ‘were providers asked to specifically consider implications for their practice’). The third element that did not have an ‘*unclear*’ response option (i.e. the rationale for using A&F) was extracted by the reviewers as a description of the rationale in the paper from the reviewer’s viewpoint and was summarized *post hoc* (see below for a description of this process). Four elements included a

specified description: what the feedback was about, the comparison provided, the rationale for using A&F, and the number of times the feedback was given. For these elements, the summary categories were developed *post hoc* by two team members (HC, KC) and confirmed by a third team member (JB). This process involved considering the range of specified responses and coding them into categories that best reflected the data.

Descriptive statistics (e.g. frequencies, ranges, median) for each element were calculated. Quality of reporting was calculated by summarizing the total numbers of ‘unclear’ for each of the 14 elements that included an *unclear* response option; interrater reliability was analysed using the kappa statistic.

RESULTS

The elaborated A&F design elements, frequencies (n [%]), and examples of who the A&F was delivered to, what A&F information was delivered, when, why, and how much, are shown in Table 2.

Table 2. A&F design elements, frequencies (n [%]), and examples of who the A&F was delivered to, what A&F information was delivered, when, why, and how much

Elements	Frequency	Examples
Who the A&F was delivered to (2)		
1. Was it delivered to:		
a) Individual providers only	Yes: 72 (51%)	Sent to individual physicians by mail
b) Groups of providers only	Yes: 25 (18%)	Given to practice group at a meeting
c) Individuals and groups of providers	Yes: 23 (16%)	Sent to both
	Unclear: 20 (14%)	
2. Was the A&F given to the person in whom the practice change was desired?	Yes: 130 (92%) No: 6 (5%) Unclear: 4 (3%)	Given to service provider (yes) versus to directors of services or a hospital administrator (no)
What A&F information was delivered? (10)		
3. Processes of care?	Yes: 111 (79%)	Prescription rates, vaccination

Elements	Frequency	Examples
	No: 9 (6%) Unclear: 20 (15%)	rates, compliance with guidelines
4. Patient outcomes?	Yes: 20 (14%) No: 102 (73%) Unclear: 18 (13%)	Identification of patients with stroke, patient compliance
5. Other?	Yes: 45 (32%) No: 95 (68%)	Costs of laboratory tests ordered
6. Performance of individual provider?	Yes: 81 (58%) No: 50 (36%) Unclear: 9 (6%)	Individual prescription ordering
7. Performance of provider group?	Yes: 90 (64%) No: 38 (27%) Unclear: 12 (9%)	Entire practice group prescription rates
8. Individual patient cases?	Yes: 35 (25%) No: 95 (68%) Unclear: 10 (7%)	Patients identified who did not receive a particular preventive care action
9. Aggregate of patient cases?	Yes: 114 (81%) No: 15 (11%) Unclear: 11 (8%)	Percent of the patients in the practice who did not receive guideline consistent care
10. Was the specific behaviour to be changed identified?	Yes: 124 (89%) No: 9 (6%) Unclear: 7 (5%)	A&F on unnecessary test ordering in order to reduce test ordering (yes) versus A&F on costs of tests in order to reduce unnecessary test ordering (no)
11. What was the comparison provided?		
a) Others previous performance alone	Yes: 68 (49%)	The average of the group
b) Standardized guideline alone	Yes: 15 (11%)	A guideline
c) Own previous performance alone	Yes: 6 (4%)	Own performance in previous quarter
d) Others previous performance + standardized guideline	Yes: 6 (4%)	Both the average of the group and a guideline
e) Other	Yes: 9 (6%) Unclear: 36 (26%)	Internal standard
12. Graph presented?	Yes: 47 (36%) No: 73 (52%) Unclear: 20 (14%)	Line graph plotting monthly rates of lab test ordering for a practice group
When was A&F delivered? (1)		
13. What was the lag time?		
a) Days	Yes: 6 (4%)	Data from previous week
b) Weeks	Yes: 22 (16%)	Data from previous month
c) Months	Yes: 46 (33%)	Data from previous year
d) Years	Yes: 3 (2%)	Data from prior to previous year
e) Mix	Yes: 2 (1%)	Mix

Elements	Frequency	Examples
	Unclear: 61 (44%)	
Why was A&F delivered? (1)		
14. Rationale for using A&F ¹	Empirical only: 51 (37%)	A single study or previous Cochrane Review of A&F cited.
	Intuitive: 39 (28%)	Un-referenced statement about A&F being effective.
	No Rationale: 37 (26%)	A&F appears for the first time in the intervention description.
	Theory: 13 (9%)	Using Social Cognitive Theory constructs to design the intervention.
How was A&F delivered? (2)		
15. Face to face?	Yes: 62 (44%) No: 68 (49%) Unclear: 10 (7%)	Ground rounds, seminars, team meetings (compared to a mailout).
16. Providers asked to consider implications of A&F on their practice?	Yes: 32 (23%) No: 108 (77%)	Meetings to make actionable plans based on A&F results
How much A&F was delivered? (1)		
17. How much was given?		
a) Once	Yes: 33 (24%)	One time mailout
b) Twice	Yes: 21 (15%)	Feedback every 6 months for 1 year
c) Three	Yes: 13 (9%)	Feedback every 2 weeks for 6 weeks
d) Four	Yes: 13 (9%)	Feedback every 2 weeks for 6 weeks
e) > Four ²	Yes: 27 (19%) Unclear: 33 (24%)	Quarterly feedback for 1 year Every 2 weeks for one year

Notes: ¹Empirical only – statement that A&F was chosen based on empirical or synthesis that it is effective. Reference required.

Intuitive - general statements that A&F is worthwhile to include but not specified to be based on a named construct or theory; for example, A&F has not been used in this context so we are going to use it.

No rationale - A&F not mentioned until it appears in the intervention

Theory – a stated theory was used as a rationale for the intervention

² Range was 5-78.

The kappa statistic ranged from .88 to .37 with a median of .62. For 8 items, the kappa statistic was > .6 (substantial agreement), and for five items, the kappa statistic was > .4

(moderate agreement) [9]. The low kappa for the remaining item is likely the result of a high prevalence of 'yes' responses [10]; raters agreed on 87% of the ratings.

Quality of reporting

Clear reporting of the 14 elements ranged from 56% to 97% with a median of 89%. Elements with the clearest reporting were whether the A&F was given to the person in whom the practice change was desired (97% clear), if the A&F addressed the behaviour to be changed (95% clear), and if the A&F was about a provider's individual performance (94% clear). Elements that demonstrated the poorest reporting were the lag time between the audit and feedback components (56% clear), the nature of the comparison (74% clear), and the total number of times A&F was given (76% clear).

Description of current A&F interventions

Who: A&F was primarily given to individuals only (51%), with 18% being given to a group and 16% given to both the group and the individuals in the group. The majority of the time (92%), the A&F was given to the target person rather than, for example, an administrator who was not the intervention target.

What: Rarely was feedback given on patient outcomes (14%); instead feedback was mostly about processes of care (79%) such as rate of antibiotic prescription. However, 32% of the time, feedback content included information other than processes of care and patient outcomes. Most of this 'other' content included patient level data ([19/45] e.g. names of specific patients at risk for a condition) or cost data ([13/45] e.g. the costs of the prescribed medications). More than half of the studies (58%) used feedback that provided information about an individual's own behaviour or patient cases. Feedback mostly presented aggregated patient data (81%), rather than feedback about individual patient's care (25%). The most common comparison used was to peers' performance or 'others' previous performance'

(49%). Fifteen percent included a standardized guideline as a comparator, and 4% used the person's own previous performance. Feedback mostly identified the specific behaviour to be changed (86%). An example of feedback that did not identify the behaviour to be changed included giving feedback on the total cost of inappropriate antibiotic prescription as opposed to the provider's rate of inappropriate antibiotic prescriptions. Graphical representation of the data was found in only 36% of the interventions.

When: Lag time (the time between the collection of data for feedback and the provision of that data) was most commonly months (33%). Rarely was the A&F provided based on a fast turnaround such as days or weeks.

Why: The rationale for using A&F ranged from empirical evidence only in 37% of the cases, an intuitive rationale in 28% of the cases and no rationale in 26% of the cases. As has been reported elsewhere, the remaining 9% of the cases used theory as the rationale for the intervention [6]. An example of an intuitive rationale was a statement that explained the rationale for why A&F might lead to behaviour change but was stated without a reference to empirical evidence or to a named theory or construct, e.g. "We thought feedback on generic prescribing to be less threatening to physicians than offering judgments on proper or improper use of drugs" [11] (p 194) or "The hypothesis underlying the hospital-specific feedback was that physicians and nurses were more likely to change their behaviour if made aware that their practices fell short in providing best care as defined by clinically meaningful quality criteria agreed upon in advance by credible authorities in their hospitals" [12] (p 1637). In the cases where no rationale was stated, A&F was not mentioned at all until it appeared in the intervention description.

How: A&F was given face-to face-in about half of the studies (44%). In only 23% of the cases were providers explicitly asked to consider the implications that A&F had on their practice.

How much: A&F was delivered on one occasion 24% of the time, 15% twice, 9% three times, 9% four times, and 19% more than four times (range 5-78).

DISCUSSION

As an initial step towards the ultimate goal of facilitating the design and delivery of consistently effective A&F, we have developed a preliminary list of 17 design elements organized according to the six categories of who the A&F was delivered to, what A&F information was delivered, when, why, how, and how much. Our quantification of related reporting limitations in primary studies indicated a range of clear reporting of 14 design elements from 56% to 97%. A&F most often includes a report of processes of care (79%) performed by either the provider group (64%) and/or individual providers (58%), based on aggregate patient data (81%), and with a comparison to the performance of others within the group (49%). Less common was providing data on individual patient cases (25%), the use of a guideline for the comparison (15%), and the use of graphical elements to describe the data (36%).

Quality reporting of interventions is essential for replication of study findings, understanding how and why an intervention might work, and for translating effective interventions into practice [13]. The proliferation of reporting guidance in the previous decade is, in a large part, a response to these needs. For some elements in our findings, reporting was clearer than for others. For instance, 24% were unclear as to the total number of times feedback was given. Van Hoof and colleagues quantified reporting quality of another commonly used implementation intervention, educational outreach [14]. Based on 25 identified design elements, reporting quality for each element ranged from 0-100%. The reporting of A&F interventions in our review did not show a similarly wide range; however, none of our elements reached 100% reporting. In both of these studies of design elements for

common implementation interventions, additional work is needed to confirm the list of elements.

While guidance exists for the reporting of interventions [13 15 16], none include the full range of elements we have proposed here for a detailed description of an A&F intervention. For example, the Template for Intervention Description and Replication (TIDieR) checklist includes reporting on the frequency with which the intervention is given, and to specify if provided individually or in a group, but does not specify reporting the nature of the comparison in A&F or whether it is based on processes of care or patient outcomes [13]. The development of reporting guidance for every type of implementation intervention is potentially unwieldy, but this study suggests a need to improve methods to achieve the detail necessary for reporting A&F interventions.

Our process identified 17 elements in this family of interventions generally referred to as 'A&F'. Our findings are consistent with the ICEBeRG framework and additionally enables a more precise and complete description of A&F interventions. The ICEBeRG element of content is consistent with our element of what will be delivered, intensity and duration with how much, and method of delivery with how. The only ICEBeRG element not included in our list of elements was context but given the lack of clarity as to the definition of context in the ICEBeRG list and the limited understanding of context in the field of health care practice improvement, we see this as a future area of focus. Ongoing efforts to examine the meaning of context [17] will potentially aid in this regard.

Further work to examine the most and least endorsed responses for the 17 elements as compared to what is known about optimal A&F could provide an opportunity to examine where the typical A&F intervention might be sub-optimal and the generation of hypotheses for further testing. For example, theoretical perspectives indicate that A&F is more effective if the comparison is empirically-supported, such as a guideline, and if the intervention

includes multiple modalities (e.g., a graph plus text) [20]. The typical A&F found in this study indicated only 15% of studies using a guideline for the comparison and the use of graphical elements to describe the data in only 36% of the studies. Attention should be paid to the basis for the comparison and the presence of multiple modalities in A&F. Further studies should investigate best practices for multiple modalities and the identification of optimal comparisons. Current best evidence suggests that A&F will be more effective when delivered on more than one occasion [1], yet 24% of the studies in this review delivered A&F on only one occasion. Increased attention to how many times A&F is delivered is needed by A&F designers.

In this review, the reported rationale for using A&F was rarely based on a theory or related constructs (9%); in 26% of trials, investigators reported no rationale for A&F at all. To date, theoretical predictions for designing and delivering optimal A&F interventions have rarely been incorporated into the design of A&F interventions - or at least, rarely stated in study reports [6 18]. For a full description of the extent and types of theories used in these 140 trials of A&F see Colquhoun et al [6]. Of note in our study is the 28% of studies that used intuitive, but not explicitly theory-based, hypotheses about how the A&F might work. Although these hypotheses can be tested, we would argue that having the added value of defined constructs and construct relationships found in explicit theories offers the most effective way of increasing our understanding of the mechanisms of action by which A&F works. An additional 36% of the studies used empirical evidence as the rationale for using A&F. While some of this cited evidence were systematic reviews, we observed that many were single studies implying the misguided belief that success of A&F in one context will lead to success in another.

Several limitations of this study warrant discussion. In using the Cochrane update on A&F as our data set, we were limited to randomised controlled trials only. Including other

study designs might have yielded additional information on A&F interventions. We recommend that further work in this area includes a broader criteria related to study design. Also, we only looked at A&F in research studies; A&F used in routine health care settings (i.e., not within research settings) might yield additional information. The development of our list of modifiable A&F elements was expert-driven and preliminary. While we used available expertise and literature to develop this list, further refinement and validation is required. For example, a broader modified Delphi study incorporating views from international experts, A&F intervention designers, and those who have specific expertise in the theory and practice of A&F intervention delivery could be used to further develop our preliminary list of elements. We categorized our list of elements in order to aid understanding (e.g. who was the A&F given to, what was given) but other approaches to categorization could be considered. Due to the approach we used for response options, we were only able to investigate reporting quality for 14 of the 17 design elements. We described reporting quality as the number of unclear responses for each element, as opposed to a separate question regarding clarity of reporting, which could be considered. As we used the dataset from the Cochrane review on A&F, the studies ranged in years from 1980-2011. Given the proliferation of reporting guidance in the last decade [13 19], it is possible that examining the quality of reporting in the previous decade only would indicate improvements in reporting quality. Future studies of reporting quality should consider analysis by decade.

Based on this preliminary work, we propose that A&F intervention designers should, at a minimum, explicitly consider and offer justification for their decisions regarding who the A&F will be delivered to, what information will delivered, when it will be delivered (in relation to the collection of the data), why, how, and how much. Attention to all 17 elements is recommended or at least a decision made as to the most relevant elements to consider. Equal attention to the reporting of these elements is warranted.

CONCLUSION

We have quantified reporting quality for 14 modifiable A&F elements based on 140 trials of A&F. We have also described typically employed A&F interventions for 17 elements and proposed a preliminary list of A&F elements to be considered when designing an A&F intervention. Our results provide a starting point for what should be considered in the design and reporting of A&F interventions.

Incorporating both empirical and theoretical evidence into intervention designs would facilitate understanding of how to optimize this commonly applied intervention. Future research needs to focus on further development of this set of A&F design elements, best practices for each of the elements, and the development of reporting guidance specific to A&F. Review teams should consider involving a broad range of study designs, reviews of A&F interventions delivered as a routine part of health care practice, and detailed examination of where current A&F interventions are suboptimal to guide the design of future A&F interventions. Recent suggestions have been made to consider A&F intervention elements as potential levers that should be considered and manipulated together [20]. For example, if A&F can only be delivered once, other elements should be considered for enhancement. A robust set of A&F design elements would facilitate these decisions, reporting quality, and the value of additional syntheses including investigation of effect modifiers.

REFERENCES

1. Ivers N, Jamtvedt G, Flottorp S, et al. Audit and feedback: effects on professional practice and healthcare outcomes. *Cochrane Database Syst Rev* 2012;**6**
2. Jamtvedt G, Young JM, Kristoffersen DT, O'Brien MA, Oxman AD. Does telling people what they have been doing change what they do? A systematic review of the effects of audit and feedback. *Quality and Safety in Health Care* 2006;**15**(6):433-36
3. Jamtvedt G, Young JM, Kristoffersen DT, O'Brien MA, Oxman AD. Audit and feedback: effects on professional practice and health care outcomes. *Cochrane Database Syst Rev* 2006;**2**(2)
4. Ivers NM, Grimshaw JM, Jamtvedt G, et al. Growing literature, stagnant science? Systematic review, meta-regression and cumulative analysis of audit and feedback interventions in health care. *Journal of general internal medicine* 2014;**29**(11):1534-41 doi: 10.1007/s11606-014-2913-y[published Online First: Epub Date].
5. Foy R, Eccles M, Jamtvedt G, Young J, Grimshaw J, Baker R. What do we know about how to do audit and feedback? Pitfalls in applying evidence from a systematic review. *BMC health services research* 2005;**5**(1):50
6. Colquhoun HL, Brehaut JC, Sales A, et al. A systematic review of the use of theory in randomized controlled trials of audit and feedback. *Implementation science : IS* 2013;**8**:66 doi: 10.1186/1748-5908-8-66[published Online First: Epub Date].
7. Improved Clinical Effectiveness through Behavioural Research G. Designing theoretically-informed implementation interventions. *Implementation science : IS* 2006;**1**:4 doi: 10.1186/1748-5908-1-4[published Online First: Epub Date].
8. Gardner B WC, McAteer J, Eccles M, Michie S. University College London Using theory to synthesise evidence from behaviour change interventions, 2010.
9. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *biometrics* 1977:159-74
10. Viera AJ, Garrett JM. Understanding interobserver agreement: the kappa statistic. *Fam Med* 2005;**37**(5):360-63
11. Gehlbach SH, Wilkinson WE, Hammond WE, et al. Improving drug prescribing in a primary care practice. *Medical care* 1984;**22**(3):193-201
12. Lakshminarayan K, Borbas C, McLaughlin B, et al. A cluster-randomized trial to improve stroke care in hospitals. *Neurology* 2010;**74**(20):1634-42
13. Hoffmann TC, Glasziou PP, Boutron I, et al. Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. *BMJ: British Medical Journal* 2014;**348**
14. Van Hoof TJ, Miller NE, Meehan TP. Do published studies of educational outreach provide documentation of potentially important characteristics? *American journal of medical quality : the official journal of the American College of Medical Quality* 2013;**28**(6):480-4 doi: 10.1177/1062860613476335[published Online First: Epub Date].
15. Ogrinc G, Mooney S, Estrada C, et al. The SQUIRE (Standards for Quality Improvement Reporting Excellence) guidelines for quality improvement reporting: explanation and elaboration. *Quality and Safety in Health Care* 2008;**17**(Suppl 1):i13-i32
16. Boutron I, Moher D, Altman DG, Schulz KF, Ravaud P. Methods and processes of the CONSORT Group: example of an extension for trials assessing nonpharmacologic treatments. *Annals of Internal Medicine* 2008;**148**(4):W-60-W-66

17. Squires JE, Graham ID, Hutchinson AM, et al. Identifying the domains of context important to implementation science: a study protocol. *Implementation Science* 2015;**10**(1):1-9
18. Davies P, Walker AE, Grimshaw JM. A systematic review of the use of theory in the design of guideline dissemination and implementation strategies and interpretation of the results of rigorous evaluations. *Implementation science : IS* 2010;**5**(14):5908-5
19. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Annals of internal medicine* 2009;**151**(4):264-69
20. Ivers NM, Sales A, Colquhoun H, et al. No more 'business as usual' with audit and feedback interventions: towards an agenda for a reinvigorated intervention. *Implementation science : IS* 2014;**9**(1):14