# Audit and Feedback to Improve Laboratory Test and Transfusion Ordering in Critical Care:

# A Systematic Review & Assessment of Adherence to Best Practices

MADISON K FOSTER

Thesis submitted to the Faculty of Graduate and Postdoctoral Studies in partial fulfillment of the requirements for the MSc degree in Epidemiology

School of Epidemiology and Public Health

Faculty of Medicine

University of Ottawa

## ABSTRACT

Audit and feedback (A&F), provision of performance data, can be an effective means to change behaviour. We hypothesize that A&F may be useful in the complex environment of critical care to address routine ordering behaviours. This thesis describes a systematic review assessing the use and effectiveness of A&F for the improvement of intensive care unit laboratory test and transfusion ordering and an analysis of how these interventions adhere to recent suggestions for best practice. The review identified relatively few published studies, almost all of which involved multi-component interventions; these were generally found to be moderately effective. Through development of an evaluation tool, we found recent suggestions for best practice may be underutilized in existing A&F interventions. This work has identified several priorities for future research to aid in the optimization of critical care A&F interventions aimed at improving test and transfusion ordering.

## ACKNOWLEDGEMENTS

Completion of this thesis would not have been possible without the help and support from many others. I would first like to graciously thank my supervisors Dr. Jamie Brehaut and Dr. Justin Presseau. Thank you for introducing me to the field of Knowledge Translation, and for allowing me the opportunity to take on such an interesting project! Dr. Brehaut's and Dr. Presseau's expertise, guidance and wisdom provided me with the support necessary to successfully finish this thesis. Thank you for always taking the time to provide comments, answer questions and for your patience!

I would also like to thank Dr. Lauralyn McIntyre for her insight into the critical care setting and Dr. Brian Hutton for his expertise in systematic reviews. Thank you both for your invaluable feedback. To Kelly Carroll, Dr. Nicola McCleary and Maria Papoulias, thank you for your help in screening, coding, and developing the evaluation tool. I greatly appreciate the time you spent to help with this project and for enduring countless consensus meetings.

Next, I would like to thank my friends from the MSc. Epidemiology cohort. I could not have asked for a better group of classmates with whom to share this experience with. Thank you for always offering a helping hand when needed and making assignments, studying and writing much more enjoyable.

To my family and friends, who have always encouraged me to further pursue my education, thank you for your love, support and kind words of encouragement. Thank you especially to my Dad, Andy, and brother, Casson. In loving memory of Kimberley Foster.

Finally, I would like to thank my partner Amit Ubhi for his unwavering support. Thank you for the countless rides, home-cooked meals when I needed to keep working away, and for always encouraging and believing in me.

## TABLE OF CONTENTS

**LIST OF TABLES**

**Chapter 1**



**Chapter 2**



**Chapter 3**

**LIST OF FIGURES**

**Chapter 1**



**Chapter 2**



**Chapter 3**

# Chapter 1 Introductory Chapter

Laboratory tests are used extensively in healthcare; in addition to monitoring patient status, tests may be used for checking medication levels and screening for disease.[1,2] Now ingrained in medical practice,[1] these innovative tools provide countless benefits, including detection of disease,[3] and identification of specific cancers and bacterial strains so that treatments may be customized to the patient.[4,5] Laboratory tests can also provide valuable information to aid in patient management decisions; it has been hypothesized that over two thirds of patient care decisions are guided by the outcome of laboratory tests.[6–9] As laboratory results can determine treatment course and whether a patient will be admitted to or discharged from the hospital, they also have a great impact on further resource use.[1,6,8,9]

With a study from as early as the 1960's remarking an "explosive" increase in laboratory testing,[10–12] growth in healthcare testing has been observed over many years.[13] For instance, an increase of 6-7% was observed for American hospital laboratory tests over the time period of 1999 to 2006.[14] In the United Kingdom, an increase of 24.2% was seen for primary care laboratory tests over the time period of 2005 to 2009.[7] Another study found a 6-8% annual rise in tests conducted in Calgary, Alberta.[12] In Eastern Ontario (Ottawa included), just over 13 million tests were completed for the fiscal year of 2015/2016, an increase of 2.1% from the previous fiscal year.[15]

Overall, laboratory tests are associated with considerable cost. In 2007, the United States laboratory services industry was expected to yield a profit of 52 billion dollars, with an estimated 6.8 billion tests conducted per year.[14] A 2006 report stated that over 697 million laboratory tests were completed in England over the span of a year, leading to an expenditure of approximately 2.5 billion pounds.[6,7] In Canada, growth in laboratory service expenditure has been observed in several provinces. Between the fiscal years of 1996/1997 and 2001/2002, increases in costs of 8%, 8%, 14%, 15% and 34% were seen for Ontario, Manitoba, Alberta, Saskatchewan and British Colombia respectively.[16] Out of these provinces, Ontario was found to have the second highest per capita expenditure in 2001/2002 ($90.41).[16] A more recent analysis estimates that Ontario laboratory facilities conducted over 244 million tests during the fiscal year of 2013/2014, which equates to approximately 2.05 billion dollars of funding.[17]

## Inappropriate Use of Resources in Healthcare

During the late 1990s, overuse in healthcare was identified as one of several urgent quality issues by the Institute of Medicine's "National Roundtable on Health Care Quality".[18] Waste in healthcare has since been a focus of concern at dedicated workshops and conferences, such as "the Institute of Medicine (IOM) Roundtable on Value & Science-Driven Health Care" workshops[19] and an "Avoiding Avoidable Care" conference.[20,21] Of U.S. health services expenditure, an upper value of 30% has been approximated to be unnecessary.[19,20,22,23] This figure is consistent with a more recent analysis of several Canadian tests and interventions which observed that 30% may have been avoidable.[24] A commonly cited motivation behind reducing unnecessary healthcare is a financial one.[25] However, inappropriate medical care may also cause adverse events, and can affect the care of other patients by reducing the availability of resources.[22,26] Raising awareness about potentially unnecessary healthcare services has thus become an important priority for a variety of reasons.

**Inappropriate Testing**

Unnecessary testing is commonly identified as a focus for cost containment programs, due to the cascading effect of tests on further resource use.[8,27,28] While it can be difficult to articulate a general definition for an 'inappropriate' test, several have been suggested in the literature. In a study that looked at the self-reported and perceived 'unnecessary' ordering by residents, the authors used the definition of "a test that would not change management regardless of its result".[29] Several additional scenarios were identified in another article; laboratory tests that are "repeated at too frequent intervals," "ordered when clinical assessment is superior," "ordered to confirm an expected response to a routine intervention," "redundant laboratory tests ordered concurrently" and those which "[do] not affect management or prognostication".[30] A recently published textbook, focussed on laboratory utilization management, also conveyed the difficulty in providing a definition for inappropriate tests and stated that, "Although clinical guidelines exist for some types of testing, often there is no peer-reviewed literature defining appropriate test utilization… In most cases we rely on local clinical experts to provide guidance or meet with clinicians to try to reach a consensus."[27]

Evidence to suggest inappropriate ordering of laboratory tests is the observed variation between providers and institutions without discernible differences in patient outcomes.[2,11] Studies have shown variation in provider and institution test ordering (or resource use including laboratory tests) in primary, outpatient, clinic care and inpatient hospital care.[7,31–37] Some of these studies additionally showed that this variation occurs without significant differences in patient outcomes such as mortality[31,37] and length of stay.[31] This suggests that some sites and settings are capable of safely providing care with use of fewer tests, and in turn that implementation of interventions in settings with variable resource use, may be valuable in reducing unnecessary orders.[2,31]

Many studies have explored the magnitude of inappropriate ordering for a variety of tests and contexts. For example, studies have estimated the proportion of potentially inappropriate or unnecessary ordering of ionized calcium tests to be 97%,[38] arterial blood gas analyses (ABGs) to be 30%,[39] blood tests (cholesterol, hemoglobin $A_{1c}$, thyroid- stimulating hormone, vitamin $B_{12}$, vitamin D, ferritin) to be 16%,[40] thyroid tests to be 10%,[41] and serum sodium tests to be 5.1%.[42] A range of 4.5%-95% inappropriate utilization was documented by a review that included studies from 1966-1997.[1,13] A more recent 2013 systematic review and meta-analysis, identifying 38 studies between 1997 and 2012, showed a 20.6% mean rate of inappropriate overutilization.[1] A varied range in inappropriate ordering can therefore be seen across the literature. Additionally, the spread in test type and setting across these studies suggests that inappropriate ordering is a widespread issue.

Many studies have also estimated the expenditure associated with these potentially wasteful practices. One Eastern Ontario study estimated that $13.9-35.9 million dollars-worth of testing may have been redundant between 1999-2000, for several frequently used tests.[43] Another study estimated that potentially unnecessary repeated tests identified among 103,000 patients cost between $0.6-$2.2 million dollars throughout the course of one year.[40] One criticism of these types of estimations, however, is that it can be difficult to determine true expenditure reductions associated with reducing unnecessary laboratory tests.[27]

**Influences on Test Ordering**

A number of factors have been argued to contribute to the overuse of laboratory tests. Several of these factors are dependent on the physician.[44,45] Knowledge of the ordering physician is one such factor, as lack of knowledge regarding price, the fact that a test is obsolete, or that a test should not be repeated within a certain timeframe, could lead to the inappropriate ordering of a test.[2,30,45–48] Fear of medico-legal issues may lead to defensive and potentially inappropriate test ordering.[7,44] Furthermore, beliefs regarding the utility of a laboratory test, or uncertainty, may influence a provider's decision to order.[2,44,48] Habit and training have also been suggested by residents to play a role in the ordering of potentially inappropriate tests.[29]

The manner in which tests can be ordered may influence ordering behaviour. For example, presence of a computer provider order entry (CPOE) system may allow for more efficient, but potentially excessive ordering.[2] Further to this, order form format has been shown to affect test-ordering volume. For instance, the un-bundling of tests (creating separate entities) has been shown to reduce the number of individual tests ordered.[49] Similarly, systems by which tests can be ordered repeatedly through one order (standing orders) may result in unnecessary tests if the healthcare provider does not remember to stop the standing order when no longer needed.[42,46,50,51] One survey also found that the nomenclature of the tests can lead to physician uncertainty as to whether they are ordering the correct test.[45,52]

Aspects of the environment such as hospital characteristics may also affect test-ordering practices. For example, Zimmerman et al. found that teaching hospitals ordered more blood tests for ICU patients, as compared to non-teaching hospitals.[53] Certain settings, such as critical care, may order some tests routinely (based on a time interval) rather than based solely on medical necessity.[54,55] Additionally, having access to an arterial line, which facilitates blood sample collection, appears to increase the frequency of blood sampling.[56,57]

Social pressures may influence provider test ordering. Pressure from patients (or the perception that patients expect testing to be completed) may lead to ordering unnecessary tests.[7,45,58] As well, pressure (real or perceived) from superiors may lead to junior healthcare providers ordering unnecessary tests.[29,59] Finally, another factor affecting test ordering is that of the payment system,[44,47,60,61] which can vary depending on a country's funding model. For example, in payment systems where there is incentive to order particular or more tests for reimbursement, unnecessary tests may be ordered.

**Underutilization**

There are also scenarios where a laboratory test may be underutilized. For example, a study assessing an intervention to improve compliance with the Surviving Sepsis Campaign resuscitation bundle[62,63] found the average compliance for ordering lactate tests and blood cultures (in advance of treatment with antibiotics) to be 61.0% and 64.5% respectively, during intervention implementation.[62] This suggests room for improvement in terms of test ordering for severe sepsis patients.[63] In a recent systematic review and meta-analysis assessing "the landscape of inappropriate laboratory testing"[1] however, the authors note that the topic of underutilization appears to be less well studied in the published literature. The review only identified eight studies assessing underutilization, in comparison to the 38 studies which assessed overutilization.[1] This suggests that a greater number of opportunities to reduce overutilization exist, or at least have been identified, as compared to opportunities to improve underutilization.

## Initiatives for Change

Several organizations have marked judicious test ordering as a major priority through establishment of local and national initiatives. Choosing Wisely campaigns by the American Board of Internal Medicine (ABIM), and the Canadian Medical Association/University of Toronto are examples of such initiatives. These campaigns focus efforts on appropriate test use, as well as the appropriate use of treatments and services.[64,65] A key message of the Choosing Wisely Canada campaign is that "more is not always better".[66] Both campaigns put emphasis on the importance of discussions between patients and healthcare providers regarding the necessity of a treatment or test.[64,65] Currently for Choosing Wisely Canada, 38 lists (227 recommendations) have been developed by different associations, ranging across a variety of specialties (anesthesiology, critical care, family medicine, general surgery, palliative care, psychiatry, etc.).[66] **Table 1** shows examples of several Choosing Wisely Canada guidelines related to test ordering.

| Society | Summary of Guidance |
|---|---|
| **CMA's Forum on General and Family Practice Issues & College of Family Physicians of Canada** | "Don't do annual screening blood tests unless directly indicated by the risk profile of the patient."[67] |
| **Canadian Anesthesiologists' Society** | "Don't order baseline laboratory studies (complete blood count, coagulation testing, or serum biochemistry) for asymptomatic patients undergoing low-risk non-cardiac surgery."[68] |
| **Canadian Society of Endocrinology and Metabolism** | "Don't routinely test for Anti-Thyroid Peroxidase Antibodies (anti – TPO)."[69] |
| **Canadian Federation of Medical Students Fédération médicale étudiante du Québec** | "Don't suggest ordering tests or treatments pre-emptively for the sole purpose of anticipating what your supervisor would want."[70] |
| **Canadian Association of Pathologists** | "Avoid standing orders for repeat complete blood count (CBC) on inpatients who are clinically/laboratorily stable."[71] |

**Table 1. Examples of Choosing Wisely Canada guidelines related to test ordering.**

The geographical span in test and healthcare service utilization initiatives further indicates that this is a widespread issue, now recognized as a priority. In addition to Canada and the United States, sixteen other countries have instituted Choosing Wisely initiatives.[22] The reduction of unnecessary testing has also been flagged as a priority by the United Kingdom's National Institute for Health and Care Excellence (NICE), which has established guidelines regarding the appropriateness of several laboratory tests prior to elective surgery.[72] Furthermore, a 2014 Canadian Agency for Drugs and Technologies in Health (CADTH) report identified numerous other provincial and global initiatives working towards improved laboratory test ordering.[73] Provincially, Health Quality Ontario (HQO) has published recommendations in regards to several community-based laboratory tests.[74] These recommendations were put forth as a part of HQO's "Appropriateness Initiative", which aims "to develop a systematic framework for the ongoing identification, prioritization, and assessment of health interventions in Ontario for which there is possible misuse, overuse, or underuse."[74]

**Changing Practice**

Although the reduction of unnecessary testing has been identified as a priority by these organizations, it seems there is a lack of published literature documenting how these campaigns, guidelines and initiatives have impacted ordering practices. A 2014 CADTH report highlighted laboratory utilization initiatives and noted this gap in the published literature, stating that this makes it difficult to determine which are successful.[73] A few studies have assessed changes in practice before versus after publication of the Choosing Wisely guidelines. One observational study used insurance claims data to assess changes in the ordering of low-value imaging for back pain after implementation of Choosing Wisely guidance and found overall a small (3.8%) change in practice.[75] An earlier study was also completed in 2015, which assessed changes in claims data for seven services (including imaging, lab testing, medication and antibiotic prescribing) each of which was associated with a Choosing Wisely guideline.[76] This study also found small, significant decreases for some of the resources (1.1%-1.5%).[76] Authors of both papers however reflect on the fact that further strategies will be required to increase compliance with these guidelines.[75,76]

Implementation of best practices can be difficult. Surveys have shown awareness/familiarity among medical professionals (primary care physicians, medical and surgical specialists) with the Choosing Wisely campaign, with proportions ranging from 21% to 66.7%.[58,77] However, studies such as those in a review have indicated that a multitude of other barriers to using guidelines exist.[78] Beyond knowledge and familiarity, factors such as physician disagreement with the guideline, and discordant patient preferences, may contribute to non-adherence.[78] The "current medical malpractice system", "patient requests", "[recommendations] by specialists", and "lack of time for shared decision making" were also rated as barriers to addressing overutilization, by 43.8% or more of surveyed primary care physicians.[79] Furthermore, it has been proposed that habitual behaviours are challenging to alter.[47,78,80] As reducing overutilization requires providers to decrease a behaviour they have likely become familiar with, designing interventions to change this practice may require additional consideration.

There is a vast literature assessing different interventions that might be used to implement the campaigns mentioned above. For instance, a systematic review of reviews concluded that 'active' Continuing Medical Education (CME) interventions (i.e. academic detailing) tended to yield better results than 'passive' interventions (i.e. printed educational material) in terms of behaviour change.[81] Several reviews have specifically assessed the effectiveness of a variety of clinician behaviour change interventions for the improvement of provider laboratory test ordering. A review by Kobewka et al.[82] assessed laboratory test ordering in all healthcare settings, while systematic reviews by Cadogan et al.[45] and Thomas et al.[8] focussed specifically on test ordering in primary care. All reviews identified a variety of interventions to have been tested in the literature, including one or a combination of A&F, education, incentive and penalty, and "system-based" or administrative interventions.[8,45,82] Moreover, all reviews[8,45,82] found a variety of intervention strategies to be effective in changing provider behaviour, with Kobewka et al.'s results showing multifaceted interventions to have a greater effect than that of single component interventions.[82] None of the reviews conducted meta-analyses, due to heterogeneity between studies, and all found that the literature showed a large range in effect size.[8,45,82] Both Kobewka et al. and Cadogan et al.'s reviews comment on the overall low quality of the included studies, suggesting further work remains to improve study conduct, with use of more rigorous study design methods.[45,82] Additional suggestions for intervention

implementation in the literature include gaining "leadership level support" and taking context-specific considerations into account by conducting a barriers assessment.[2,45]

**Transfusions: A Similar Story**

While transfusions are a life-saving therapy, recent initiatives have aimed to reduce their use, as transfusions are associated with non-trivial risks such as nosocomial infection and transfusion reactions.[30,83–91] Even though the risk of blood-borne diseases, such as Human Immunodeficiency Virus (HIV), have been reduced through screening, there is still concern for increased risk of infections due to Transfusion Related Immune Modulation (TRIM).[84,88] Other complications include Transfusion-Related Acute Lung Injury (TRALI) and Transfusion-Associated Circulatory Overload (TACO).[88]

The serious potential risks associated with transfusion, and limited supply of blood products have led to interest in assessing whether transfusions may be given at lower hemoglobin levels, in non-bleeding patients (i.e. treatment where transfusion is 'triggered' by a lower haemoglobin value).[84,89] A Cochrane review assessing 23 studies comparing liberal and restrictive transfusion strategies found that there were no significant differences in patient morbidity or mortality.[89] The results indicate that restrictive RBC transfusion strategies may be safe to institute for some patient groups.[89] Use of a restrictive transfusion policy is also beneficial in that it promotes appropriate use of a limited and precious resource, to ensure it is available for those whom it will provide benefit to.[89]

Studies from China, England and Iran have found inappropriate rates of RBC transfusion, ranging between 15%-30.9% of transfusions and 19-48% of patients.[92–94] More locally, an audit assessing a sample of transfusions completed in 10 Ontario hospitals during 2015, saw rates of inappropriate ordering ranging between 0%-68% (overall inappropriate rate= 22%).[95] As seen in **Table 2**, both the American and Canadian Choosing Wisely initiatives have published many guidelines related to transfusion orders, as there is evidence suggesting a restricted strategy over a liberal one.[89] Similarly, inappropriate ordering of other blood components (platelets, plasma and cryoprecipitate) has been described in the literature .[92,93,96–99] Inappropriate use of blood products thus represents another gap in the translation of evidence to practice, and it appears further efforts will be required to rectify this.

We identified two systematic reviews that focused on the application of behaviour change interventions to improve transfusion ordering behaviour. The first review[100] was published in 2002, and included studies that used single or multifaceted interventions to reduce inappropriate transfusion. The intervention categories as coded by the authors included education, retrospective audits, prospective audits, an algorithm, request forms, and "patient-specific decision support."[100] This review found that all types of interventions were capable of improving transfusion ordering behaviour, however as all but one used an uncontrolled before-after design, were subject to several limitations. In 2005, the same research group conducted and published another systematic review of the literature.[101] This review included studies aiming to generally reduce transfusion (rather than focussing on the reduction of inappropriate transfusion). In this review, interventions were coded as one or more of the following: education, guideline, audit/feedback, audit/approval, or form. Similar to their previous review, the authors found all different types of interventions to be effective in changing provider behaviour, leading to either an overall reduction in transfusion ordering or in inappropriate transfusion ordering. All but two of the studies used an uncontrolled before-after design. Thus, it was again concluded that future work should focus on the comparison of

different interventions, using a controlled study design.[101] Similar to the systematic reviews assessing the test ordering literature, neither was able to conduct a meta-analysis due to study outcome heterogeneity.

| Choosing Wisely (An Initiative of the ABIM Foundation) | |
|---|---|
| American College of Obstetricians and Gynecologists | "Don't routinely transfuse stable, asymptomatic hospitalized patients with a haemoglobin level greater than 7-8 grams."[102] |
| American Association of Blood Banks | "Don't transfuse red blood cells for iron deficiency without hemodynamic instability."[103] |
| American Association of Blood Banks | "Don't transfuse more units of blood than absolutely necessary."[103] |
| Critical Care Societies Collaborative – Critical Care | "Don't transfuse red blood cells in hemodynamically stable, non-bleeding ICU patients with a haemoglobin concentration greater than 7g/dL."[86] |
| American Society of Hematology | "Don't transfuse more than the minimum number of red blood cell (RBC) units necessary to relieve symptoms of anemia or to return a patient to a safe haemoglobin range (7 to 8 g/dL in stable, non-cardiac in-patients)."[104] |
| American Society of Anesthesiologists | "Don't administer packed red blood cells (PRBCs) in a young healthy patient without ongoing blood loss and haemoglobin of ≥6 g/dL unless symptomatic or hemodynamically unstable."[105] |
| Society of Hospital Medicine- Adult Hospital Medicine | "Avoid transfusions of red blood cells for arbitrary haemoglobin or hematocrit thresholds and in the absence of symptoms of active coronary disease, heart failure or stroke."[106] |
| **Choosing Wisely Canada** | |
| Canadian Critical Care Society, Canadian Association of Critical Care Nurses, Canadian Society of Respiratory Therapists | "Don't routinely transfuse red blood cells in hemodynamically stable ICU patients with a haemoglobin concentration greater than 70 g/L (a threshold of 80 g/L may be considered for patients undergoing cardiac or orthopaedic surgery and those with active cardiovascular disease)."[87] |
| The Canadian Hematology Society<br>Société Canadienne d'Hématologie | "Don't transfuse patients based solely on an arbitrary haemoglobin threshold."[107] |
| Canadian Society of Palliative Care Physicians | "Don't transfuse red blood cells for arbitrary haemoglobin or hematocrit thresholds in the absence of symptoms, or if no benefit was perceived from previous transfusions."[108] |
| Canadian Society for Transfusion Medicine<br>Société Canadienne de Médecine Transfusionnelle | "Don't transfuse blood if other non-transfusion therapies or observation would be just as effective."[109] |
| Canadian Society for Transfusion Medicine<br>Société Canadienne de Médecine Transfusionnelle | "Don't transfuse more than one Red cell unit at a time when transfusion is required in stable, non-bleeding patients."[109] |
| Canadian Society of Internal Medicine | "Don't transfuse red blood cells for arbitrary hemoglobin or hematocrit thresholds in the absence of symptoms, active coronary disease, heart failure or stroke."[110] |

**Table 2. Examples of Choosing Wisely guidelines related to administration of red blood cells.**

**The Intensive Care Unit Setting**

In intensive care units (ICUs), critically ill patients are rigorously monitored and treated by teams of healthcare providers.[111–113] The extensive use of technology also plays a central role in ICU care.[113] Due to the severity of patient illness, the ICU is a fast-paced and high pressure environment.[114,115] This creates a stressful workplace, not only emotionally (due to the requirement to make difficult decisions quickly), but also as a result of physical and professional factors.[115] Poor lighting, alarms with low sensitivity and similar sounds for different warnings, poorly placed equipment, and a multitude of cords and tubes have been cited as physical factors adding to the stressful environment of the ICU.[111,115] Additionally, the literature frequently mentions the fact that ICUs produce a large amount of patient data (i.e. vital signs and laboratory data), and that this can be difficult for individual providers to process.[111,116] Furthermore, the ICU is a complicated setting in that teams are interdisciplinary, composed of individuals from various professional backgrounds.[112] Teamwork, collaboration, cohesion and communication are thus important in the ICU.[112,115] Finally, long hours, on-call and night shifts are commonly mentioned as contributing to taxing work conditions.[111,115] Working in this stressful environment can have a psychological toll on providers and result in burnout.[111,115,117,118] While the ICU is a highly complex environment,[112] making it a difficult area to institute interventions, it also highlights the importance of conducting research to make improvements in this setting. There is also interest in studying the intensive care setting due to the fact that it is very resource intensive.[53] In fact, it has been estimated that of a hospital's expenditure up to 30% may be ICU-related.[119,120] As seen in a 2015 report put forth by the Ministry of Health and Long-Term Care, hospital laboratories are conducting almost half (47%) of the laboratory tests in Ontario.[17]

Healthcare providers in the intensive care setting frequently order laboratory tests, to monitor patient status.[30] Tests are used heavily in this setting due to the severity of patient illness (or unstable condition) to closely monitor changes in status,[30,121] as well as due to the team-based nature (i.e. tests may be requested by each consultant).[57] Some tests may also be ordered routinely (based on a time interval).[86] While this information can be valuable in treating and monitoring patients, excessive phlebotomy for laboratory testing can lead to harmful side effects.[30] Iatrogenic anemia is one such complication.[30] Though a variety of other factors, such as surgery, bleeding and nutrition, may contribute to the development of anemia, phlebotomy has also been found to play a role.[30,88,122] The literature shows a wide range in daily blood loss values for ICU patients due to phlebotomy. A 2007 review of the literature cited a range between 25-40 mL.[30] Meanwhile, a recent study published by Ullman et al. cited a range of 41.5-377 mL from their review of the literature, and a median volume of 38 mL from their own study of an Australian adult ICU.[122] Anemia appears to increase the risk for unfavourable outcomes among patients with specific diagnoses, such as chronic obstructive pulmonary disease, chronic kidney disease, and cardiovascular disease.[30,88,123]

Nosocomial infections can also be acquired through phlebotomy,[30,124] and are especially dangerous if antimicrobial-resistant.[125] Furthermore, excessive blood draws contributing to anemia may lead to the patient requiring a transfusion,[126] a treatment which can cause further adverse events.[30] Thus, while laboratory testing is an important tool in intensive care, due to the potential risks and downstream consequences that testing may have, it is important that we assess current practice and ensure it aligns with the evidence-base. Similar to other settings, studies have shown inappropriate use of tests in intensive care. As previously mentioned, the study by Melanson et al. showed that almost one third (30%) of the Arterial Blood Gas (ABG) tests ordered in the hospital were potentially unnecessary; ABGs were most frequently ordered in the ICU.[39] A 2017 study completed in an Ontario ICU compared predicted 'essential' blood tests

(determined by the attending physicians the day prior) to actual orders.[127] The authors found that a large proportion of the tests actually completed had not been predicted to be 'essential' (51.3%), and they also saw underutilization of the 'essential' tests (15.6%).[127] Despite the study design's inherent limitation, in that physicians were surveyed the day prior, leaving room for error, the results still suggest that test ordering could be further optimized.[127]

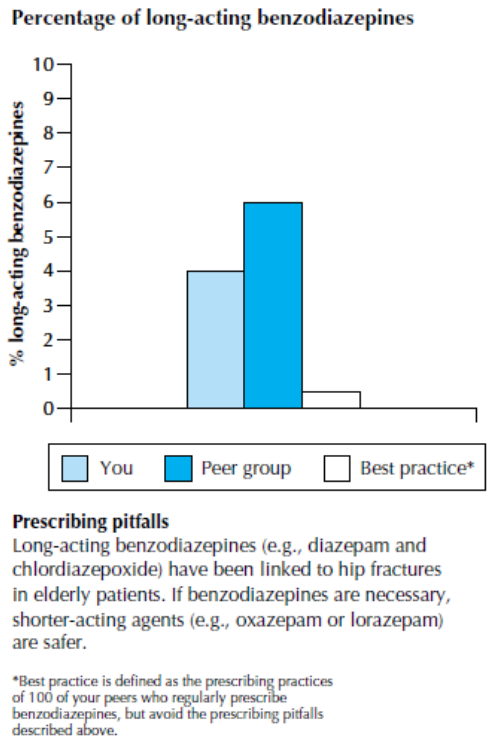
## PROPOSED SOLUTION

### Audit & Feedback

With the overarching goal of improving the quality of care, one focus of Knowledge Translation (KT) and Implementation Science is to investigate approaches used to incorporate evidence-based results into practice.[128,129] There are many different types of KT activities, and one commonly used strategy for improving practice is Audit and Feedback (A&F). The definition of A&F provided by the Effective Practice and Organisation of Care (EPOC) Review Group is: *"A summary of health workers' performance over a specified period of time, given to them in a written, electronic or verbal format. The summary may include recommendations for clinical action"*.[130] Audit and feedback interventions involve the collection of performance data for specific behaviours of an individual and/or group.[131] This data is then subsequently provided back to the individual and/or group to encourage them to change their behaviour, in hopes of improving their performance.[131] As suggested by the EPOC definition, the mode by which A&F interventions are provided can vary widely. For instance, feedback may be provided through a mailed report,[131] e-mail, within a desktop or web platform,[132] during a staff meeting,[131] or at a seminar/presentation.[133] Feedback may provide a comparator (past performance, peers, a standard or goal), and as noted in the EPOC definition, the feedback may provide suggestions as to how the individual or group may go about changing their behaviour to improve their performance.[130,131] **Figure 1A** provides a simple example of a feedback report from a study that intended to reduce inappropriate benzodiazepine prescriptions.[134]

A&F is most effective as an iterative intervention.[135] By delivering feedback more than once, a feedback loop is created (displayed in **Figure 1B**), such that recipients' behaviour is collected and fed back to them, allowing them to assess progress on their performance.[135] It has been suggested that providers' evaluation or "self-audit" of their own performance is imperfect.[131,136(p1095)] Feedback thus aids in highlighting any discrepancies in performance that they may or may not be aware of.[131,137] For instance, a provider may not be aware of the frequency with which they order a test inappropriately. Feedback describing the amount by which they order a test inappropriately should direct a provider's attention to the discrepancy in practice.[131,137] If allowed the opportunity to modify their behaviour, iterative feedback can provide recipients with data on how their efforts have affected performance, and this may continue in a cycle.[135] Adoption of electronic record systems in healthcare provides the capacity to collect and store routine data, supply easily accessible data, and has improved the potential sustainability of iterative feedback interventions.[135,138,139]

**A)**




**B)**

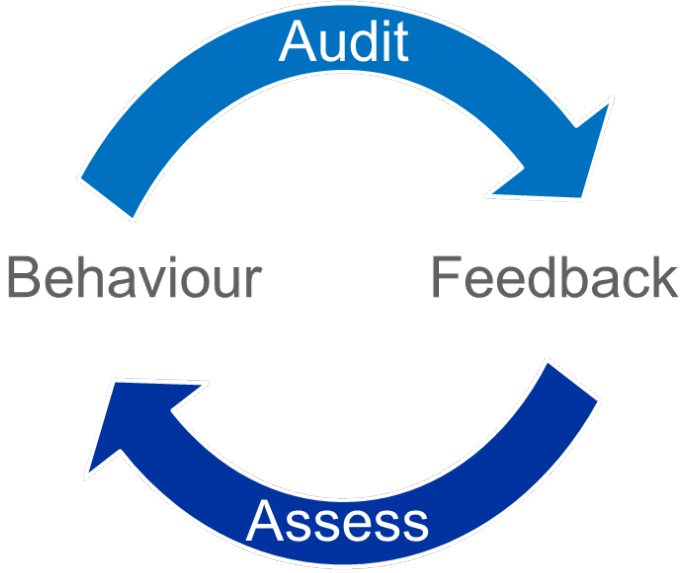



**Figure 1: A)** A sample feedback report provided to physicians during an intervention study aiming to reduce inappropriate benzodiazepine prescriptions.[134] **B)** The Audit & Feedback Loop. Auditing an individual or group's behaviour and subsequently delivering feedback to them allows for the individual/group to assess their performance.[135] Upon the next opportunity, they may modify their behaviour. Continuation of auditing and provision of feedback allows for a cycle of performance improvement.[135]

## Feedback in the ICU

Context may influence whether or not an intervention is successful;[140–142] thus, as with any healthcare setting, unique intricacies of the ICU environment may present challenges to implementing change. Setting specific factors, such as the critical condition of the patient population, multidisciplinary teams, and limited resources and healthcare provider time may impact which strategies are effective for changing practice.[143–145] A&F may be an especially suitable intervention for the ICU setting due to its inherent adaptability and capacity for tailoring to various levels and settings. First, it may be delivered through a variety of modes allowing for integration into work flow. Feedback is also flexible in that it may be given at the individual or team level, which addresses the fact that the ICU provides team-based care. The mass amount of patient data produced[111,116] also makes the ICU a practical setting for the application of A&F. The data on clinician behaviours such as test and transfusion ordering is now usually recorded electronically,[139,146] which allows for implementation of feasible and sustainable A&F interventions, as compared to chart review audits.[13,135,138] Furthermore, A&F would highlight how frequently resources are ordered. Bringing this

information to the attention of the healthcare providers may be helpful, as tests may at times be ordered out of routine or habit.[39,80]

National initiatives using audit to improve practice in the ICU setting in fact currently exist for Canada and the United Kingdom.[147] Similarly, contribution of data on quality indicators to the Dutch National Intensive Care Evaluation (NICE) registry allows provision of feedback to the ICUs.[148,149] While these initiatives provide examples of A&F in the ICU setting, it remains unclear whether this class of intervention has been successful in changing ICU provider test or transfusion ordering behaviour. We were unable to identify any systematic reviews that focus specifically on test and transfusion ordering in the ICU. Furthermore, we note that relatively few studies within Kobewka et al.'s review[82] targeted the intensive care setting. The potential challenges of intervening in the ICU should not preclude us from trying, as studies completed in other settings of care may not be generalizable to the ICU, due to its unique characteristics. Given the volume of tests and transfusions completed in the ICU[30,88] and the potential for improvement in care, there is a need to further understand how we can best support ICU clinicians in appropriate test and transfusion ordering.

**The Theories That Explain How Audit and Feedback Works**

A 2012 Cochrane review assessing the efficacy of healthcare Audit and Feedback (A&F) interventions found that overall, A&F has a positive effect on provider performance, with a large range in effect across included studies.[131] The authors further assessed whether A&F intervention study effect sizes were improving over time, and saw stabilization as of 2003, which suggests researchers have not been improving upon their interventions (despite conducting 32 trials between then and 2009).[150] Trials comparing different types of A&F would help to inform how best to optimize this intervention, and thus there have been calls to limit the conduct of A&F versus no intervention ("usual care") studies which may have limited added value for this already large trial literature.[131,150,151] The question remains: how to optimize this class of intervention?[135]

Theory can provide an important source of information on improving A&F. Studying theories relevant to A&F is imperative to better understanding potential mechanisms of action, and thus determining design and delivery methods to optimize this class of intervention.[150] It has been found that few A&F studies report using theory to develop their interventions.[152] Furthermore, for implementation scientists who do use theory, it appears there is no standard approach to determining which theories are most useful.[153]

While there is a paucity of high quality evidence about which theoretical mechanisms of action should be employed for A&F in healthcare, we can look to those that have been explored in other disciplines.[131,154] As depicted in **Figure**, **control theory** proposes that a system compares performance ("input function" or "perception") to a standard ("reference value").[155] Should performance differ from the standard, the system will aim to close this gap ("reduce the discrepancy") by making adjustments (the "output function" or "behaviour").[155] This behaviour change, as well as other changes outside of the system ("disturbances") may influence the environment, and affect performance. A comparison is made again, and the loop continues. Control theory, when applied to human behaviour, suggests that humans have an internal need to self-regulate.[155] This theory thus proposes that providing an individual with data, showing that their performance is below that of a standard (presence of a discrepancy), should lead to a change in their behaviour to improve their performance (decreasing the discrepancy).[137]

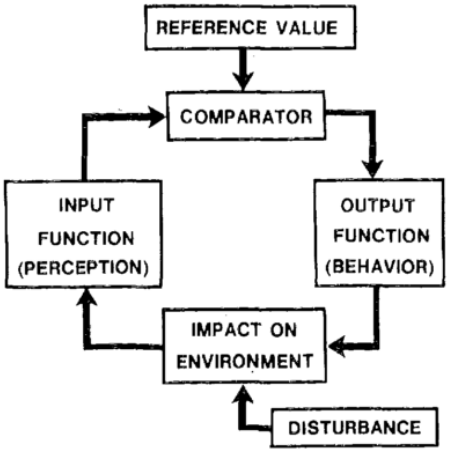



**Figure 2: ″The negative feedback loop″**[155]

**Social psychology** also provides helpful insight into potential mechanisms of action for feedback. **Social cognitive theory** suggests individuals "self-[monitor]" and set inherent goals.[156] Goals motivate individuals to work towards a target, as stipulated by **goal theory**.[137,157] Feedback in turn presents recipients with their behavioural data, allowing them to make a comparison to that of their goal and determine the distance to it (the discrepancy).[137] As it has been suggested that providers conduct imperfect evaluations of their own performance, feedback helps to present how far away recipients actually are from their goal.[131] A further component of social cognitive theory is that self-efficacy affects how an individual reacts to negative performance feedback; high self-efficacy may result in continued motivation to reaching the goal, while low self-efficacy can lead to goal abandonment.[137,156] The application of goal theory to feedback varies from that of control theory (an aim "to reduce the discrepancy"), in that the approach to closing the gap between the performance feedback and goal is an aim to "achieve".[137] Additionally, social cognitive theory takes into account "pre-existing cognitive structures and self-beliefs."[156] Empirical work proposes that setting challenging yet attainable goals leads to greater outcomes as opposed to very simple or non-specific, "do your best" goals.[157,158] This notion can be applied to A&F by providing aspirational comparators, such as the performance of the highest achievers, rather than the overall group average.[159]

**Feedback Intervention Theory** (FIT), draws on constructs from several other theories, including those previously discussed.[137] As seen in **Table 3**, FIT similarly suggests that behaviour is influenced by gaps between performance and a standard.[137] Furthermore, as proposed by control theory, FIT suggests a hierarchal arrangement of goals, wherein attention may be concentrated on one of the various levels.[137,155] FIT adds on to these original theoretical constructs by additionally proposing that feedback redirects and focuses a person's attention on the discrepancy in their performance.[137]

| Feedback Intervention Theory |
| --- |
| a) "Behavior is regulated by comparisons of feedback to goals or standards" |
| b) "Goals or standards are organized hierarchically" |
| c) "Attention is limited and therefore only *feedback-standard gaps* that receive attention actively participate in behavior regulation" |
| d) "Attention is normally directed to a moderate level of the hierarchy" |
| e) "FIs change the locus of attention and therefore affect behavior" |

**Table 3. Feedback Intervention Theory (FIT)[137]**

While these theories provide general potential mechanisms of action for A&F, there are also a variety of modifiable elements of feedback that may alter the effectiveness of the intervention.[135,160] Within healthcare, the application of A&F has been found to be highly heterogeneous.[131,160] There are thus many aspects of A&F that could be adjusted to further improve the effectiveness of these interventions,[135,160] albeit not providing the entire picture of all mechanisms.[135] Brehaut et al. set out to identify modifiable elements hypothesized to optimize A&F interventions, to provide guidance for A&F developers.[135] The suggestions were developed through assessment of the literature and theoretical perspectives, as well as through expert interviews.[135] As seen in **Table 4**, the suggestions cover four domains: "Nature of the desired action", "Nature of the data available for feedback", "Feedback display", and "Delivering the feedback intervention".[135] Although some suggestions remain to be empirically tested in a healthcare setting, and further work is required to elucidate the mechanisms, these suggestions provide a starting point on which to begin more efficient, cumulative research.[135,150] More specifically, these suggestions may be used as testable hypotheses, applied during the design phase of trials or retrospectively to reviews, to compare various types of A&F.[150] It is especially important to pose hypotheses and study the potential mechanisms of action for more effective A&F in unique settings such as the ICU, as findings from other settings may not generalize to a more complex environment.

**Rationale**

We propose that audit and feedback could be a useful intervention to improve performance in the ICU setting; specifically, to improve the appropriateness of laboratory test and transfusion ordering. Although previous systematic reviews have addressed the use of various interventions for the improvement of test ordering across healthcare settings[82] or in primary care,[8,45] and transfusion ordering,[100,101] the degree to which A&F has been applied to these practices in the context of intensive care, and how effective it has been, is unclear. We thus conducted a systematic review including search terms to identify A&F studies completed in the ICU setting. The most recent systematic search pertaining to laboratory testing across settings was conducted in 2013, and in 2003 for transfusion ordering.[82] It is therefore likely that new studies using A&F in the ICU may have been published since this time. Further to this, our inclusion and exclusion criteria differed from that of Kobewka et al.'s, Wilson et al.'s and Tinmouth et al.'s (in terms of intervention scope, setting, and in some cases the target population and comparator). As A&F can be operationalized in a variety of ways, we were also interested in taking a more in-depth look at how the A&F interventions were implemented. We thus set out to conduct a systematic review of the literature, to determine the extent

to which A&F has been employed in the intensive care setting, specifically for the purpose of improving provider ordering of tests and transfusions. We were also looking to qualitatively assess various components of the feedback, such as whether A&F was used as a single mechanism intervention, or in concert with other interventions. If appropriate, we planned to conduct a meta-analysis to assess the effect size of the A&F interventions in changing test or transfusion ordering.

| Nature of the Desired Action | Nature of the Data Available for Feedback |
|---|---|
| 1. Recommend actions that are consistent with established goals and priorities.<br><br>2. Recommend actions that can improve and are under the recipient's control.<br><br>3. Recommend specific actions. | 4. Provide multiple instances of feedback.<br><br>5. Provide feedback as soon as possible and at a frequency informed by the number of new patient cases.<br><br>6. Provide individual rather than general data.<br><br>7. Choose comparators that reinforce desired behaviour change. |
| **Feedback Display** | **Delivering the Feedback Intervention** |
| 8. Closely link the visual display and summary message.<br><br>9. Provide feedback in more than one way.<br><br>10. Minimize extraneous cognitive load for feedback recipients. | 11. Address barriers to feedback use.<br><br>12. Provide short, actionable messages followed by optional detail.<br><br>13. Address credibility of the information.<br><br>14. Prevent defensive reactions to feedback.<br><br>15.Construct feedback through social interaction. |

**Table 4. Brehaut and colleagues' 15 suggestions for improved audit and feedback interventions[135]**

In light of the recent guidance published by Brehaut and colleagues,[135] we were also interested in developing an evaluation tool, so that we could assess how well existing A&F interventions adhere to suggestions for best practice. The guidance by Brehaut et al.[135] frames suggestions generally, and is not inherently usable as a means for assessing existing reports of A&F interventions in its current form, hence the need for a tool. The great variation in how feedback is operationalized in terms of design and delivery has raised the question of which design elements and delivery methods may improve the effectiveness of A&F.[131,135,160] Through application of our tool to A&F intervention studies we aimed to provide insight into which suggestions have been followed in the existing literature. Furthermore, we aimed to discover gaps in the literature by identifying which suggestions have not been followed previously.

## Aim & Objectives

The overall aim of this research was to gain a better understanding of the state of use of A&F in the ICU setting. Additionally, we were interested in developing an evaluation tool that would allow us to assess whether existing A&F interventions were consistent with suggestions for best practice. Our work provides a foundation on which to further assess modifiable elements of A&F interventions, and testable hypotheses as to how to improve their effectiveness. Optimization of these interventions could lead to improved test and transfusion ordering, which could in turn greatly impact both the quality of patient care and healthcare resource use.[30,88,89]

**Objective I:** Conduct a systematic review of the literature to identify audit and feedback intervention trials taking place in the intensive care setting, which have been employed for the purpose of improving provider test or transfusion ordering.

**Objective II:** Develop, pilot and employ an evaluation tool to assess consistency of existing audit and feedback interventions with suggestions for best practice.

## Overview of Methodology & Structure

The research of this thesis has been structured in a manuscript format, which consists of an introductory chapter, two draft manuscripts, and a discussion chapter. The first draft manuscript (Chapter 2) describes a systematic review of the literature, the purpose of which was to assess the extent to which audit and feedback interventions have been used in the intensive care setting, to improve test and transfusion ordering. Identification of these articles allowed for the qualitative analysis of the interventions and a summary of their effectiveness in changing provider behaviour. The draft manuscript provides an overview of the methodology; search strategy development, information sources, screening, data extraction and analysis. The second draft manuscript (Chapter 3) reports on development, piloting and employment of an evaluation tool designed to assess consistency of existing A&F interventions with best practices. Chapter 4 contains a discussion and integration of the results from Chapters 2 and 3, as well as a discussion of the implications and significance of these results.

## Chapter 2 (Draft Component Manuscript 1):

## Audit and Feedback to Improve Laboratory Test and Transfusion Ordering in Critical Care: A Systematic Review

### PREFACE

Chapter two describes a systematic review intended to identify publications assessing the use of Audit and Feedback (A&F) in the critical care setting for the purpose of improving provider test or transfusion ordering, and to summarize intervention characteristics and effectiveness. Our interest in this topic came from the Eastern Ontario Regional Laboratory Association's (EORLA) desire to make changes to their current feedback strategy to providers on laboratory test ordering, and their appeal for guidance.

We chose to focus specifically on the critical care setting as providers in the ICU are frequent users of hospital laboratory services.[1–8] We were also interested in assessing the extent to which A&F has been applied to the related issue of transfusion ordering; excessive blood tests can contribute to iatrogenic anemia, which in turn can increase a patient's risk of requiring a transfusion.[9,10] Transfusion poses further non-trivial risks to the patient, and blood components are a precious medical resource.[10,11] Improving appropriateness of test and transfusion ordering can have beneficial impacts on the quality of patient care, resource supplies, and expenditure.[9–12] As both practices were included in the Choosing Wisely recommendations put forth by the Critical Care Societies Collaborative,[13] we thought it germane to assess and summarize how A&F has been employed to address these issues to date.

Contributions: JCB and JP were responsible for the conception of this project and provided guidance and expertise throughout the entire project. MF drafted the manuscript, and JCB, JP, NM, LM and BH provided critical input and aided in the revision of the manuscript. MF and KC completed title and abstract screening, and KC provided confirmation for inclusion and exclusion of full text articles screened by MF. Data extraction and quality assessment were completed by MF and NM. The guarantor of this review is MF.

Notes: Appendices for documented changes from the protocol, the PRISMA checklist, justification for choice of outcomes and changes in secondary outcomes, the Medline search strategy, additional details on the search strategy and sample, the reporting quality assessment, and study definitions of appropriateness can be found at the end of this chapter. Ethics approval was not required for this study as the systematic review relied solely on published studies.

Publication Citation: Foster M, Presseau J, McCleary N, Carroll K, McIntyre L, Hutton B, Brehaut J. "Audit and Feedback to Improve Laboratory Test and Transfusion Ordering in Critical Care: A Systematic Review." (In preparation).

**Audit and Feedback to Improve Laboratory Test and Transfusion Ordering in Critical Care: A Systematic Review**

Madison Foster, MSc (candidate)[1,2*], Justin Presseau, PhD[1,2,3], Nicola McCleary, PhD,[2] Kelly Carroll, MA,[2] Lauralyn McIntyre,[1,4] Brian Hutton[1,2,5], Jamie Brehaut, PhD[1,2]

[1]School of Epidemiology and Public Health, University of Ottawa, 451 Smyth Road, Ottawa, Ontario, K1H 8M5, Canada

[2]Ottawa Hospital Research Institute, Clinical Epidemiology Program, The Ottawa Hospital, General Campus, 501 Smyth Road, Centre for Practice Changing Research, Box 201B, Ottawa, Ontario, K1H 8L6, Canada

[3]School of Psychology, University of Ottawa, 136 Jean-Jacques Lussier, Vanier Hall, Ottawa, Ontario, K1N 6N5, Canada

[4]Department of Critical Care Medicine, The Ottawa Hospital, General Campus, 501 Smyth Road, Ottawa, Ontario, K1H 8L6, Canada

[5]Ottawa Hospital Research Institute, Knowledge Synthesis Unit, The Ottawa Hospital, General Campus, 501 Smyth Road, Centre for Practice Changing Research, Box 201B, Ottawa, Ontario, K1H 8L6, Canada

**Author E-mail List:**
Madison Foster: madfoster@ohri.ca
Justin Presseau: jpresseau@ohri.ca
Nicola McCleary: nmccleary@ohri.ca
Kelly Carroll: kecarroll@ohri.ca
Lauralyn McIntyre: lmcintyre@ohri.ca
Brian Hutton: bhutton@ohri.ca
Jamie Brehaut: jbrehaut@ohri.ca

***Corresponding Author:** Madison Foster (madfoster@ohri.ca), C/O Kelly Carroll, The Ottawa Hospital, General Campus, 501 Smyth Road, Centre for Practice Changing Research, Box 201B, Ottawa, Ontario, K1H 8L6, Canada

## ABSTRACT

Laboratory tests and transfusions are sometimes ordered inappropriately, particularly in the critical care setting which sees frequent use of both. Audit and Feedback (A&F) is a potentially useful strategy for modifying these healthcare provider behaviours, but its application to improving test and transfusion ordering in critical care is not well understood. We aimed to conduct a systematic review of the literature on A&F interventions for improving test or transfusion ordering in the critical care setting.

Five relevant databases, two registries and the bibliographies of relevant articles were searched. We included critical care studies that assessed the use of A&F targeting healthcare provider behaviours, alone or in combination with other interventions to improve test and transfusion ordering, as compared to historical practice, no intervention, or another healthcare behaviour change intervention. Studies were included only if they reported laboratory test or transfusion orders, or the appropriateness of orders, as outcomes. There were no restrictions based on study design, date of publication or follow-up time. Intervention characteristics and absolute differences in outcomes were summarized. The quality of individual studies was assessed using a modified version of the Effective Practice and Organisation of Care Cochrane Review Group's criteria.

We identified sixteen studies; 13 uncontrolled before-after studies, one randomized controlled trial, one controlled before-after study and one controlled clinical trial (quasi-experimental). These studies described 17 interventions; most interventions (88%) were multifaceted with an A&F component. Feedback was most often provided in a written format only (41%), more than once (53%), and most often only provided data aggregated to the group-level (41%). Most studies saw a change in the hypothesized direction, but not all studies provided statistical analyses to formally test improvement. Overall study quality was low, with studies often lacking a concurrent control group.

Our review suggests that A&F can be effective in the context of critical care, but further research is required to characterize approaches that optimize the effectiveness in this setting alongside more rigorous evaluation methods.

**Registration:** PROSPERO CRD42016051941

**Keywords:** Audit, Feedback, Intensive Care, Critical Care

## BACKGROUND

Laboratory testing is an important and high volume medical resource that facilitates disease detection and monitoring of patient status.[14] However, lab testing is prone to inappropriate use,[14] with estimates suggesting that 20-30% of tests ordered are low-value, i.e. unnecessary, not indicated, or potentially harmful.[14,15] While the tests themselves directly comprise only 4% of overall hospital expenditure, they are thought to be important in up to 70% of subsequent healthcare decisions and their related expenditures, and thus represent an important area for quality improvement.[16,17]

Critical care is one setting where tests are ordered often,[9] and where there is concern of overuse contributing to clinically important poor outcomes in vulnerable patients.[1–6,18–20] Blood loss can contribute to iatrogenic anemia.[9,10] Subsequent red blood cell (RBC) transfusions[9,10] can be associated with non-trivial risks such as transfusion-associated circulatory overload (TACO), transfusion-related acute lung injury (TRALI) and transfusion-related immunomodulation (TRIM).[10,21] The potential risks and downstream consequences associated with laboratory testing and transfusion, in addition to increased expenditure and limited blood resources, all provide motivation to reduce inappropriate use.[9–11,22]

Audit and Feedback (A&F), provision of clinical performance data, represents a potentially low cost and sustainable strategy[23,24] for improvement of test and transfusion practices in the critical care setting. A Cochrane review has demonstrated its widespread effectiveness across a range of clinical behaviours.[25] It is a broadly used intervention, familiar to most healthcare providers. Test and transfusion ordering are increasingly documented electronically, providing accessible data to produce feedback reports at a reasonable cost.[24,26–29] A&F interventions in the context of test ordering in various clinical settings show a 22% relative risk reduction in test volume.[30] To date however, no review has examined the effectiveness of A&F interventions to modify these behaviours in the complex, team-based critical care setting.

## OBJECTIVES

- **Objective I:** To review how A&F interventions have been implemented in the critical care setting to improve the appropriateness of laboratory test and transfusion ordering

- **Objective II:** To summarize the effectiveness of these interventions as compared to usual care or other interventions

## METHODS

### Protocol & Registration

We used the Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols (PRISMA-P checklist)[31] to draft our protocol, which was registered with the International Prospective Register of Systematic Reviews (PROSPERO: CRD42016051941).[32,33] All deviations to the protocol (**Appendix A**) were minor and were implemented prior to the start of data extraction.

### Eligibility Criteria

**INCLUSION:** Studies with the following PICOS characteristics were included in the review (**Appendix B** provides justifications for our choice of outcomes):

*Population:* Studies that targeted healthcare professionals (physicians, nurses, phlebotomists, or respiratory therapists) ordering laboratory tests or blood transfusion components (Red Blood Cells (RBCs), platelets, plasma, or cryoprecipitate) for patients in an intensive care unit (ICU). Articles targeting healthcare professionals ordering laboratory tests or blood transfusion components for patients in a non-ICU setting were excluded.

*Intervention:* Studies assessing Audit and Feedback (A&F) interventions, defined as *"Any summary of clinical performance of health care over a specified period of time. The summary may also have included recommendations for clinical action. The information may have been obtained from medical records, computerised databases, or observations from patients."*[34] We also included multifaceted interventions that included an A&F component (e.g. A&F paired with educational sessions).

*Comparator:* Studies that compared A&F interventions to usual care (no intervention; historical or concurrent), or any other single or multifaceted behavioural intervention that did not involve A&F (e.g. education, incentives, reminders, or systems-based changes).

*Outcomes:* Primary outcomes included the number of laboratory tests or transfusions ordered. Secondary outcomes included: the appropriateness of ordered laboratory tests or transfusions (for example as judged by the clinical context, or as compared to specified guidelines), length of stay (LOS), mortality, infection, and laboratory test or blood product expenditure.

*Study Design:* We included Randomized Controlled Trials (RCTs), Controlled Clinical Trials (CCTs), observational studies (Controlled Before-After studies (CBAs), Interrupted Time Series studies (ITSs), and Uncontrolled Before-After studies (UBAs)).

*Setting*: We assessed studies that implemented interventions in an intensive care setting. All types of hospitals (i.e. academic, community) and ICUs (i.e. surgical, medical, pediatric, neonatal, etc.) were included. Studies implementing interventions across multiple settings (i.e. hospital-wide) were only included if ICU-specific data was reported for the primary outcome.

**EXCLUSION*:*** No time restrictions or year or language filters were used. We excluded conference abstracts, commentaries and letters to the editor, as well as studies not published in English to maintain feasibility. Previous literature suggests such language restrictions do not greatly affect review conclusions.[35] Studies implementing interventions across multiple settings, but not reporting ICU-specific data for the primary outcome were excluded.

**Search Strategy Development and Information Sources**

Our Medline (1946) search strategy (**Appendix C**) was developed with help from an information specialist. The strategy was then peer reviewed by a second, independent information specialist, as recommended by the Centre for Reviews and Dissemination [36–38] (**Appendix D** provides further details). Medical Subject Headings (MeSH terms) and title and abstract terms ('.tw') were chosen for the general categories 'Laboratory Tests', 'Transfusions', 'Intensive Care', and 'Audit and Feedback'. This template strategy was translated for use in the remaining databases, Embase (1947), EBM Reviews- Cochrane Central Register of Controlled Trials, CINAHL (1981), and PsycINFO (1806). These searches were run on October 28th, 2016, starting from database conception. The trial registries 'ClincalTrials.gov' and International Standard Registered Clinical/soCial sTudy Number (ISRCTN) were additionally searched on December 23rd, 2016 to identify any relevant ongoing trials, using the search terms 'intensive care' and 'feedback'. The bibliographies of included articles and relevant systematic reviews[25,30,39–41] were also hand searched to identify any further articles meeting the inclusion criteria.

**Study Records**

**Data Management:** Citations retrieved from the search were imported into the reference manager software program *Mendeley Desktop 1.17.12* (Mendeley Ltd., London, UK) for de-duplication, then imported into *Covidence,*[42] a web-based platform for screening.

**Selection Process:** The titles and abstracts of unique citations identified from electronic database searches were screened by two independent reviewers (MF and KC), and registry citations were screened by one reviewer (MF). Conflicts were resolved through discussion or reference to a third independent reviewer (JCB, JP). Full text articles were screened by one reviewer (MF), and justifications for inclusion or exclusion were confirmed by a second member of the research team (KC).

**Data Collection Process:** Data was extracted by two independent reviewers (MF and NM) using a standardized data extraction form implemented in Microsoft Excel 2011. One reviewer piloted the form on the first five articles and only minor refinements were required. Conflicts between data extraction forms were identified by one reviewer (MF), and consensus was reached between reviewers through discussion. If reviewers were not able to come to an agreement, a third reviewer (JCB, JP) was consulted to reach consensus.

**Data Extracted**

We extracted several feedback intervention details based on characteristics described in the most recent Cochrane review[25] (format type, interval between reports (frequency)) and recently published guidance for the optimization of A&F[24] (type of data, specificity of data, number of reports, mode of delivery). We also extracted details about other intervention components (if applicable), study design, type of control (i.e. historical, concurrent), type of ICU, type of patient (if applicable), type of laboratory test or blood

component targeted, study participants (i.e. healthcare provider type), number of participants, follow-up time points, study country, funding, year of publication, and each study's definition for an appropriate test or transfusion (if applicable).

### Risk of Bias

Two independent reviewers (NM and MF) assessed the methodological quality of studies using a modified version of the EPOC Review Group's quality criteria,[34] used by Kobewka et al.[30] (**Appendix E**). At the present time, there is not enough evidence to pick an appropriate cut-off to differentiate between high and low-quality studies. Furthermore, Cochrane recommends researchers avoid a scaled approach, and instead advocates for complete reporting of quality criteria.[43] We have thus presented results for each criteria item, and have not excluded any studies from our qualitative review. Reviewers were not blinded during data extraction or quality assessment. Cohen's Kappa[44] was determined to evaluate inter-rater reliability.

### Data Synthesis and Analysis

Because of high heterogeneity in study designs, methods, outcomes and variable reporting formats, we deemed meta-analysis to be inappropriate. Tables of study characteristics, intervention characteristics, and intervention effects were prepared to describe the set of included studies; absolute differences have been calculated for study outcomes. Our results have otherwise been reported as per the PRISMA guidelines, and a PRISMA checklist has been completed to document the inclusion of all critical elements of this review (**Appendix F**).[45]

## RESULTS

### Study Selection

**Figure 1** describes our screening process. Starting from 2,364 citations (extracted from electronic databases on October 28th, 2016 and registries December 23rd, 2016), after removal of duplicates and two rounds of screening, 16 unique studies (described within a set of 17 publications)[21,46–61] were identified for inclusion (**Note**: Merlani et al.[49] and Diby et al.[50] are publications assessing different aspects of the same study).

### Study Characteristics

**Table 1** describes characteristics of the included studies (n=16). Ten of the 16 studies (63%) included transfusion outcomes,[21,53,55,56,58–63] eight studies (50%) included test ordering outcomes,[47–52,54,55,61] while two studies included both.[55,61] Of the studies including test ordering outcomes, six aimed to reduce overall test ordering[47,49–52,54,61] and four aimed to improve the appropriateness of tests; one aimed to increase compliance with a sepsis bundle,[55] one aimed to improve compliance with arterial blood gas guidelines (an algorithm),[49,50] one aimed to improve compliance with standards for practice in the ICU,[48] and one aimed to reduce 'unordered' tests (tests with no written order).[51] Of the studies including transfusion ordering outcomes, three aimed to reduce the overall number of transfusions[56,61,63] while seven aimed to improve the appropriateness of transfusions.[21,53,55,58–60,62] Of those assessing appropriateness, two aimed to improve compliance with a bundle;[21,55] three assessed appropriateness as per guidelines or a protocol involving a transfusion 'trigger' (defined level(s) at which to transfuse) and sometimes other patient factors;[58–60] and

one study assessed appropriateness as per guidelines but included an additional category based on clinical context, "inconsistent with guidelines yet appropriate for ICU".[62] The remaining study used a combination of transfusion 'triggers' and an audit of clinical factors, however, several transfusion triggers were noted in the publication and it was not entirely clear which were used to specify appropriateness.[53] Further details on the criteria used to assess appropriateness can be found in **Appendix G**.

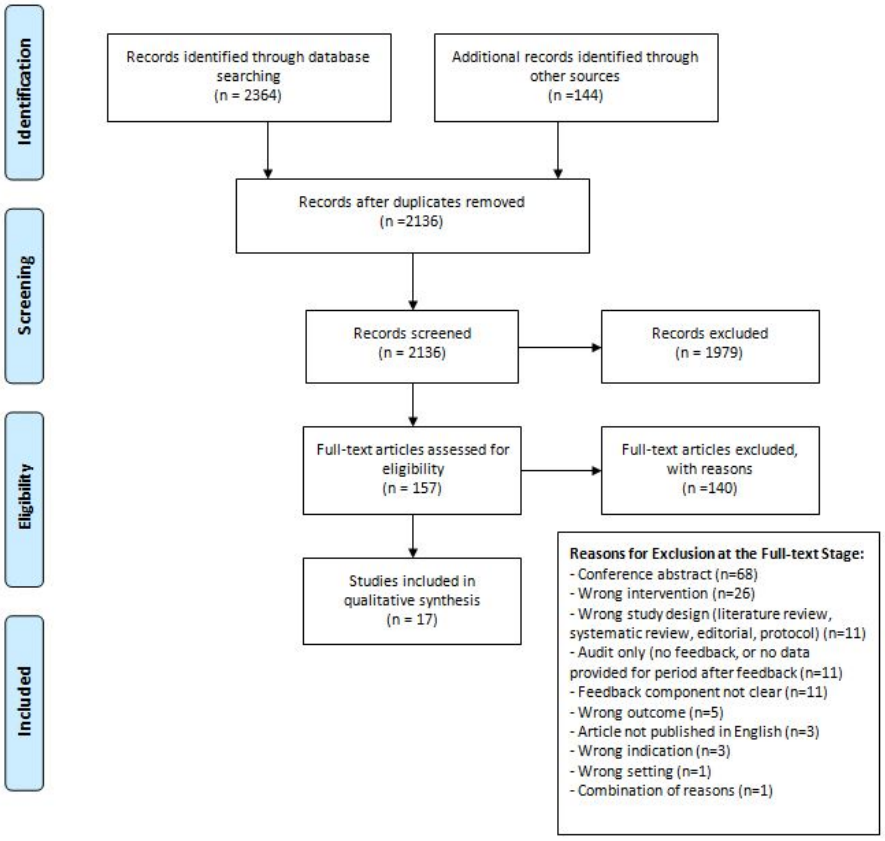



**Figure 1.** PRISMA Flow Diagram outlining the selection of citations for inclusion in the qualitative analysis.[45] A list of the excluded full text articles, sorted by reason for exclusion, can be found in **Appendix H**.

Most studies (81%) used an uncontrolled before-after design.[47,49–55,58–63] Only one randomized controlled trial (RCT),[48] one controlled clinical trial (CCT) with a quasi-experimental comparative design,[21] and one controlled before-after design were identified.[56] Most (56%) were conducted in America,[48,51,55,56,58–61,63] and two (13%) were conducted in Canada.[47,62] Half (50%) of the included studies did not report their source of funding;[21,47,51,52,54,59,60,63] four studies (25%) reported government grant funding.[48,55,58,61] Most studies (56%) were conducted in a single ICU,[47,49–51,53,55,56,58,59,62] while four studies (25%) were conducted in multiple ICUs at a single centre.[21,54,60,63] Most (69%) took place at academic hospitals.[21,49–56,59,61,62] The year of publication ranged from 1988 to 2016, and study duration ranged from 25 weeks to 4 years.

**Assessment of Study Quality**

A Cohen's Kappa of 0.67 was computed for inter-rater reliability (**Appendix I**), representing "substantial agreement" as per Landis and Koch, but just meeting the cut-off for "suggesting that … conclusions tentatively be made" as per Krippendorff.[44] As such, reviewers discussed all disagreements to reach a consensus. **Appendix E** describes the quality of included studies (n=16). Overall quality of the studies was judged to be poor; 94% of studies[21,47,49–56,58–63] scored 4 or lower on the 8-9 criteria (risk of contamination was often not applicable). Most studies reported similar providers between groups (94%),[21,47–52,54–56,58–63] and used an objective primary outcome measure or blinded for the primary outcome assessment (88%; 13 studies[21,47,48,51–56,58,60,61,63] and one study[62] respectively). However, most studies lacked a concurrent control group (88%)[47,49–56,58–63], did not use time series analysis (100%), provided an insufficient amount of detail to allow for replication (100%), and did not report the number of tests per patient (56%).[21,48,51,53,54,58,60,62,63]

**Range of A&F Interventions**

There was a range of A&F interventions (n=17) used in the 16 included studies.[A] As shown in **Table 2**, most interventions were multifaceted (88%),[21,47–51,53–56,59–63] including A&F and one or more additional components (i.e. education, guidelines, opinion leaders, financial incentives, checklists or administrative interventions). Seven interventions (41%) reported providing feedback in a written format only,[21,49,50,52,54,56,60,61] four (24%) provided at least verbal feedback,[47,53,58,59] and three (18%) reported providing both written and verbal feedback.[21,48,51] Four interventions (24%) provided feedback only once,[48,53,54,63] nine (53%) provided feedback more than once,[21,49,50,52,55,56,58,60,61] and in four cases it was unclear or the feedback was provided variably (24%).[47,51,59,62] Where reported, feedback was provided daily in one study,[52] weekly in two (12%),[55,58] monthly in three (18%),[21,49,50,61] and at various instances in four (24%).[21,56,59,60] Feedback most often provided data on group performance only, in seven of the interventions (41%);[21,48–50,54,61–63] three interventions provided both group and individual feedback (18%),[21,58,60] and one intervention only clearly reported providing individual feedback (unclear if group data was provided).[47] Feedback recipients were most commonly multiple groups of healthcare providers (HCPs) (29%),[48–51,59,62] or physicians only (24%).[52,53,58,60]

[A] One study compared two different types of feedback; proportions for the feedback intervention characteristics have therefore been calculated using a denominator of 17.

**Table 1. Summary of Study Characteristics (n= 16 studies)[a]**

| Clinical Behaviour Targeted | Number of Studies (%) |
|---|---|
| **Laboratory Test Ordering** | **8 (50.0%)** |
| Multiple, miscellaneous or unspecified tests | 3 (18.8%) |
| ABG | 2 (6.3%) |
| Lactate & blood cultures | 1 (6.3%) |
| Superficial cultures | 1 (6.3%) |
| Blood work | 1 (6.3%) |
| **Transfusion Ordering** | **10 (62.5%)** |
| RBCs | 6 (37.5%) |
| FP/FFP | 3 (18.8%) |
| All (RBC, FFP, platelets, cryoprecipitate) | 1 (6.3%) |
| **Study Design** | |
| Uncontrolled Before After | 13 (81.3%) |
| Cluster Randomized Controlled Trial | 1 (6.3%) |
| Controlled Clinical Trial[b] | 1 (6.3%) |
| Controlled Before After | 1 (6.3%) |
| **Data Collection** | |
| Prospective | 5 (31.3%) |
| Retrospective | 4 (25.0%) |
| Mixed | 3 (18.8%) |
| Unclear | 4 (25%) |
| **Funding** | |
| Not Reported | 8 (50%) |
| Government Grant | 4 (25%) |
| Institutional[c] and Non-profit Grants | 2 (12.5%) |
| Institutional[c] | 1 (6.3%) |
| No Funding | 1 (6.3%) |

| Country | Number of Studies (%) |
|---|---|
| USA | 9 (56.3%) |
| Canada | 2 (12.5%) |
| Finland | 1 (6.3%) |
| Germany | 1 (6.3%) |
| Israel | 1 (6.3%) |
| The Netherlands | 1 (6.3%) |
| Switzerland | 1 (6.3%) |
| **Number of Sites** | |
| Single centre, single ICU study | 9 (56.3%) |
| Single centre, multi-ICU study | 4 (25.0%) |
| Multicentre study | 2 (12.5%) |
| Single centre, # ICUs unclear | 1 (6.3%) |
| **Hospital Type** | |
| Teaching | 11 (68.8%) |
| Not Reported | 3 (18.8%) |
| Other: Veteran's Administration Medical Centre | 1 (6.3%) |
| **ICU Type** | |
| Surgical | 2 (12.5%) |
| Neonatal | 2 (12.5%) |
| Cardiac Surgery | 3 (18.8%) |
| Neurosurgical | 1 (6.3%) |
| Medical | 1 (6.3%) |
| Mixed Patient Population | 2 (12.5%) |
| Multiple Types of ICUs | 3 (18.8%) |
| Not Specified | 2 (12.5%) |

a) Proportions were calculated for the 16 studies, rather than the 17 publications. Totals may be slightly greater or less than 100% due to rounding; b) The control group was another type of A&F; c) 'Institution' refers to both hospitals and academic institutions. *Abbreviations:* ABG=Arterial Blood Gas; FP= Frozen Plasma; FFP= Fresh Frozen Plasma; ICU= Intensive Care Unit; RBC= Red Blood Cell

**Table 2. Description of Feedback**

| Study & Country | Format | Delivery | Data Specificity | Data Included in Feedback | Instances of Feedback Provided | Frequency/ Interval | Other Intervention Components | Feedback Recipients |
|---|---|---|---|---|---|---|---|---|
| Solomon 1988 USA[46] | NR | Unclear ("reported") | Group; unclear if individual | "…it was determined that 43% of the transfusions were unjustified. The results of this audit were reported" | 1 | N/A | •Education<br>•Guidelines<br>•Administrative (new request form, policy) | Unclear ("leaders of the surgical and medical attending staff") |
| Paes 1994 Canada[47] | Verbal, unclear if written component (NR) | "continuing medical education rounds" & unclear ("direct encounters", "direct, immediate feedback") | Individual; unclear if group | Unclear; "information obtained from this audit", "direct positive and negative performance feedback", "direct, immediate feedback about the policy" | Unclear | Unclear | •Education<br>•Administrative (protocol/policy)<br>•Opinion leader<br>•"Barriers" (ordering this test required a justification and conversation with the laboratory consultant; colleagues were encouraged to challenge inappropriate orders) | Unclear (all staff types mentioned) |
| Hendryx 1998 USA[48] | Written and verbal | Face-to-face feedback meeting, reports | Group, unclear if individual | **Face-to-face:** Unclear; "reviewed the findings, and offered concrete, practical suggestions for improvement"<br>**Reports:** "percentage of processes successfully done, number of patients treated and their length of stay and discharge status, and occurrence of nosocomial events" | 1 | N/A | •Education (newsletter, seminars)<br>•Telephone consultation service | All providers |
| Merlani 2001 & Diby 2005 Switzerland[49,50] | Written | "time series charts, displayed on walls, and published in the unit information bulletin" | Group | Adherence, ABGs per patient day | 20 | Monthly | •Education<br>•Guidelines (Algorithm) | Physicians, physicians in-training, nurses, nurses in-training |
| Beland 2003 USA[51] | Written and verbal | "in-service training sessions", handouts, posters | NR | "findings of the audit"; "laboratory charges"; "rate of unordered tests" | Unclear | NR | • Guidelines<br>• Opinion Leader<br>• Discussion on reducing hospital costs to save nurse positions<br>• "new processes" | Nurses & unclear ("medical staff", "healthcare staff members") |

| Study & Country | Format | Delivery | Specificity | Data Included in Feedback | Instances of Feedback Provided | Frequency/ Interval | Other Intervention Components | Feedback Recipients |
|---|---|---|---|---|---|---|---|---|
| Wisser 2003 Germany[52] | Written | "sent together with the laboratory results" | NR (patient-level data) | "cumulative diagnostic blood loss" | Unclear (multiple) | Daily | Unclear | Physicians |
| Petäjä 2004 Finland[53] | Verbal | "presented and discussed at a staff meeting" | NR | Unclear; "Results of PI and PII", "justifications of and goals for change" | 1 | N/A | • Administrative (on-line auditing system) | Physicians, physicians in-training |
| Calderon-Margalit 2005 Israel[54] | Written | Letter; "sent to the wards and reviewed with senior medical staff" | Group; unclear if individual (NR) | "overall institutional reduction in requests for all clinical biochemistry tests, as well as data on their specific ward's reduction in testing" | 1 | N/A | • Education<br>• Administrative (policy) | Unclear ("heads of all the wards", "senior medical staff") |
| Schramm 2011 USA[55] | NR | NR | NR | "compliance with the sepsis resuscitation bundle" | ~84 | Weekly | •Education & Order Set (also at baseline)<br>•Sepsis Response Team activation | Unclear ("healthcare providers") |
| Masud 2011 USA[56] | Written, unclear if verbal component (NR) | Letters & unclear ("sharing data") | NR | "number of units transfused", "transfusions…outside of the recommended guidelines", "outcomes" | Unclear (multiple) | **Feedback:** Monthly & quarterly<br>**Educational Letter:** Depends on recipient | • Education<br>• Formation of transfusion committees | Unclear, Educational letters: Physicians |
| Arnold 2011 Canada[57] | NR | NR | Group; unclear if individual | "general rates of inappropriate FP use"; "rates of inappropriate FP use after each of their weeks on service" | Unclear | NR | •Education<br>•Administrative (request form required indication, prompt if not completed) | Physicians, nurses |
| Beaty 2013 USA[58] | Verbal, unclear if written component (NR) | "publicly at a weekly cardiac surgical division meeting" | Group & individual | Protocol adherence (exact details unclear) | 17 | Weekly | • Administrative (Protocol/ restriction of who could order)<br>*Note: Only A&F alone intervention | Physicians, physicians in-training |

| Study & Country | Format | Delivery | Specificity | Data Included in Feedback | Instances of Feedback Provided | Frequency/ Interval | Other Intervention Components | Feedback Recipients |
|---|---|---|---|---|---|---|---|---|
| Gutsche 2013 USA[59] | Verbal, unclear if written component (NR) | "feedback interviews and re-education" | NR | Unclear | Unclear; "in the case of guideline noncompliance" | Variable (depends on recipient) | • Guideline<br>• Education<br>• Administrative (closing of the unit) | Physicians, physicians in-training, nurses, other (physician assistants) |
| Yeh 2015 USA[60] | Written | Email & reports | Individual & group | Details of transfusion events; summaries of transfusion activity | **Individual:** Unclear (variable, depends on recipient; 16 were sent in total)<br><br>**Group:** 6 | **Individual:** Unclear (depends on recipient; "within 72 hours of transfusion")<br><br>**Group:** Monthly | • Education | Physicians, physicians in-training |
| Murphy 2016 USA[61] | Written | Reports | Group ("Unit-level") | "Change in utilization" (ABGs & RBCs) | 12 | Monthly | • Education<br>• Opinion Leaders<br>• Financial Incentives | Unclear ("ICUs") |
| Borgert 2016 Netherlands[21] | **Arm 1:** Written<br><br>**Arm 2:** Written and verbal | **Arm 1:** Emailed report, posters<br><br>**Arm 2:** Emailed report, posters, "face-to-face contact" (report) | **Arm 1:** Group<br><br>**Arm 2:** Group and individual | **Arm 1:** "Compliance levels per team"<br><br>**Arm 2:** "Compliance levels per team"; "Compliance levels of the complete bundle and compliance per element" | **Group:** 4<br><br>**Individual:** Unclear (for every transfusion ordered); overall= 40 "face-to-face contact" and 84 e-mails | **Group:** Monthly<br><br>**Individual:** Varied but "within 72 hours after each RBC transfusion" | • Education<br>• Bundle/Checklist | Nurses |

***Abbreviations:*** NR= Not Reported.

**Table 3a. Summary of laboratory test and transfusion ordering and appropriateness**

| Study | Design | Change sought in primary outcome | Absolute Δ Time 1- Time 2 | Reported p-value | Absolute Δ Time 1- Time 3 | Reported p-value |
|---|---|---|---|---|---|---|
| Solomon 1988 | UBA | Decrease in use of FFP /month for the SICU and MICU (units not reported) | -79 | NR | | |
| Paes 1994 | UBA | Decrease number of superficial cultures per patient | -1.72 | NR | | |
| Hendryx 1998 | RCT | Improve$^{\phi}$ process compliance for 'lab work' (%) | Treatment =+17%<br>Control = -7% | <0.0001 | | |
| Merlani 2001 & Diby 2005 | UBA | Decrease median # ABGs (per patient day) | -1.7 | <0.001 | -3.4 | <0.001 |
| | | Improve$^{\phi}$ average adherence to guideline (%) | +15% | <0.0001 | +27% | <0.0001 |
| Beland 2003 | UBA | Decrease & Improve$^{\Sigma}$: | | | | |
| | | Total # of 'blood work' tests per patient | +144 | NR | +214 | NR |
| | | Unordered 'blood work' tests per patient | +15 | NR | +6 | NR |
| Wisser 2003 | UBA | Decrease number of tests (various) per patient | -8 | NR | | |
| Petäjä 2004[a] | UBA | Improve$^{\otimes}$: | | | | |
| | | FFP (transfusions per patient) | -0.74 | NR* | +0.03 | NR* |
| | | Platelets (units per patient) | -0.05 | NR* | -0.47 | NR* |
| | | | | | **Δ Time 2 – Time3** | |
| | | Distribution of pre-transfusion platelet counts | | | Presented graphically | 0.452 |
| | | Distribution of pre-transfusion prothrombin time values | | | Presented graphically | <0.001 |
| | | Audited + Prothrombin time value > 39% | | | -17% | <0.0001 |
| | | Audited + Prothrombin time value > 49% | | | -9.7% | <0.0001 |
| | | All transfusions + Prothrombin time value > 49% | | | -6.9% | <0.0001 |
| Calderon-Margalit 2005 | UBA | Decrease clinical Biochemistry Test orders per 100 hospital days for ICUs (mean volume per 4-month period) (**TARGET**) | -613.1 (-5579) | 0.009 | | |
| | | Hematology Test orders per 100 hospital days for ICUs (mean volume per 4-month period) (**Not targeted**) | +34.2 (+707) | NS (NR) | | |

| Study | Design | Change in Primary Outcome Sought | Absolute Δ Time 1- Time 2 | Reported p-value | Absolute Δ Time 1- Time 3 | Reported p-value |
|---|---|---|---|---|---|---|
| Schramm 2011[b] | UBA | Improve[Δ] # of compliant episodes: | | | | |
| | | Lactate measured (%) | +15.8% | <0.001 | +21.6% | <0.001 |
| | | Blood cultures before antibiotics (%) | +5.3% | <0.001 | +10.0% | <0.001 |
| | | Appropriate RBC transfusion (%) | +3.8% | 0.397 | +3.1 | 0.397 |
| Masud 2011[c] | CBA | Decrease proportion of CABG patients receiving transfusion (total blood product use) | Δ 2006-2007= -9.9%<br>Δ 2007-2008= -6% | NR<br>NR | Δ 2006-2008= -15.9% | <0.005 |
| | | Decrease Volume (Units) for CVICU Patients | | | | |
| | | All products | Δ 2007-2008= -2,288 | NR | | |
| | | Red Cells | Δ 2007-2008= -870 | NR | | |
| | | Platelets (concentration) | Δ 2007-2008= -566 | NR | | |
| | | Plateletpheresis | Δ 2007-2008= -53 | NR | | |
| | | Fresh Frozen Plasma | Δ 2007-2008= -660 | NR | | |
| | | Cryoprecipitate | Δ 2007-2008= -139 | NR | | |
| Arnold 2011 | UBA | Improve[ϑ]: | | | | |
| | | Number of frozen plasma (FP) requests per patient | -0.36 | NR* | -0.19 | NR* |
| | | Inappropriate FP requests | T2 reported graphically | NR | -14% | 0.09 |
| | | FP requests consistent with guidelines | T2 reported graphically | NR | -1% | 0.86 |
| | | FP requests inconsistent with guidelines yet appropriate for the ICU | T2 reported graphically | NR | +15% | 0.04 |
| Beaty 2013[d] | UBA | Improve[ϕ] Odds Ratio (Risk of RBC transfusion above a Hgb threshold of 8gm/dL determined by univariate logistic regression) | T2 OR= 0.52 | 0.003 | T3 OR= 0.37 | < 0.001 |
| | | Improve proportion of RBC units with a Hgb threshold of ≥ 8gm/dL | Reported graphically (decrease) | <0.001 | Reported graphically (decrease) | <0.001 |

| Study | Design | Change in Primary Outcome Sought | Absolute Δ Time 1- Time 2 | Reported p-value | Absolute Δ Time 1- Time 3 | Reported p-value |
|---|---|---|---|---|---|---|
| Gutsche 2013 | UBA | Improve appropriateness[ϕ]; assessed proportion of patients receiving unnecessary RBC transfusion (%) | -6.6% | 0.016 | | |
| Yeh 2015 | UBA | Improve[ϕ]: | | | | |
| | | RBC Transfusions (U per event) | -0.4 | NR* | Unclear; -0.53 to -0.73 | NR* |
| | | Hgb trigger >8.0 g/dL | -23%, | <0.001 | -8%, | 0.44 |
| | | Over-transfusion rate (post-transfusion Hgb >10 g/dL) | -8%, | 0.004 | -5%, | 0.50 |
| | | Mean pre-transfusion Hgb trigger (g/dL) | -0.5 | <0.001 | -0.3 | 0.068 |
| Murphy 2016 | UBA | Decrease mean ABG orders per encounter | -1.6 | < 0.05 | -1.6 | <0.05 |
| | | Decrease mean RBC unit orders per encounter | -0.1 | <0.05 | -0.1 | <0.05 |

*Abbreviations:* ABGs= Arterial blood gases, CBC= Complete Blood Count, FFP= Fresh Frozen Plasma, LFT= Liver function tests, MICU= Medical Intensive Care Unit, NR= Not Reported, NS= Not Significant, PT/PPT= prothrombin time/partial thromboplastin time, RBCs= Red blood cells, SICU= Surgical Intensive Care Unit, T1= Time 1 (baseline), T2= Time 2 (implementation), T3= Time 3 (follow-up); **a)** Study reported appropriateness data for T2 and T3 combined (not shown), T2: After audit system activated, T3: Audit system + Post-Feedback; **b)** P-values for comparison of all three periods; **c)** 2005-2006= Standard care, 2007= Standard care/ Education and A&F ("Educational initiative began in late 2007"), 2008= Education and A&F "fully implemented"; **d)** T2: Weekly, group feedback; T3: Weekly, individual feedback as a group. *Measures of Appropriateness:* Δ = Bundle; ϕ= Guidelines/Algorithm/Protocol/Standards for Practice; Σ= aimed to reduce 'unordered tests' (tests with no written order); ϑ= Guidelines + Clinical Context; ⊗ = Combination of transfusion triggers and audit of patient factors (specifics unclear). *Not the aim of the study. **Note: p-values are those reported in studies.**

**Table 3b. Summary of transfusion ordering and appropriateness for comparative A&F study**

| Study | Design | Change in Primary Outcome Sought | Absolute Δ Arm 2- Arm 1 | Reported p-value |
|---|---|---|---|---|
| Borgert 2016[e] | CCT | **Improve[Δ]:** | | |
| | | **Number of transfused RBCs (per patient)** | | |
| | | Implementation | -36 (-0.6) | 0.0025 |
| | | Post-implementation | -54 (-0.6) | <0.001* |
| | | **Transfusion bundle compliance (%)** | | |
| | | Implementation | +31% | <0.001* |
| | | Post-implementation | +36% | <0.001 |

**e)** Comparison of monthly, group A&F (Arm 1) to monthly, group A&F plus timely individual A&F (Arm 2). *Measures of Appropriateness:* Δ = Bundle. **Note: p-values are those reported in studies, with one exception *article reported a p <0.000, this has been corrected to <0.001.**

**Table 4a. Secondary Outcomes**

| Study | Outcome | Absolute Δ Time 1- Time 2 | Reported p-value | Absolute Δ Time 1- Time 3 | Reported p-value |
|---|---|---|---|---|---|
| Paes 1994 | Costs for superficial cultures | -$21.49/patient | NR | | |
| Hendryx 1998 | Mean Total LOS (days) | Treatment= -3.2<br>Control = -0.6 | NS | | |
| | Mean ICU LOS (days) | Treatment= -2.1<br>Control = -0.3 | NS | | |
| | Mortality Rate (*unclear if ICU or hospital) | Treatment = +0.02<br>Control = -0.14 | NS | | |
| | Mean Infectious Nosocomial Events per 100 ICU Days | Treatment = +0.1<br>Control= -1.0 | NS | | |
| Merlani 2001 & Diby 2005 | Unit population, Mean stay (days) | -0.3 | 0.26 | -0.3 | 0.26 |
| | Unit population, Mortality (%) | +0.1% | 0.80 | -0.5% | 0.80 |
| | Savings (per patient day) | Pilot period: SFr 34.8 (or £14.15) | | Consolidation period: SFr 68.4 (or £27.81) | |
| Beland 2003[a] | Average LOS (days) | +4.9 | NR | +9.4 | NR |
| | Total charge for unordered tests (average cost per patient per day in ICU) | +$4,564.80 (+$42.32) | NR | +$3,246.75 (-$21.91) | NR |
| Petäjä 2004[b] | LOS (NR, Calculated) Days of Care/All Admissions (days/patient) | +0.1 | NR | -1.2 | NR |
| Schramm 2011 | Median ICU LOS (Days) | 0 | 0.010 | 0 | 0.010 |
| | Hospital mortality (%) | -1.6% | 0.029 | -8.3% | 0.029 |

| Study | Outcome | Absolute Δ Time 1- Time 2 | Reported p-value | Absolute Δ Time 1- Time 3 | Reported p-value |
|---|---|---|---|---|---|
| Masud 2011[c] | Observed: Expected operative mortality index | Δ 2006-2007= -0.07 | NR | Δ 2006-2008= -0.12 | NR |
| | for isolated CABG | Δ 2007-2008= -0.05 | NR | | |
| | Average LOS for CVICU patients (Days) | Δ 2007-2008= -0.21 | NR | | |
| | Estimated expense for all blood products | Δ 2007-2008 = -$928,125 | NR | | |
| | CRBSI | Δ 2007-2008= +0.3 | NR | | |
| | VAP Incidence | Δ 2006-2007=0 | NR | Δ 2006-2008=-1 | NR |
| | | Δ 2007-2008=-1 | NR | | |
| | Surgical Site Infection Rate | | | | |
| | CBGB risk 0,1 | Δ 2006-2007= -0.66 | NR | Δ 2006-2008= -0.2 | NR |
| | | Δ 2007-2008= 0.46 | NR | | |
| | CBGB risk 2 | Δ 2006-2007= -2.49 | NR | Δ 2006-2008= -3.43 | NR |
| | | Δ 2007-2008= -0.94 | NR | | |
| | CBGC risk 1 | Δ 2006-2007= -5.13 | NR | Δ 2006-2008= -5.13 | NR |
| | | Δ 2007-2008= 0 | NR | | |
| | CBGC risk 2,3 | Δ 2006-2007= 0 | NR | Δ 2006-2008= 0 | NR |
| | | Δ2007-2008= 0 | NR | | |
| Arnold 2011 | ICU Mortality (%) | -4% | 0.76 | -9% | 0.76 |
| | Hospital Mortality (%) | -4% | 0.90 | +2% | 0.90 |
| Beaty 2013[d] | CSICU LOS (Days) | | | | |
| | Non-transfused (n=368) | -0.1 | 0.21 | -0.5 | 0.21 |
| | Transfused (n=144) | -1.4 | 0.22 | 0 | 0.22 |
| | Total Hospital LOS | | | | |
| | Non-transfused (n=368) | 0 | 0.11 | -1 | 0.11 |
| | Transfused (n=144) | -3 | 0.36 | +1 | 0.36 |
| | Observed In-Hospital Mortality | -4.8%, | 0.02 | -5.5% | 0.02 |

| Study | Outcome | Absolute Δ Time 1- Time 2 | Reported p-value | Absolute Δ Time 1- Time 3 | Reported p-value |
|---|---|---|---|---|---|
| Gutsche 2013 | Mean ICU LOS (hours) | -1.5 | 0.90 | | |
| | Mean Hospital LOS (days) | -0.9 | 0.24 | | |
| | 30 Days-Mortality | -1.5% | 0.42 | | |
| Yeh 2015 | Mortality (%) | +3% | 0.60 | 0% | 0.60 |
| | Median ICU LOS (Days) | -1 | 0.57 | 3 | 0.57 |
| | Median Hospital LOS (Days) | +1.5 | 0.48 | +9 | 0.48 |
| Murphy 2016[e] | ICU Mortality Rate (%) | -1.7%, | <0.05 | -1.2%, | <0.05 |
| | Hospital Mortality Rate (%) | -1.8% | <0.05 | -1.5% | <0.05 |
| | Mean ICU length of stay (units NR) | -0.1 | NS | 0 | NS |
| | Estimated total gross direct cost savings | | | $1,942,735 | |
| | Estimated net cost savings (accounting for incentive pay-out) | | | $1,544,095 (or $772,048 per year) | |

**Secondary Outcomes; length of stay (LOS), mortality, infection and expenditure or savings.** *Abbreviations*: ABGs= Arterial Blood Gases, CABG= Coronary Artery Bypass Grafting, CRBSI= catheter-related bloodstream infection, CSICU= cardiac surgery ICU, CVICU= cardiovascular ICU, ICU= intensive care unit, NR= Not Reported, NS= Not Significant, RCBs= Red Blood Cells, T1= Time 1 (baseline), T2= Time 2 (implementation), T3= Time 3 (follow-up), VAP= ventilator-associated pneumonia; **a)** NR whether average represents mean or median; **b)** T2: After audit system activated; T3: Audit system + Post-Feedback; **c)** CRBSI = hospital level, VAP = "incidences of VAP for randomly sampled quarters", surgical site infections = cardiovascular surgery service; **d)** T2: Weekly, group feedback; T3: Weekly, individual feedback as a group; Unclear if LOS values are means or medians; **e)** Estimated overall savings from reduction in ABGs, RBCs and Chest X-rays. ***Note: No relevant secondary outcomes reported for:** Solomon 1988, Wisser 2003, Calderon-Margalit 2005. ***Note: p-values are those reported in studies.**

**Table 4b. Secondary Outcomes for Comparative A&F Study**

| Study | Outcome | Absolute Δ Arm 2- Arm 1 | Reported p-value |
|---|---|---|---|
| Borgert 2016[f] | **Median ICU LOS (Days)** | | |
| | Implementation | 0 | p= 0.63 |
| | Post-implementation | +3 | p=0.57 |
| | **ICU Mortality (%)** | | |
| | Implementation | +0.8% | p=0.92 |
| | Post-implementation | -4.2% | p=0.57 |

**f)** Comparison of monthly, group A&F (Arm 1) to monthly, group A&F plus timely individual A&F (Arm 2). *Measures of Appropriateness:* Δ = Bundle. **Note: p-values are those reported in studies.**

**Summary of Studies on Improving Test Ordering**

**Table 3a** summarizes test and transfusion ordering (or appropriateness) outcome data from the included studies. Six studies aimed to reduce test ordering.[47,49–52,54,61] Five of these studies reported decreases (range: -1.6 mean tests per encounter, -1.72 to -8 tests per patient, -1.7 to -3.4 median tests per patient day, -613.1 tests per 100 hospital days) .[47,49,50,52,54,61] All three of the studies that tested significance found these reductions to be statistically significant.[49,50,54,61]

Four studies aimed to improve the appropriateness of test orders (as per compliance with a bundle,[55] guidelines (an algorithm),[49,50] standards for practice,[48] or whether the test had a written order[51]). Three of these studies reported statistically significant increases in compliance (range: +5.3 to +27%).[48–50,55] The remaining study reported a decrease in the proportion of inappropriate tests, however, upon assessing the number of overall tests per patient and inappropriate tests per patient, we noted undesired increases in both outcomes (range: +144 to +214 total tests/patient; +6 to +15 unordered tests per patient). No statistical test was reported.[51]

**Summary of Studies on Improving Transfusion Orders**

Three studies sought to reduce transfusion orders. All three reported decreases (range: -0.1 mean RBC unit orders/encounter, -79 FFP use/month [units not reported], -6 to -15.9% of patients receiving transfusion [all products], -2,288 units [all products]/year ); one reported a statistically significant difference,[61] one reported a statistically significant decrease for a subset of patients (overall significance not reported),[56] and one did not report a statistical test.[63]

Seven studies[21,53,55,58–60,62] aimed to improve the appropriateness of transfusion orders (as per compliance with a bundle,[21,55] a protocol/guideline,[58–60] guidelines plus clinical context,[62] and a combination of transfusion triggers and audit of patient factors [specifics unclear][53]). Outcomes included the over-transfusion rate, the odds of an inappropriate transfusion, the proportion of patients receiving inappropriate orders, the threshold at which a transfusion was given, the proportion of transfusions with an inappropriate threshold, or compliance with a bundle. Two studies saw significant decreases (range: OR of inappropriate transfusion 0.37-0.52; proportion of patients receiving unnecessary transfusion -6.6%);[58,59] one saw significant reductions during the intervention period and non-significant reductions at follow-up (range: -8 to -23% inappropriate transfusions; -5 to -8% over-transfusion rate; -0.3 to -0.5 g/dL mean pre-transfusion trigger);[60] one saw a significant reduction for one transfusion outcome, but no significant difference for another (-6.9% to -17% in proportion of transfusions over specific triggers; distribution of pre-transfusion platelet counts: p= 0.452);[53] and one saw a non-significant increase in compliance (range: +3.1 to +3.8% compliant episodes of transfusion).[55] Another study saw non-significant decreases for both inappropriate transfusions and transfusions consistent with guidelines (-14% and -1% respectively).[57] As described in **Table 3b**, the final included study was a head-to-head comparison of different types of A&F and found the enhanced strategy (timely individual + monthly group feedback) to significantly improve compliance of transfusions as compared to the monthly, group A&F (range: +31% to +36% bundle compliance).[21]

**Table 3a** also describes A&F in light of different comparators. Fourteen studies (88%) compared multifaceted interventions to usual care.[21,47–51,53–56,59–63] In most cases, data were only reported for the baseline and post-intervention periods, thus not enabling direct assessment of A&F components only. Nine of these studies[48–50,53–56,59–61] saw a statistically significant change in the hypothesized direction for at least

one of the outcomes (range: +15 to +27% in compliance, +5.3% to +21.6% compliant episodes, -0.1 to -1.6 orders/encounter, -1.7 to -3.4 median tests per patient day, -613.1 tests/ 100 hospital days, -6.9% to -17% in proportion of transfusions over specific triggers, -23% in inappropriate transfusions, -8% in over-transfusion rate, -0.5 g/dL mean pre-transfusion trigger, -6.6% in patients receiving unnecessary transfusion, -15.9% of patients receiving transfusion); three[47,52,63] reported changes in the hypothesized direction but did not report the significance (range: -1.72 to -8 tests per patient; -79 FFP use/month [units not reported]), and one[57] saw a statistically significant increase in transfusions "inconsistent with guidelines yet appropriate for the ICU" (+15% in requests), but non-significant decreases in both inappropriate (-14% in requests) and "consistent with guidelines" transfusions (-1% in requests). One study[51] did however provide a comparison of A&F alone versus usual care prior to implementing additional intervention components; undesired increases were seen for both overall (+144 tests per patient) and inappropriate tests per patient (+15 unordered tests per patient) (significance not reported). The only study[58] to implement a sole A&F intervention, saw a significant decrease in the odds and proportion of inappropriate transfusion (OR 0.37-0.52).

**Additional Outcomes**

Additional outcomes of interest, including length of stay, mortality, infection and expenditure, are summarized in **Tables 4a and 4b**. Length of stay (ICU or hospital) and mortality (ICU or hospital) outcomes were reported in totals of 11 studies[21,48–51,53,55,56,58–61] and 10 studies,[21,48–50,55,56,58–62] respectively. A statistically significant reduction in LOS measure was reported in only 1 of the 7 studies where it was tested.[55] Statistically significant decreases in mortality were found in 3 of the 8 studies in which it was tested.[55,58,61] In the two studies that reported infection rates, one saw no statistical difference,[48] and the other did not report statistical tests.[56] Savings or expenditure was reported in 5 studies,[47,49–51,56,61] however no statistical tests were reported.

**DISCUSSION**

A&F is known to be an effective component of interventions to improve practice,[25] and it is suggested to be a feasible strategy due to the availability of electronic health data.[24,26,27,29,64] However, relatively little work has explored how this change strategy can be effectively implemented in a critical care setting. Our systematic review yielded 16 studies, the majority of which showed positive effects, though their overall quality and rigour of design were assessed to be relatively weak.

Of the sixteen included studies, only one[58] assessed A&F alone as the sole intervention; the remaining studies assessed the effects of A&F alongside a range of intervention components (and in one case it was unclear if there were additional components). That most studies used a multifaceted intervention was reasonable, as previous literature has suggested that these interventions are more effective than single component interventions.[30,65–67] While the lack of simple comparison studies would seem to prevent us from directly assessing the effectiveness of A&F, some investigators have argued that the substantial literature (the latest Cochrane review included 140 trials[25]) demonstrates A&F's effectiveness, and negates the need for further testing of this intervention on its own.[68] Instead, the assessment of the conditions and mechanisms under which A&F is most effective is argued to be more likely to improve effectiveness of interventions.[25,68,69] Future primary studies may therefore consider the application of theory, process evaluations, and methods to compare different intervention component combinations to facilitate

identification of those that are most effective and to better understand the potential mechanisms.[70,71] Syntheses of the literature of the sort we report here are another way to advance work in this field.

Our review points to some mechanisms by which A&F might be made more effective in this context. Two studies in our review[21,58] suggest enhancing group feedback with individual feedback may improve intervention effectiveness. This is in line with a previous meta-analysis which found combined group and individual feedback yielded a larger effect size than either type of feedback alone.[29] Recent guidance around A&F[24] also suggests that provision of individualised feedback whenever possible is more likely to be effective, as group-level feedback is easier for an individual to discount. In the critical care context, both levels of feedback may be preferable, in that it addresses the team-based nature of critical care,[72,73] but still provides specific data for individual practitioners.

In 8 of the 17 interventions, feedback was either presented only once, it was not clearly specified how often feedback was provided, or the feedback was provided variably (only when an inappropriate order was placed).[47,48,51,53,54,59,62,63] The finding that not all A&F interventions provide iterative feedback suggests the important notion of the feedback loop[24] is overlooked in some cases. Recent guidance[24] recommends that feedback be provided multiple times, in order to close the feedback loop (i.e. a provider identifies a practice gap(s) based on the first instance of feedback, makes a change, and then needs subsequent instances of feedback to understand whether the practice change has resulted in improved outcomes).

While we were ideally interested in studies that aimed to reduce inappropriate tests and transfusions, it can be difficult to both define and adjudicate whether these resources are used appropriately.[17,41] Thus, some studies aim to reduce inappropriate orders, but simply measure the overall reduction in tests or blood components. For instance, in our small sample, six studies (37.5%) did not assess appropriateness. Albeit difficult to define, it is important for intervention studies to assess 'appropriateness.' This is needed to ensure that the tests and transfusions reduced are in fact unnecessary, and that underuse and patient harm does not occur, especially in the context of the ICU. The remaining 10 studies (62.5%) assessed appropriateness, with the majority identifying 'appropriateness' as compliance with guidelines or protocols.

The limited evidence we could find pertaining to patient length of stay (LOS) and mortality showed few significant differences. In part, this may be due to a lack of reporting on patient outcomes, an issue that has also been identified in other reviews.[30]

While our use of a subset of quality items precluded computing an overall quality score for each study, we found studies in this area lacking on important quality indicators. Many studies lacked a concurrent control group, and only one study used randomization. No time-series analyses were identified. Interventions were rarely described adequately to allow for replication. Lack of an appropriate control group and time-series analysis makes interpretation of study results difficult, as any effect seen may simply be due to coincidence, Hawthorne effects, seasonal differences or another undocumented change.[74–76] Non-randomized studies are at risk of introducing selection bias.[43] Furthermore, poor reporting of intervention details makes synthesis and replication more difficult.

**Strengths & Limitations**

We conducted the first comprehensive review of A&F interventions for improvement of test and transfusion ordering in critical care. We searched five electronic databases and two registries, as well as reference lists

of relevant articles and systematic reviews. Our search strategy was developed and peer reviewed with guidance from library information specialists, and screening, data extraction, and the risk of bias assessment were completed by two independent reviewers. Furthermore, in addition to summarizing the effectiveness of these interventions, our review is the first to assess characteristics of the A&F interventions in light of recent best practice guidance.[24]

Our study has limitations that warrant consideration. Inconsistency in reporting and differences in intervention component nomenclature complicated our categorization of intervention types. Using standard intervention categories and terms (such as those outlined by the Effective Practice and Organisation of Care (EPOC) taxonomy[77] or the Expert Recommendations for Implementing Change (ERIC) project[78]), reporting guidelines (such as the Template for Intervention Description and Replication (TIDieR) checklist[79]), and online access to more detailed descriptions of the interventions, may facilitate comparisons between studies in future reviews. While we worked hard to be comprehensive, some relevant studies may not have been included in our review as not all publications provide the relevant information in the abstract. Considerable work aiming to improve test and transfusion ordering may be conducted as quality improvement initiatives, and thus be less frequently published or more difficult to identify in electronic searches.[80,81] Finally, another limitation of this review (and others) is the potential for publication bias. We note that many of the included studies showed desired, albeit weak effects, which may suggest that studies that have positive and/or significant findings may be more likely to be submitted and published. Due to the heterogeneity in outcomes we were not able to assess the potential for publication bias by funnel plot, as Cochrane suggests asymmetry statistical tests be conducted with no less than ten studies.[82] Future updates to this review, however, may be able to address this issue.

**Guidance for Future Research**

Our research helped to identify several ways in which to build upon the extant literature. Use of more rigorous study designs, such as randomized controlled trials or cluster randomized controlled trials, would help to produce a higher quality evidence base around A&F interventions in the critical care setting. Greater focus on head-to-head trials of different types of A&F to study potential mechanisms of action and whether theory-informed suggestions for best practice help to optimize this intervention would advance this literature.[24,25,68] To allow for more robust and conclusive synthesis techniques such as meta-analysis and network meta-analysis, primary studies should employ comparative designs measuring and reporting on common outcomes (e.g. the number of laboratory tests per patient). Furthermore, adoption of consistent[77,78] and thorough reporting practices,[79] improved access to feedback templates, and development of core outcome sets would enable research teams to produce cumulative knowledge. Measurement and reporting of core patient outcomes and cost data will also help to assess whether these interventions are safe and sustainable.

**Implications & Conclusions**

This study showed that A&F is potentially effective in the critical care setting, but interventions are typically inconsistent with best practice recommendations for A&F interventions, and lack important indicators of study quality. In the majority of cases, A&F was implemented as one part of a multi-component intervention, limiting our ability to determine which components were contributing to the overall success. Additionally, the majority of studies in our sample were uncontrolled, leaving the results prone to bias.[75]

More research focussed on the optimization of A&F in critical care is warranted; initial signals of efficacy, and the lack of consistency with best practices, suggest these types of intervention can be improved. Future work should focus on understanding the mechanisms by which this intervention works,[24,83] particularly in this team-based environment. Assessment of whether interventions designed with more best practice recommendations[24] in place are more effective, would help to advance this literature, as would tools for standardizing and codifying the description and reporting of A&F interventions. This will help us to develop hypotheses as to how A&F may be optimized for the improvement of test and transfusion ordering in the critical care setting.

### Abbreviations

ABGs= Arterial Blood Gases, CABG= Coronary Artery Bypass Grafting, CBC= Complete Blood Count, CRBSI= catheter-related bloodstream infection, CSICU= cardiac surgery ICU, CVICU= cardiovascular ICU, FFP= Fresh Frozen Plasma, FP= Frozen Plasma, ICU= intensive care unit, LFT= Liver function tests, NR= Not Reported, NS= Not Significant, PT/PPT= prothrombin time/partial thromboplastin time, RBCs= Red Blood Cells, T1= Time 1 (baseline), T2= Time 2 (implementation), T3= Time 3 (follow-up), VAP= ventilator-associated pneumonia

## DECLARATIONS

### Ethics Approval and Consent to Participate

Not applicable.

### Consent for Publication

Not applicable.

### Availability of Data and Material

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

### Competing Interests

The authors declare that they have no competing interests.

### Funding

MF received a Queen Elizabeth II scholarship for her Master's thesis. MF also received a University of Ottawa Graduate Studies Scholarship and held a graduate studentship with the Ottawa Hospital Research Institute. Funding bodies had no role in the design of the study, collection, analysis, interpretation of data, or in the writing of the manuscript.

### Authors' Contributions

JCB and JP were responsible for the conception of this project and provided guidance and expertise throughout the entire project. MF drafted the manuscript, and JCB, JP, NM, LM and BH provided critical input and aided in the revision of the manuscript. MF and KC completed title and abstract screening, and

KC provided confirmation for inclusion and exclusion of full text articles screened by MF. Data extraction and quality assessment were completed by MF and NM. The guarantor of this review is MF.

**Acknowledgements**

The authors would like to thank Mish Boutet and Risa Shorr for their help in developing and reviewing the search strategy.

**APPENDIX A: Documented Changes from Original Protocol**

| Number | Section | Change(s) |
|---|---|---|
| 1 | Eligibility criteria, Population | - Specified physicians at all levels<br>- Added the healthcare professional examples phlebotomists and respiratory therapists<br>- Added the example cryoprecipitate |
| 2 | Eligibility criteria, Comparator | - Specified ***behavioural*** intervention<br>- Provided examples: (such as education, incentives, reminders, or systems-based changes).<br>- Due to the limited number of studies identified, we broadened our criteria to allow for inclusion of studies comparing different types of A&F. |
| 3 | Eligibility criteria, Secondary Outcomes<br><br>&<br><br>Outcomes and Prioritization | - Specified the appropriateness of **ordered laboratory tests or transfusions** (previously 'services') and added: 'for example as judged by the clinical context, or as compared to specified guidelines'<br>- The secondary outcome extubation failure was removed prior to completion of screening/ the start of data extraction. We decided to narrow our outcome focus to ensure manageable data extraction.<br>- We removed the secondary outcome transfusions (for laboratory test interventions) as this was already captured in our primary outcome.<br>- We removed the secondary outcome 'healthcare provider perceptions, beliefs and motivation to reduce inappropriate test and transfusion ordering' prior to completion of screening/ the start of data extraction. We decided to narrow our outcome focus to ensure manageable data extraction. |
| 4 | Eligibility criteria, Study Design | - Since our protocol registration we have specified that uncontrolled before-after studies are included in our review. Due to low study yield, we decided to broaden the scope of possible designs. Since our protocol registration we have also specified that conference abstracts, editorials and letters to the editor have been excluded. |
| 5 | Eligibility criteria, Setting | - We switched 'critical care setting' with 'intensive care setting' to reduce ambiguity |
| 6 | Eligibility criteria | - Added: 'As we anticipated relatively few studies to have been conducted specifically in ICUs, we included studies regardless of study year or follow-up time. To ensure we were aware of the full extent of the literature, we did not use year or language filters.' |
| 8 | Selection Process | - Originally, the full text articles were to be screened by two independent reviewers, however due to limited resources, full text articles were instead screened by one reviewer (MF). The justifications to include and exclude articles were confirmed by another member of the research team (KC). |
| 9 | Selection Process | - We originally said: "**After data extraction**, both reviewers will discuss whether pooling of the data is appropriate, and if so, which studies should be included in the meta-analysis." However, prior to data extraction the research team determined that the sample was highly heterogeneous and deemed meta-analysis inappropriate. |

| | | | |
|---|---|---|---|
| 10 | Data Collection | - | Data was originally to be extracted by one reviewer (MF) and confirmed by a second independent reviewer. However, due to changes in availability, data was extracted by two independent reviewers. |
| 11 | Data Collection | - | We had originally said: "using a piloted data extraction form. The data extraction form will be created and piloted using 4 A&F, non-ICU intervention studies (2 laboratory test studies and 2 transfusion studies). Non-ICU setting studies will be used as we expect the ICU setting literature to be limited and therefore do not want to use the available ICU data for piloting." However, due to resource limitations data extraction was first completed by one reviewer, and it was determined that only minor adjustments were required. |
| 12 | Data Items | - | Had originally said: "If more than one study arm contains a feedback component, the most complex arm will be assessed." However, we decided we would assess all study arms with an A&F component. |
| 13 | Individual Study Quality and Risk of Bias | - | While we anticipated that most of the studies would be of an observational nature (CBAs, ITSs), we had not anticipated that the majority would be non-controlled studies (UBAs). We had originally outlined that the methodological quality would be assessed using criteria specific to each study design using the EPOC Review Group's quality criteria (including 'Consumer Involvement' for all studies).[34] However, criteria is only provided for studies with a concurrent control group (RCT, CCT, CBA, ITS). We thus decided to instead use the quality criteria outlined by Kobewka et al.,[30] which is a modified version of the EPOC quality criteria. The similar reviews, which we had previously identified (by Cadogan et al.[16] and Kobewka et al.[30]) both used adapted, but different versions of the EPOC quality criteria. As Kobewka et al.'s review[30] did not restrict based on study design and also included studies without a concurrent control group, we decided to use their modified version of quality criteria. In Kobewka et al.'s review,[30] one set of criteria was used for all study designs. While we had originally stated that we would also assess the risk of bias for any RCTs using the Cochrane collaboration's tool[43], we decided to instead simplify matters, and stay consistent with Kobewka et al.[30] Two additional items from the EPOC Review Group's quality criteria[34] "Blinded assessment of primary outcome(s)" and "Reliable primary outcome measure(s)" were also assessed. |
| 14 | Individual Study Quality and Risk of Bias | - | We originally said: "The reviewers will be trained to assess risk of bias and methodological quality through reading of the educational materials associated with each tool.[43,53]" This was no longer applicable. |
| 15 | Synthesis of Results & Summary Measures | - | We had originally planned to enter the extracted data into RevMan5.3 software[84] to generate weighted summary measures and forest plots for presentation of the results. The mean difference ± 95% confidence limits were to be used to summarize continuous outcomes, and the Odds Ratio (OR) ± 95% confidence limits were to be used to summarize categorical outcomes. This was no longer applicable. |
| 16 | Synthesis of Results & Summary Measures | - | We originally planned to create the results tables after the data extraction consensus meeting. However, due to limited resources the results tables were constructed prior to the consensus meeting and updated after the consensus meeting. |

| | | |
|---|---|---|
| | | - We also planned to use the summary of study characteristics table to assess qualitative study heterogeneity (in terms of study type, type of control, ICU type, feedback intervention, and comparator), and had said "if two or more studies are judged to be similar… pooling of the outcomes will be deemed appropriate." However, during article selection, we determined that the study sample was too heterogeneous for pooling (in terms of interventions and outcomes), and meta-analysis was deemed inappropriate. |
| 17 | Synthesis of Results & Summary Measures | - We originally said: "The information obtained from the quality and risk of bias analyses will aid in the determination of whether meta-analysis is appropriate." However, this was no longer applicable as we decided meta-analysis would be inappropriate prior to the quality assessment of the articles. |
| 18 | Additional Analyses: Sensitivity and Subgroup Analysis | - We had originally said "For RCTs, if meta-analysis is deemed appropriate, the risk of bias assessment will be used to complete sensitivity analyses. This will be done by completing analyses without studies with a high risk of bias for selective reporting and incomplete outcome data." This was no longer applicable. |
| 19 | Additional Analyses: Sensitivity and Subgroup Analysis | - This entire section was removed.<br>- As meta-analysis was deemed inappropriate, planned sensitivity and sub-group analyses could not be completed. These categories had included: study quality, type of funding, ICU type, study country (to look at possible differences in payment systems), intervention type (multifaceted versus single mechanism). (*Note: ROB was not included as we decided to no longer assess this- see change #13.) |
| 20 | Meta-Biases | - This entire section was removed.<br>- We had originally planned to produce a funnel plot to assess the potential for publication bias across studies using the RevMan5.3 software.[84] This was no longer applicable as meta-analysis was not completed.<br>- The previously planned sensitivity analyses, to be completed by removing studies with a high risk of bias for selective reporting and incomplete outcome data, could not be conducted as meta-analysis was deemed inappropriate. (*Also note: ROB was not assessed- see change #13.) |
| 21 | Risk of Bias Across Studies: Confidence in Cumulative Evidence | - The overall confidence in cumulative evidence was to be assessed using the GRADE scoring system (Grading of Recommendations, Assessment, Development and Evaluations).[85] Due to the fact that meta-analysis was deemed inappropriate, and due to limited resources, we decided to forego GRADE scoring. |
| 22 | Reporting of Results | - We had originally planned to first discuss the effectiveness of each intervention, and then compare and contrast the heterogeneity between study characteristics. We have instead provided an overview, and have grouped discussions based on intervention type (i.e. A&F alone, multifaceted).<br>- Higher quality studies were to be discussed first for each outcome. Overviews were instead provided.<br>- We had originally planned to discuss our primary and secondary outcomes for each study, and group studies based on comparison type as per Ivers et al.[25]; this was done for our primary outcome, an overview was provided for secondary outcomes. |

## APPENDIX B: Justification for Choice of Outcomes & Changes in Secondary Outcomes

*Justifications*

The primary outcome of interest for this review was the number of laboratory tests or transfusions. We anticipated that this outcome could be reported in a variety of formats:

- The number of appropriate laboratory tests or transfusions (or compliance with guidelines)
- The number of inappropriate laboratory tests or transfusions
- The overall number of laboratory tests or transfusions ('appropriateness' not measured)

We chose our primary outcome as the overall number of laboratory tests or transfusions, because not all studies measure the appropriateness of services.[86] Secondary outcomes included appropriateness of services (where possible, along with definitions of "appropriateness"), Length of Stay (LOS), mortality, infections, and laboratory or blood product expenditure. We chose to extract the patient outcomes LOS, mortality and infections, as it is important to ensure patient safety is not affected by the intervention. We also extracted how the interventions affected laboratory test or blood product expenditure, as studies aiming to improve the appropriate use of medical resources tend to highlight the negative financial implications of overuse[87] and the opportunity for cost-reduction.[88,89]

*Changes*

Since registration of our protocol, but prior to completion of screening and the start of data extraction, we removed the secondary outcomes 1) extubation failure and 2) healthcare provider perceptions, beliefs and motivations to reduce inappropriate test and transfusion ordering, to narrow the focus and ensure manageable data extraction.

**APPENDIX C: Medline Search Strategy**

1 exp Intensive Care Units/
2 exp Critical Care/
3 Critical Illness/
4 (intensive care or icu or nicu).tw.
5 (critical* adj2 (ill* or care)).tw.
6 1 or 2 or 3 or 4 or 5
7 Clinical Audit/
8 exp Medical Audit/
9 Nursing Audit/
10 audit*.tw.
11 exp Management Audit/
12 feedback/ or formative feedback/ or feedback, psychological/
13 feedback*.tw.
14 (feed* adj2 back?).tw
15 benchmark*.tw.
16 "utilization review"/ or "concurrent review"/
17 Peer Review, Health Care/
18 (utili?ation review or "usage review" or data review).tw.
19 "Quality of Health Care"/
20 or/7-19
21 6 and 20
22 exp Blood Transfusion/
23 transfus*.tw.
24 ((rbc or red blood or erythrocyte or plasma or platelet) adj2 therap*).tw.
25 (blood product? adj2 therap*).tw
26 exp Diagnostic Services/
27 "Diagnostic Techniques and Procedures"/
28 exp Laboratories/
29 Diagnostic Tests, Routine/
30 ((lab or laboratory or diagnostic test$) adj3 (use$ or utili$ or requisition$ or usage)).tw.
31 lab work*.tw.
32 laboratory work*.tw.
33 exp Clinical Laboratory Techniques/
34 or/22-33
35 21 and 34

**APPENDIX D: Additional Details on Search Strategy & Sample**

*Detailed Search Strategy Development*

An initial systematic review search strategy was developed with help from a library information specialist. Medical Subject Headings (MeSH terms) and title and abstract terms ('.tw') were chosen for the general categories 'Laboratory Tests', 'Transfusions', 'Intensive Care', and 'Audit and Feedback'. The key words and MeSH terms of relevant primary studies and systematic reviews were assessed during the development process. The initial search strategy was then reviewed by a second, independent library information specialist. The second librarian made several recommendations such as:

- The removal of several acronym title and abstract terms (for example MICU, SICU, TICU).
- The removal of terms relating to quality improvement (for example 'Quality Improvement/', 'Quality Assurance, Health Care/', '(Quality adj2 Control).tw')
- Replacing the title and abstract terms 'lab*' and 'test*' with 'lab', 'laboratory' or 'diagnostic test' and combining with 'use', 'utilisation', 'requisition' or 'usage'.
- The addition of the MeSH term 'Critical Illness/'
- Combining transfusion related title and abstract terms with 'therapy'

After checking whether the search strategy was able to pick up 14 target articles, the librarian additionally suggested the addition of '"Quality of health care"/' and 'lab work*.tw'. With the addition of these terms the strategy was capable of capturing 13 of the 14 target articles.

The study team then met to discuss any modifications, and agreed on the final strategy. The Medline search strategy was translated to the remaining databases (Embase, Cochrane Library, CINAHL, and PsychInfo). The trial registries 'ClincalTrials.gov' and ISRCTN were additionally searched, using the search terms 'intensive care', 'feedback', and 'transfusion' or 'lab'.

*Additional Details on Search Results*

An additional 144 citations were also identified by searches within ClinicalTrials.gov (89), the ISRCTN registry (50), hand-searching the references of included articles (2), and various means (i.e. searching for the full text publication of a citation from a letter to the editor, finding a full text publication for a conference abstract, and identification by one of the reviewers) (3).

*Challenges Encountered*

We found that studies containing an audit and feedback component did not always refer to this component as 'audit and feedback'. The inconsistency in reporting of intervention components has been noted previously in the literature.[78] Thus, studies that did not include words in the abstract related to 'intensive care', 'audit' or 'feedback' may not have been picked up. However, we did include MeSH and text words related to 'critical care', 'critical illness', 'utilization review' and 'quality of health care' in hopes that this would pick up additional studies.

In many cases, we were not able to determine intervention characteristics due to lack of reporting, which prevented us from being more complete in our assessment of feedback characteristics. Reporting guidelines,

or the use of online supplemental appendices with examples of the feedback template (or a more detailed description), may help to ensure adequate reporting of all pertinent intervention details.

*Additional Study Characteristics*

The number of patients in each study ranged from 28 to 22,567 and the number of health care providers (HCPs) per study ranged from 29 to 120, though this variable was only reported in 3 studies.[21,49,50,59]

The patient population across studies varied widely, including (where specified): in-patients (intensive care units, internal medicine, cardiology, obstetrics and gynecology, pediatrics, hemato-oncology, surgery)[54]; adult patients admitted to cardiothoracic (2), neuroscience (2), medical (2), and surgical (1) ICUs[61]; patients in ICU from various departments (visceral surgery/gynecology/obstetrics, cardiovascular surgery, and internal medicine wards (gastroenterology, nephrology, oncology, cardiology))[52]; patients from the "medical and surgical services" (including ICUs)[63]; "adult patients admitted to a … medical-surgical ICU"[62]; surgical ICU patients[49,50]; adult patients in a surgical ICU who were "stable, low-risk"[60]; cardiac surgery patients (adult)[58,59]; coronary artery bypass graft (CABG) cardiac surgery and cardiovascular ICU patients ("including other cardiovascular surgical, vascular, and cardiology patients")[56]; "neurosurgery, surgery and trauma admissions"[51]; "mechanically ventilated intensive care unit (ICU) patients"[48]; severe sepsis or septic shock patients (≥18)[55]; newborns[53]; "newborns admitted to a level 2 neonatal unit that provides care to low to moderately high-risk neonates" and "for the assessment and potential treatment of sepsis".[47]

Only one study[21] reported the average number of years in practice (Arm 1 median =10, Arm 2 median=14) and average age (Arm 1 median=41, Arm 2 median= 44) for recipients, though another study[49] did specify the proportion of nurses in their first (15%) or second year (15%) of training.

## APPENDIX E: Reporting Quality Assessment

### *Criteria Items*

We assessed the quality of included studies using the seven criteria items reported by Kobewka et al.[30] (modified from the Effective Practice and Organization of Care (EPOC) group).[34] We also assessed two additional items from the EPOC Review Group's quality criteria: 1) "Blinded assessment of primary outcome(s)" and 2) "Reliable primary outcome measure(s)".[34(pp20-21)] The EPOC Review Group's Data Collection Checklist[34] was used as a guide for the applicable components. The quality assessment criteria were assessed as 'Yes', 'No' or 'Unclear'. While Kobewka et al.[30] assessed criteria items as either 'Yes' or 'No', the reviewers felt an 'Unclear' category was necessary.

1. "Were patients similar between groups?
2. Were those ordering the tests [or transfusions] similar between groups?
3. Was there a concurrent control group?
4. Was the intervention described adequately enough to be replicated?
5. Was there a risk of contamination between experimental and control groups?
6. Were the results reported per patient instead of per institution or per physician?
7. Was a time-series analysis conducted?"[30]
8. "Assessment of primary outcome blinded [or objective]?
9. Primary outcome measure reliable?"[34]

**Individual Quality Assessment of Included Studies**

| Study | Patients were similar between groups | Providers were similar between groups | Was there a concurrent control group? | Intervention was described in sufficient detail to be replicated | Was there risk of contamination between groups? | Was reduction in test use reported per patient? | Was time-series analysis performed? | Assessment of primary outcome blinded (or primary outcome objective)? | 1° Outcome measure reliable? |
|---|---|---|---|---|---|---|---|---|---|
| **Solomon, 1988** | Unclear | Yes | No | No | N/A | No | No | Yes | Unclear |
| **Paes, 1994** | Unclear | Yes | No | No | N/A | Yes | No | Yes | Yes |
| **Hendryx, 1998** | Yes | Yes | Yes | No | No | No | No | Yes | Yes |
| **Merlani, 2001 & Diby, 2005**[a] | Yes[a] | Yes | No | No | N/A | Yes[a] | No | Yes & Unclear | Unclear |
| **Beland, 2003** | Unclear | Yes | No | No | N/A | No | No | Yes | Unclear |
| **Wisser, 2003** | Unclear | Yes | No | No | N/A | Yes | No | Yes | Unclear |
| **Petäjä, 2004** | Unclear | Unclear | No | No | N/A | No | No | Yes | Yes & Unclear |
| **Calderon-Margalit, 2005** | Unclear | Yes | No | No | N/A | No | No | Yes | Yes |
| **Schramm, 2011** | No | Yes | No | No | N/A | Yes | No | Yes | Unclear |
| **Masud, 2011** | Unclear | Yes | No[b] | No | N/A[b] | Yes & No | No | Yes | Yes |
| **Arnold, 2011** | Yes | Yes | No | No | N/A | No | No | Yes[c] | No |
| **Beaty, 2013** | Unclear | Yes | No | No | N/A | No | No | Yes | Yes |
| **Gutsche, 2013** | No | Yes | No | No | N/A | Yes | No | Unclear | Unclear |
| **Yeh, 2015** | Yes | Yes | No | No | N/A | No | No | Yes | Unclear |
| **Murphy, 2016** | No | Yes | No | No | N/A | Yes | No | Yes | Yes |
| **Borgert, 2016** | Yes | Yes | Yes | No | Unclear | No | No | Yes | Unclear |

a) Used the most definitive answer from Merlani et al.[49]; b) Assessed for the data extracted; c) Blinded; all others coded as 'yes' were found to have objective outcomes; Note: Questions 1-7 as per Kobewka et al.[30] and 8-9 as per Cochrane Effective Practice and Organisation of Care Review Group (EPOC)[34]

**Summary of Study Quality**

| Quality Criteria | Number of Studies (%) | Quality Criteria | Number of Studies (%) |
|---|---|---|---|
| **1. Were patients similar between groups?** | | **6. Were the results reported per patient instead of per institution or per physician?** | |
| Yes | 5 (31%) | Yes | 6 (38%) |
| No | 3 (19%) | No | 9 (56%) |
| Unclear | 8 (50%) | Yes & No | 1 (6%) |
| **2. Were those ordering the tests [or transfusions] similar between groups?** | | **7. Was a time-series analysis conducted?** | |
| Yes | 15 (94%) | No | 16 (100%) |
| Unclear | 1 (6%) | | |
| **3. Was there a concurrent control group?** | | **8. Assessment of primary outcome blinded (or objective)?** | |
| Yes | 2 (12.5%) | Yes | 14 (88%) |
| No | 14 (87.5%) | Unclear | 1 (6%) |
| **4. Was the intervention described adequately enough to be replicated?** | | Yes and Unclear | 1 (6%) |
| No | 16 (100%) | **9. Primary outcome measure reliable?** | |
| **5. Was there a risk of contamination between experimental and control groups?** | | Yes | 6 (38%) |
| No | 1 (6%) | No | 1 (6%) |
| Unclear | 1 (6%) | Unclear | 8 (50%) |
| Not Applicable | 14 (88%) | Yes and Unclear | 1 (6%) |

Note: Questions 1-7 as per Kobewka et al.,[30] questions 8-9 as per Cochrane Effective Practice and Organisation of Care Review Group (EPOC)[34]

**APPENDIX F: PRISMA Checklist**[45]

| Section/topic | # | Checklist item | Reported on page # |
|---|---|---|---|
| **TITLE** | | | |
| Title | 1 | Identify the report as a systematic review, meta-analysis, or both. | 17 (2) |
| **ABSTRACT** | | | |
| Structured summary | 2 | Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number. | 17 (2) |
| **INTRODUCTION** | | | |
| Rationale | 3 | Describe the rationale for the review in the context of what is already known. | 18 (3) |
| Objectives | 4 | Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS). | 18-19 (3-4) |
| **METHODS** | | | |
| Protocol and registration | 5 | Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number. | 19 (4) |
| Eligibility criteria | 6 | Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale. | 19 (4) |
| Information sources | 7 | Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched. | 20 (5) |
| Search | 8 | Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated. | 44 (29) |
| Study selection | 9 | State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis). | 20 (5) |
| Data collection process | 10 | Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators. | 20 (5) |
| Data items | 11 | List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made. | 20 (5) |
| Risk of bias in individual studies | 12 | Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis. | 21 (6) |
| Summary measures | 13 | State the principal summary measures (e.g., risk ratio, difference in means). | 21 (6) |
| Synthesis of results | 14 | Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., $I^2$) for each meta-analysis. | N/A |

| Section/topic | # | Checklist item | Reported on page # |
|---|---|---|---|
| Risk of bias across studies | 15 | Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies). | N/A |
| Additional analyses | 16 | Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified. | N/A |
| **RESULTS** | | | |
| Study selection | 17 | Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram. | 22 (7) |
| Study characteristics | 18 | For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations. | 21-24 (6-9) |
| Risk of bias within studies | 19 | Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12). | 23 and 48-49 (8 and 33-34) |
| Results of individual studies | 20 | For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot. | a) 25-33 (10-18) b) N/A |
| Synthesis of results | 21 | Present results of each meta-analysis done, including confidence intervals and measures of consistency. | N/A |
| Risk of bias across studies | 22 | Present results of any assessment of risk of bias across studies (see Item 15). | N/A |
| Additional analysis | 23 | Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]). | N/A |
| **DISCUSSION** | | | |
| Summary of evidence | 24 | Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers). | 34-35 (19-20) |
| Limitations | 25 | Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias). | 36-37 (21-22) |
| Conclusions | 26 | Provide a general interpretation of the results in the context of other evidence, and implications for future research. | 37-38 (22-23) |
| **FUNDING** | | | |
| Funding | 27 | Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review. | 38 (23) |

**APPENDIX G: Definitions of Appropriateness**

| Study | Definition of Appropriateness |
|---|---|
| Solomon 1988 | N/A- Appropriateness not assessed |
| Paes 1994 | N/A- Appropriateness not assessed |
| Hendryx 1998 | Compliance with standards for practice:<br>"The data collection protocol was derived employing objective indicators established by the Task Force on Guidelines, Society for Critical Care Medicine. Using the indicators as a guide, the university specialist team developed more specific standards based on current practice. The protocol reflects basic processes of ICU care that should be delivered regardless of technological sophistication or patient mix."[48] |
| Merlani 2001 & Diby 2005 | Adherence to guidelines/ algorithm, developed in-house:<br>"The pilot guideline, which included a comprehensive time frame and the use of pulse oximetry, was devised as an algorithm based on three pathways corresponding to pH, PaO2, and PaCO2 and included a minimum number of three mandatory analyses per day for safety reasons. The guideline was approved by the two other unit consultants. To incorporate the opinion and the experience of the users we amended the pilot version of the guideline ten months later. In this consolidated version, the time frames were widened and the daily mandatory tests removed. We added a list of clinical settings in which the algorithm should not be applied. The consolidated guideline was validated by four critical care experts from outside the unit."[49] |
| Beland 2003 | Aimed to reduce 'unordered' tests (tests with no written order) |
| Wisser 2003 | N/A- Appropriateness not assessed |
| Petäjä 2004 | Unclear; various cut-offs mentioned |
| Calderon-Margalit 2005 | N/A- Appropriateness not assessed |
| Schramm 2011 | Adherence to a bundle; "…modified from recommendations by the International Surviving Sepsis Campaign and the Institute for Health-care Improvement. Each element of the resuscitation bundle was expected to be completed within 6 hrs. Compliance with each element of the bundle as well as the overall compliance using all-or-none approach was recorded."[55]<br><br>Lactate: "Measured before or within 1 hr after blood culture"[55]<br>Blood Culture: "Drawn before antibiotics administered"[55]<br>Red blood cell administration: "Transfused if hematocrit <30% and ScvO2 <70% or mixed venous O2 <65% despite fluid resuscitation"[55] |
| Masud 2011 | N/A- Appropriateness not assessed |

| Study | Definition of Appropriateness |
|---|---|
| Arnold 2011 | "The appropriateness of each FP request was adjudicated … based on clinical data (bleeding, use of other blood products, and planned or recently performed invasive procedures) and laboratory parameters (INR, partial thromboplastin time, hemoglobin level, and platelet count). Frozen plasma requests were adjudicated as inappropriate if they were inconsistent with published guidelines and considered to be unnecessary in the clinical context, consistent if they matched guideline recommendations, and appropriate for the ICU yet inconsistent with guidelines if the request fell outside published recommendations but was felt to be reasonable given the unique requirements of the critically ill."[62] *Note: Also provided a table with specific criteria as well as a reference to Crosby E, Ferguson D, Hume H, et al. Guidelines for red blood cell and plasma transfusion for adults and children. Can Med Assoc J 1997;156:S1-S24. |
| Beaty 2013 | Adherence to a protocol; Transfusion trigger of Hgb < 8 gm/dL |
| Gutsche 2013 | Compliance with a clinical practice guideline, developed in-house: "All transfusions associated with a hemoglobin (hgb) < 7.0 mg/dL were considered to be in compliance with the guideline. If a patient was transfused with an hgb from 7 mg/dL to 7.9 mg/dL, then the chart was analyzed for evidence of organ ischemia or pressor requirement. Transfusion associated with an hgb from 7 mg/dL to 7.9 mg/dL without evidence of organ ischemia, shock, pressor requirement, or hemorrhage as evidence by chest tube output >250 mL/hour or documentation of alternate evidence of hemorrhage was considered guideline noncompliance. All transfusions for an hgb >8.0 mg/dL were considered guideline noncompliant if there was no evidence for hemorrhage as evidence by a chest tube output >250 mL/hour or documentation of alternate evidence of hemorrhage."[59] |
| Yeh 2015 | Compliance with guidelines ("stable, low-risk patients"[60]):<br>• "Inappropriate transfusions (Hgb trigger > 8.0 mg/dL)"[60]<br>• "Over-transfusion (defined as post-transfusion Hgb > 10.0 g/dL)"[60] |
| Murphy 2016 | N/A- Appropriateness not assessed |
| Borgert 2016 | Adherence to a bundle ("all-or-none (AON)-approach"[21])<br><br>1) "Is the haemoglobin (Hb) result considered reliable?"[21]<br>2) "Have you verified if the Hb transfusion threshold was reached?"[21]<br>3) "Have you verified if informed consent was obtained?"[21]<br>4) "Is the identity of the patient checked by two persons independently before transfusion?"[21]<br>5) "Is the blood product checked by two persons independently before transfusion?"[21] |

***Note: Reference, table and figure numbers from original articles have been removed to avoid confusion.**

## APPENDIX H: Excluded Full-Text Articles Sorted by Reason for Exclusion

| Title | First Author | Year of Publishing | Exclusion Reason | Total Number |
|---|---|---|---|---|
| A clinical audit for improving utilization of tests and reducing costs in surgical wards and intensive care unit | Vezzani, A | 2009 | Conference Abstract | 68 |
| A multifaceted strategy to reduce inappropriate use of fresh frozen plasma transfusions in the intensive care unit | Arnold, DM | 2009 | Conference Abstract | |
| A multi-site audit of transfusion administration practice with use of an online tool to capture results | Owens, W | 2011 | Conference Abstract | |
| A Retrospective Analysis on the Effectiveness of a Maternal Hemorrhage Plan...Proceedings of the 2015 AWHONN Convention | Sincore, TJ | 2015 | Conference Abstract | |
| A retrospective study of transfusion practices in a pediatric intensive care unit with an operating blood management program | Hassan, NE | 2010 | Conference Abstract | |
| An interdisciplinary program for improving the recognition and treatment of severe sepsis | Oxman, D | 2013 | Conference Abstract | |
| Appropriate regulation of routine laboratory testing can reduce the costs associated with patient stay in intensive care | Goddard, K | 2011 | Conference Abstract | |
| Audit of compliance with the severe sepsis resuscitation bundle in patients admitted to ICCU | Irving, J | 2010 | Conference Abstract | |
| Beyond the "bundle": Interventions to decrease catheter associated bloodstream infections in a community teaching hospital | Dumigan, DG | 2012 | Conference Abstract | |
| Blood management in the ICU: Changing transfusion practice in critically ill patients | Norgaard, A | 2010 | Conference Abstract | |
| Blood stream infections in paediatric critical care: Getting the diagnosis right - A quality improvement project | McCluskey, J | 2014 | Conference Abstract | |

| | | | | |
|---|---|---|---|---|
| Changing clinical practice of central line culture investigation in a regional intensive care unit category: Clinical lesson | O'Hare, P | 2011 | Conference Abstract | |
| Clinical audit system in implementing Surviving Sepsis Campaign guidelines in patients with peritonitis | Valiveru, RC | 2014 | Conference Abstract | |
| Clinical utility of endotracheal tube cultures from neonates in a neonatal intensive care unit: Completion of an audit cycle category: Clinical lesson | Yew, P | 2011 | Conference Abstract | |
| Consolidating blood draws in children as a blood conservation technique | Mack, E | 2013 | Conference Abstract | |
| Control of platelet transfusions in a teaching hospital setting | Copplestone, A | 2012 | Conference Abstract | |
| Cryoprecipitate prospective audit program: Impact and limitations | Osegueda, V | 2013 | Conference Abstract | |
| Evaluation of the appropriateness of frozen plasma usage in the era of prothrombin complex concentrates: A retrospective study | Shih, AW | 2013 | Conference Abstract | |
| Impact of clinical resource management: Laboratory optimization is associated with reduced cost and improved patient outcomes | Shin, AY | 2013 | Conference Abstract | |
| Impact of remote electronic monitoring and tele-intensive care unit based algorithm in monitoring packed red cell transfusion behavior for anemia of critical illness: Longitudinal multi-year experience from a single community health system in the United S | Li, N | 2014 | Conference Abstract | |
| Implementation of a blood management program in the intensive care unit (ICU) | Umezawa Makikado, LD | 2013 | Conference Abstract | |
| Implementation of a massive haemorrhage protocol: The legnano experience | Novelli, CAE | 2014 | Conference Abstract | |
| Implementation of a massive Transfusion Protocol at a university medical center | Zantek, ND | 2009 | Conference Abstract | |

| | | | | |
|---|---|---|---|---|
| Implementing transfusion practice guidelines - The Austrian approach | Gombotz, H | 2009 | Conference Abstract | |
| Improved computerized order entry for pRBC transfusion associated with decreased product utilization | Wool, G | 2015 | Conference Abstract | |
| Improved outcomes using multidisciplinary teams to implement sepsis bundles | Seoane, L | 2011 | Conference Abstract | |
| Improving a health care system's critical lab value reporting process through a multi-disciplinary quality team | McCollum, D | 2015 | Conference Abstract | |
| Improving patient outcomes following emergency laparotomy: Assessing the impact of quality improvement measures based on NELA recommendations | D., Pachter | 2016 | Conference Abstract | |
| Improving the efficient use of platelet transfusions in critically ill patients with decision support integrated into the electronic patient record | A., Thoppil | 2011 | Conference Abstract | |
| Incidence of nosocomial blood stream infections, antibiotic resistances and blood culture ordering and testing practices: A Thuringia-wide prospective population-based quality management project (AlertsNet) | Mikolajczyk, R | 2013 | Conference Abstract | |
| Incorporating the evidence using cpoe and dashboards: Implementation of SCCM adult red blood cell transfusion clinical guidelines | Luehr, E | 2010 | Conference Abstract | |
| Institution-wide quantification of iatrogenic blood loss using a novel informatics-driven methodology | Ledingham, DL | 2010 | Conference Abstract | |
| Introduction of a sepsis screening tool and care bundle using a moulage-based training program to improve recognition and management of severe sepsis | Stephens, T | 2012 | Conference Abstract | |
| Investigating the frequency and volume of blood sampling in critical care patients in an attempt to reduce iatrogenic anaemia | Laird, AE | 2011 | Conference Abstract | |

| | | | | |
|---|---|---|---|---|
| Making the Sepsis Six count on a high-risk pregnancy unit, delivering an improvement in sepsis care | Pritchard, N | 2016 | Conference Abstract | |
| Management of maternal sepsis in a large UK District General Hospital: Audit results, interventions and introduction of a regional audit tool | Katakam, N | 2014 | Conference Abstract | |
| Monitoring compliance with transfusion guidelines in hospital departments by electronic data capture | Norgaard, A | 2014 | Conference Abstract | |
| Monitoring the compliance with transfusion guidelines in hospital wards | Norgaard, A | 2013 | Conference Abstract | |
| Moving a bone marrow transplant unit towards a high reliability unit | Mott, B | 2014 | Conference Abstract | |
| National comparative audit of blood and component use in cardiac surgery | Allard, S | 2013 | Conference Abstract | |
| Optimizing transfusion in the Intensive Care Unit (ICU) after cardiac surgery: Combining a transfusion algorithm based upon thrombelastography (TEG) with business change management intervention in both doctors and nurses | Jepsen, K | 2011 | Conference Abstract | |
| Overview of blood component usage at a central hospital | Truus, R | 2011 | Conference Abstract | |
| Paediatric community acquired pneumonia-improving management | R., Robertson | 2015 | Conference Abstract | |
| Pediatric blood sparing: A joint pediatric-laboratory quality improvement initiative to reduce pediatric blood sample volumes for laboratory testing | Baffa, A | 2009 | Conference Abstract | |
| Point of care coagulation and platelet function testing: Implementation of a new service in a tertiary cardiac surgery unit | Pearse, B | 2013 | Conference Abstract | |
| Potential of improving transfusion practice in critical care | Noel, S | 2010 | Conference Abstract | |

| | | | | |
|---|---|---|---|---|
| Process management of sepsis. the implementation of a modified triage tool (Septic) in an inner city emergency department and its effects on the management of sepsis | O'Connor, G | 2011 | Conference Abstract | |
| Real time monitoring of blood transfusion in intensive care following cardiac surgery | Ng, CSH | 2011 | Conference Abstract | |
| Red cell transfusion in critical care: An audit on recent British Society of Haematology Guidelines | Kendrick, K | 2014 | Conference Abstract | |
| Red cell transfusions on patients during and after a critical care admission: An audit into current practice at the Queen's Medical Centre, Nottingham | Redding, N | 2014 | Conference Abstract | |
| Reducing platelet (PLT) and red blood cell (RBC) utilization in a 325-bed hospital | Sutton, BC | 2013 | Conference Abstract | |
| Reduction in inappropriate red cell transfusion through prospective computerized order auditing | Desrosiers, KP | 2013 | Conference Abstract | |
| Retrospective analysis of platelet transfusion practice over 10 years in an ICU in the United Kingdom | Connor, DM | 2010 | Conference Abstract | |
| Setting up a patient blood management programme | Wood, E | 2013 | Conference Abstract | |
| Severe sepsis: Craigavon area hospital | McKeague, R | 2014 | Conference Abstract | |
| Surviving sepsis: Improving the early treatment and recognition in acute medical patients using an audit proforma | Revill, A | 2010 | Conference Abstract | |
| Sustainable improvement in transfusion practices through pre transfusion audit in Brazil | Lazar, AS | 2011 | Conference Abstract | |
| The effects of audit and research in postpartum haemorrhage: Benefits for all! | Moses, T | 2016 | Conference Abstract | |

| | | | | |
|---|---|---|---|---|
| The impact of critical care nurse training on thromboelastography usage to guide perioperative blood component transfusion in a cardiothoracic critical care unit | Shah, A | 2012 | Conference Abstract | |
| The transfusion safety officer: An effective tool in patient blood management | Levine, RL | 2015 | Conference Abstract | |
| Towards reducing inappropriate ICU blood transfusions: Combining education and electronic reminders | Minik, O | 2015 | Conference Abstract | |
| Tracking ventilator bundle compliance and ventilator-associated events | Frisch, J | 2016 | Conference Abstract | |
| Trends in best practice adherence in a large cohort of ICUS: 2005-2010 | Badawi, O | 2010 | Conference Abstract | |
| Use of coagulation screening in the critical care unit | Rice, A | 2012 | Conference Abstract | |
| Usefulness of sepsis screening tools and education in recognizing the burden of sepsis on hospital wards | Galtrey, EJ | 2015 | Conference Abstract | |
| Utlization of red cell concentrate at the National Hospital of Sri Lanka for a period of four years | Adikarama, Y | 2009 | Conference Abstract | |
| Weekly feedback with identification of physician-specific behaviour improves adherence to blood utilization protocol in cardiac surgery | Beaty, CA | 2012 | Conference Abstract | |
| Employing quality improvement methodology in sepsis: An electronic sepsis order set further improves compliance with the Surviving Sepsis Campaign 3-hour bundle | Rossi, S | 2014 | Conference Abstract | |
| A major haemorrhage protocol improves the delivery of blood component therapy and reduces waste in trauma massive transfusion. | Khan, S | 2013 | Wrong intervention | 26 |
| A utilization management intervention to reduce unnecessary testing in the coronary care unit. | Wang, TJ | 2002 | Wrong intervention | |

| | | | | |
|---|---|---|---|---|
| Algorithmic and consultative integration of transfusion medicine and coagulation: a personalized medicine approach with reduced blood component utilization. | Brown, RE | 2011 | Wrong intervention | |
| An audit of catheter specimen testing practices in the ICU. | Curran, E | 1997 | Wrong intervention | |
| Can the cost of distal vascular reconstruction be reduced without sacrificing quality? Analysis of 500 cases | Choi, DS | 2000 | Wrong intervention | |
| Changing practices of red blood cell transfusions in infants with birth weights less than 1000 g. | Maier, RF | 2000 | Wrong intervention | |
| Computerized quality assurance of decisions to transfuse blood components to critically ill patients. | Pentti, J | 2003 | Wrong intervention | |
| Current red blood cell transfusion practices. | Goodnough, LT | 1996 | Wrong intervention | |
| Eliminating needless testing in intensive care - An information-based team management approach | Roberts, DE | 1993 | Wrong intervention | |
| Evidence-based red cell transfusion in the critically ill: quality improvement using computerized physician order entry. | Rana, R | 2006 | Wrong intervention | |
| High-value care in the surgical intensive care unit: Effect on ancillary resources | Ko, A | 2016 | Wrong intervention | |
| Improving guideline compliance: assessment of unit-based reminder for monitoring platelet counts post-PCI...percutaneous coronary interventions | Belletti, D | 2002 | Wrong intervention | |
| Management of anaemia and blood transfusion in critical care - implementing national guidelines in ICU. | Watson, S | 2014 | Wrong intervention | |
| Multicenter implementation of a severe sepsis and septic shock treatment bundle | Miller III, RR | 2013 | Wrong intervention | |
| Multi-modality blood conservation strategy in open-heart surgery: an audit. | Reddy, SM | 2009 | Wrong intervention | |

| | | | | |
|---|---|---|---|---|
| Overutilization of serum electrolyte determinations in critical care units. Savings may be more apparent than real but what is real is of increasing importance. | Baigelman, W | 1985 | Wrong intervention | |
| Practice guideline for arterial blood gas measurement in the intensive care unit decreases numbers and increases appropriateness of tests | Pilon, CS | 1997 | Wrong intervention | |
| Quality in practice: Preventing and managing neonatal sepsis in Nicaragua | Lopez, S | 2013 | Wrong intervention | |
| Reducing blood testing in pediatric patients after heart surgery: A quality improvement project | Delgado-Corcoran, C | 2014 | Wrong intervention | |
| Reduction of hospital resources utilization in vascular surgery: A four- year experience | Roddy, SP | 1998 | Wrong intervention | |
| Results of a collaborative quality improvement program on outcomes and costs in a tertiary critical care unit | Clemmer, TP | 1999 | Wrong intervention | |
| The impact of peer management on test-ordering behavior. | Neilson, EG | 2004 | Wrong intervention | |
| The impact of selective laboratory evaluation on utilization of laboratory resources and patient care in a level-I trauma center | Chu, UB | 1996 | Wrong intervention | |
| Clinical Nurse Specialists Lead Teams to Impact Glycemic Control After Cardiac Surgery. | Klinkner, G | 2014 | Wrong intervention | |
| The effect of nurse champions on compliance with Keystone Intensive Care Unit Sepsis-screening protocol. | Campbell, J | 2008 | Wrong intervention | |
| Trauma case management: Improving patient outcomes | Curtis, K | 2006 | Wrong intervention | |
| A prospective one year study of massive blood transfusion in an intensive therapy unit | Das, SR | 1993 | Audit only (no feedback, or no data from after feedback) | 11 |

| | | | | |
|---|---|---|---|---|
| A QI project to reduce nosocomial blood loss | Andrews, JO | 1998 | Audit only (no feedback, or no data from after feedback) | |
| An audit of fresh frozen plasma transfusion in intensive care patients. | Gunawardana, RH | 1996 | Audit only (no feedback, or no data from after feedback) | |
| Appropriate use of blood component in pediatric patients in a Venezuelan General University Hospital: Cross-sectional study | Marti-Carvajal, AJ | 2005 | Audit only (no feedback, or no data from after feedback) | |
| Assessment of deep vein thrombosis prophylaxis in surgical patients: a study conducted at Nancy University Hospital, France. | Lepaux, DJ | 1998 | Audit only (no feedback, or no data from after feedback) | |
| Developing and Pilot Testing Quality Indicators in the Intensive Care Unit | Pronovost, PJ | 2003 | Audit only (no feedback, or no data from after feedback) | |
| Evaluation of the appropriateness of frozen plasma usage after introduction of prothrombin complex concentrates: a retrospective study. | Shih, AW | 2015 | Audit only (no feedback, or no data from after feedback) | |
| Retrospective audit of out-of-hours laboratory tests in an intensive care unit | Harris, CE | 1991 | Audit only (no feedback, or no data from after feedback) | |
| Use of the laboratory in a teaching hospital. Implications for patient care, education, and hospital costs. | Griner, PF | 1971 | Audit only (no feedback, or no data from after feedback) | |
| Laboratory utilization on a university surgical service | Liptzin, BA | 1972 | Audit only (no feedback, or no data from after feedback) | |
| Monitoring compliance with transfusion guidelines in hospital departments by electronic data capture | Norgaard, A | 2014 | Audit only (no feedback, or no data from after feedback) | |
| Acute care. Testing times for diagnostics | Johnson, P | 2013 | Wrong study design | 11 |
| Beyond the boundaries: a continuum of cardiac care. | Macready, N | 1997 | Wrong study design | |

| | | | | |
|---|---|---|---|---|
| Blood transfusion: Old blood, new blood or no blood | Duggan, JM | 2011 | Wrong study design | |
| Nurses and laboratory testing: New directions in POC | Blair, CH | 2004 | Wrong study design | |
| QA in transfusion services | Stugart, N | 1982 | Wrong study design | |
| Update on neonatal blood transfusions | Seidel, W | 1993 | Wrong study design | |
| Creation, implementation, and maturation of a massive transfusion protocol for the exsanguinating trauma patient. | Nunez, TC | 2010 | Wrong study design | |
| How we provide transfusion support for neonatal and pediatric patients on extracorporeal membrane oxygenation | Yuan, S | 2013 | Wrong study design | |
| Bloodstream infections, antibiotic resistance and the practice of blood culture sampling in Germany: study design of a Thuringia-wide prospective population-based study (AlertsNet). | Karch, A | 2015 | Wrong study design | |
| A high rate of compliance with neonatal intensive care unit transfusion guidelines persists even after a program to improve transfusion guideline compliance ended | Christensen, RD | 2011 | Wrong study design | |
| The Bloodwork Police | D'Angelo, C | 2001 | Wrong study design | |
| Learning to not know: results of a program for ancillary cost reduction in surgical critical care. | Barie, PS | 1996 | Feedback component not clear enough | 11 |
| Lessons Learned: Durability and Progress of a Program for Ancillary Cost Reduction in Surgical Critical Care | Barie, PS | 1997 | Feedback component not clear enough | |
| Maintaining Quality of Care While Reducing Charges in the ICU: Ten ways. | Civetta, JM | 1985 | Feedback component not clear enough | |
| Reducing neonatal transfusions | Batton, DG | 1992 | Feedback component not clear enough | |
| The effect of respiratory care department management of a blood gas analyzer on the | Beasley, KE | 1992 | Feedback component not clear enough | |

| | | | | |
|---|---|---|---|---|
| appropriateness of arterial blood gas utilization. | | | | |
| The Surviving Sepsis Campaign: results of an international guideline-based performance improvement program targeting severe sepsis. | Levy, MM | 2010 | Feedback component not clear enough | |
| Effect of laboratory testing guidelines on the utilization of tests and order entries in a surgical intensive care unit | Kumwilaisak, K | 2008 | Feedback component not clear enough | |
| Impact of clinical guidelines to improve appropriateness of laboratory tests and chest radiographs | Prat, G | 2009 | Feedback component not clear enough | |
| Abdominal aortic aneurysm pathway: outcome analysis. | Painter, LM | 1995 | Feedback component not clear enough | |
| Implementing a program to improve compliance with neonatal intensive care unit transfusion guidelines was accompanied by a reduction in transfusion rate: a pre-post analysis within a multihospital health care system | Baer, VL | 2011 | Feedback component not clear enough | |
| A computer based intervention on the appropriate use of arterial blood gas | Bansal, P | 2001 | Feedback component not clear enough | |
| Building a business case for colorectal surgery quality improvement | Lee, KKH | 2013 | Wrong outcomes | 5 |
| Effects of availability of patient-related charges on practice patterns and cost containment in the pediatric intensive care unit | Sachdeva, RC | 1996 | Wrong outcomes | |
| Evaluation and development of potentially better practices to prevent neonatal nosocomial bacteremia. | Kilbride, HW | 2003 | Wrong outcomes | |
| Introducing intensive insulin therapy: the nursing perspective. | Preston, S | 2006 | Wrong outcomes | |

| | | | | |
|---|---|---|---|---|
| Quality improvement report: Improving early management of bloodstream infection: A quality improvement project | Minton, J | 2008 | Wrong outcomes | |
| [Lessons and impact of two audits on postpartum hemorrhages in 24 maternity hospitals of the network "Securite Naissance - Naitre Ensemble" in "Pays-de-la-Loire" area]. | Branger, B | 2011 | Full text not in English | 3 |
| [Variables determining the amount of care for very preterm neonates: the concept of medical stance]. | Burguet, A | 2014 | Full text not in English | |
| Hospital use of fresh frozen plasma | Barbolla, L | 1997 | Full text not in English | |
| Nurse-led implementation of an insulin-infusion protocol in a general intensive care unit: improved glycaemic control with increased costs and risk of hypoglycaemia signals need for algorithm revision. | Alm-Kruse, K | 2008 | Wrong indication | 3 |
| Reductions in invasive device use and care costs after institution of a daily safety checklist in a pediatric critical care unit. | Tarrago, R | 2014 | Wrong indication | |
| Blood wastage reduction: a 10-year observational evaluation in a large teaching institution in France. | Zoric, L | 2013 | Wrong indication | |
| Implementing surviving sepsis guidelines in a district general hospital | Page, I | 2011 | Wrong setting | 1 |
| Strategies for success: a PDSA analysis of three QI initiatives in critical care. | Lipshutz, AKM | 2008 | Combination of reasons | 1 |

**APPENDIX I: Quality Assessment Inter-Rater Reliability**

| Overall Agreement | | MF | | | | |
|---|---|---|---|---|---|---|
| | | Yes | No | Unclear | N/A | Total |
| NM | Yes | 36 | 6 | 7 | 0 | 49 |
| | No | 7 | 57 | 1 | 0 | 65 |
| | Unclear | 11 | 1 | 12 | 0 | 24 |
| | N/A | 0 | 1 | 0 | 14 | 15 |
| | Total | 54 | 65 | 20 | 14 | 153 |

**Total Agreement (n) =** 36 + 57 + 12 + 14 = 119

**Total Agreement (%) =** (119/153) x100= 77.8%

| Code | Calculation | Expected Frequency |
|---|---|---|
| Yes | (54 x 49)/153 | 17.3 |
| No | (65 x 65)/153 | 27.6 |
| Unclear | (20 x 24)/153 | 3.1 |
| N/A | (14 x 15)/153 | 1.4 |
| Sum | 17.3+27.6+3.1+1.4 | 49.4 |

**Cohen's Kappa**

K = (Σ agreement -Σ expected frequency)/(N-Σ expected frequency)

K= (119 – 49.4)/(153-49.4)

K= 0.67

## REFERENCES


1. Cismondi F, Celi LA, Fialho AS, et al. Reducing Unnecessary Lab Testing in the ICU with Artificial Intelligence. *Int J Med Inform*. 2013;82(5):345-358. doi:10.1016/j.ijmedinf.2012.11.017.

2. Merkeley HL, Hemmett J, Cessford TA, et al. Multipronged strategy to reduce routine-priority blood testing in intensive care unit patients. *J Crit Care*. 2016;31(1):212-216. doi:10.1016/j.jcrc.2015.09.013.

3. Raad S, Elliott R, Dickerson E, Khan B, Diab K. Reduction of Laboratory Utilization in the Intensive Care Unit. *J Intensive Care Med*. 2016:885066616651806. doi:10.1177/0885066616651806.

4. Kleinpell RM, Farmer JC, Pastores SM. Reducing Unnecessary Testing in the Intensive Care Unit by Choosing Wisely. *Acute Crit Care*. 2018;33(1):1-6. doi:10.4266/acc.2018.00052.

5. Barie PS. Phlebotomy in the Intensive Care Unit: Strategies for Blood Conservation. *Crit Care*. 2004;8 Suppl 2:S34-6. doi:10.1186/cc2454.

6. Kotecha N, Shapiro JM, Cardasis J, Narayanswami G. Reducing Unnecessary Laboratory Testing in the Medical ICU. *Am J Med*. 2017;130(6):648-651. doi:10.1016/j.amjmed.2017.02.014.

7. Melanson SEF, Szymanski T, Rogers SO, et al. Utilization of Arterial Blood Gas Measurements in a Large Tertiary Care Hospital. *Am J Clin Pathol*. 2007;127(4):604-609. doi:10.1309/ELH5BPQ0T17RRK0M.

8. Smoller BR, Kruskall MS. Phlebotomy for Diagnostic Laboratory Tests in Adults: Pattern of Use and Effect on Transfusion Requirements. *N Engl J Med*. 1986;314(19):1233-1235.

9. Ezzie ME, Aberegg SK, O'Brien JM. Laboratory Testing in the Intensive Care Unit. *Crit Care Clin*. 2007;23(3):435-465. doi:10.1016/j.ccc.2007.07.005.

10. McEvoy MT, Shander A. Anemia, Bleeding, and Blood Transfusion in the Intensive Care Unit: Causes, Risks, Costs, and New Strategies. *Am J Crit Care*. 2013;22(6 Suppl):eS1-13. doi:10.4037/ajcc2013729.

11. Carson JL, Stanworth SJ, Roubinian N, et al. Transfusion Thresholds and Other Strategies for Guiding Allogeneic Red Blood Cell Transfusion. *Cochrane Database Syst Rev*. 2016;(10):CD002042. doi:10.1002/14651858.CD002042.pub4.

12. Verbrugghe S. Initiatives to Optimize the Utilization of Laboratory Tests. Ottawa: Canadian Agency for Drugs and Technolgies in Health. 2014;(Environmental Scan; issue 44).

13. Critical Care Societies Collaborative - Critical Care. Five Things Physicians and Patients Should Question. Choosing Wisely. http://www.choosingwisely.org/societies/critical-care-societies-collaborative-critical-care/. Published 2014. Accessed April 23, 2017.

14. Zhi M, Ding EL, Theisen-Toupal J, Whelan J, Arnaout R. The Landscape of Inappropriate Laboratory Testing: A 15-year Meta-Analysis. *PLoS One*. 2013;8(11):1-8. doi:10.1371/journal.pone.0078962.

15. Australian Association of Pathology Practices Inc. *An Analysis of Pathology Test Use in Australia*. Family Medicine Research Centre, University of Sydney; 2011.

16. Cadogan SL, Browne JP, Bradley CP, Cahill MR. The effectiveness of interventions to improve laboratory requesting patterns among primary care physicians: a systematic review. *Implement Sci*. 2015;10:167. doi:10.1186/s13012-015-0356-4.

17. Lewandrowski K, Baron J, Dighe A. Utilization Management in the Clinical Laboratory: An Introduction and Overview. In: Lewandrowski K, Sluss P, eds. *Utilization Management in the Clinical Laboratory and Other Ancillary Services*. Springer, Cham; 2017.

18. Koch CG, Li L, Sun Z, et al. Hospital-Acquired Anemia: Prevalence, Outcomes, and Healthcare Implications. *J Hosp Med*. 2013;8(9):506-512. doi:10.1002/jhm.2061.

19. Salisbury AC, Amin AP, Reid KJ, et al. Hospital-acquired anemia and in-hospital mortality in patients with acute myocardial infarction. *Am Heart J*. 2011;162(2):300-309.e3. doi:10.1016/j.ahj.2011.05.021.

20. Salisbury AC, Reid KJ, Alexander KP, et al. Diagnostic Blood Loss From Phlebotomy and Hospital-Acquired Anemia During Acute Myocardial Infarction. *Arch Intern Med*. 2011;171(18):1646-1653. doi:10.1001/archinternmed.2011.361.

21. Borgert M, Binnekade J, Paulus F, Goossens A, Vroom M, Dongelmans D. Timely individual audit and feedback significantly improves transfusion bundle compliance—a comparative study. *Int J Qual Heal Care*. 2016;28(5):601-607. doi:10.1093/intqhc/mzw071.

22. Mikhaeil M, Day AG, Ilan R. Non-essential blood tests in the intensive care unit: a prospective observational study. *Can J Anesth Can d'Anesthésie*. 2017;64(3):290-295. doi:10.1007/s12630-016-0793-9.

23. Ivers NM, Desveaux L, Presseau J, et al. Testing feedback message framing and comparators to address prescribing of high-risk medications in nursing homes: protocol for a pragmatic, factorial, cluster-randomized trial. *Implement Sci*. 2017;12(1):86. doi:10.1186/s13012-017-0615-7.

24. Brehaut JC, Colquhoun HL, Eva KW, et al. Practice Feedback Interventions: 15 Suggestions for Optimizing Effectiveness. *Ann Intern Med*. 2016;164(6):435-441. doi:10.7326/M15-2248.

25. Ivers N, Jamtvedt G, Flottorp S, et al. Audit and Feedback : Effects on Professional Practice and Healthcare Outcomes ( Review ). *Cochrane Database Syst Rev*. 2012;(6):1-227. doi:10.1002/14651858.CD000259.pub3.

26. Lapinsky SE, Holt D, Hallett D, Abdolell M, Adhikari NKJ. Survey of information technology in Intensive Care Units in Ontario, Canada. *BMC Med Inform Decis Mak*. 2008;8:1-6. doi:10.1186/1472-6947-8-5.

27. Patel S, Rajkomar A, Harrison JD, et al. Next-generation audit and feedback for inpatient quality improvement using electronic health record data: a cluster randomised controlled trial. *BMJ Qual Saf*. 2018:[Epub ahead of print]. doi:10.1136/bmjqs-2017-007393.

28. Pendry K. The use of big data in transfusion medicine. *Transfus Med*. 2015;25(3):129-137.

29. Hysong ST. Meta-Analysis: Audit and Feedback Features Impact Effectiveness on Care Quality. *Med Care*. 2009;47(3):356-363.

30. Kobewka DM, Ronksley PE, McKay JA, Forster AJ, Van Walraven C. Influence of educational, audit and feedback, system based, and incentive and penalty interventions to reduce laboratory test utilization: A systematic review. *Clin Chem Lab Med*. 2015;53(2):157-183. doi:10.1515/cclm-

2014-0778.

31. Shamseer L, Moher D, Clarke M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. *BMJ*. 2015;350:g7647. doi:10.1136/bmj.g7647.

32. Centre for Reviews and Dissemination- University of York. PROSPERO International Prospoective Register for Systematic Reviews- Guidance notes for registering a systematic review protocol with PROSPERO. 2016:1-23. www.crd.york.ac.uk/prospero.

33. Foster M, Presseau J, Carroll K, Brehaut J. Evaluation of Feedback Interventions for the Reduction of Inappropriate Laboratory Tests and Transfusions in Intensive Care Units: A Systematic Review Protocol. PROSPERO. http://www.crd.york.ac.uk/PROSPERO/display_record.asp?ID=CRD42016051941. Published 2016.

34. Cochrane Effective Practice and Organisation of Care Review Group (EPOC). *Data Collection Checklist*. Ottawa; 2002.

35. Morrison A, Polisena J, Husereau D, et al. The Effect of English-Language Restriction on Systematic Review-Based Meta-Analyses: A Systematic Review of Empirical Studies. *Int J Technol Assess Health Care*. 2012;28(2):138-144. doi:10.1017/s0266462312000086.

36. Spencer AJ, Eldredge JD. Roles for librarians in systematic reviews: a scoping review. *J Med Libr Assoc*. 2018;106(1):46-56. doi:10.5195/JMLA.2018.82.

37. Institute of Medicine (US) Committee on Standards for Systematic Reviews of Comparative Effectiveness Research. Standards for Finding and Assessing Individual Studies. In: Eden J, Levit L, Berg A, Morton S, eds. *Finding What Works in Health Care: Standards for Systematic Reviews*. Washington, DC: National Academies Press (US); 2011. doi:10.17226/13059.

38. CRD (Centre for Reviews and Dissemination). *Systematic Reviews: CRD's Guidance for Undertaking Reviews in Health Care*. York, UK: York Publishing Serivces, Ltd.; 2009. doi:10.1016/S1473-3099(10)70065-7.

39. Tinmouth A, Macdougall L, Fergusson D, et al. Reducing the Amount of Blood Transfused: A Systematic Review of Behavioral Interventions to Change Physicians' Transfusion Practices. *Arch Intern Med*. 2005;165(8):845-852. doi:10.1001/archinte.165.8.845.

40. Damiani G, Pinnarelli L, Sommella L, et al. Appropriateness of fresh-frozen plasma usage in hospital settings: a meta-analysis of the impact of organizational interventions. *Transfusion*. 2010;50(1):139-144. doi:10.1111/j.1537-2995.2009.02371.x.

41. Wilson K, MacDougall L, Fergusson D, Graham I, Tinmouth A, Hébert PC. The effectiveness of interventions to reduce physician's levels of inappropriate transfusion: what can be learned from a systematic review of the literature. *Transfusion*. 2002;42(9):1224-1229. doi:10.1046/j.1537-2995.2002.00185.x.

42. Vertitas Health Innovation Ltd. Covidence. https://www.covidence.org/. Accessed October 19, 2016.

43. Higgins J, Altman D, Sterne J (editors). Chapter 8: Assessing risk of bias in included studies. In: Higgins J, Green S, eds. *Cochrane Handbook for Systematic Reviews of Interventions*. Version 5.

1.0 (updated March 2011). The Cochrane Collaboration; 2011. Available from www.handbook.cochrane.org.

44. Hallgren KA. Computing Inter-Rater Reliability for Observational Data: An Overview and Tutorial. *Tutor Quant Methods Psychol*. 2012;8(1):23-34. doi:10.20982/tqmp.08.1.p023.

45. Moher D, Liberati A, Tetzlaff J, Altman D, The PRISMA Group. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLoS Med*. 2009;6(7):e1000097. doi:10.1371/journal.pmed1000097.

46. Solomon DH, Hashimoto H, Daltroy L, Liang MH. Techniques to Improve Physicians' Use of Diagnostic Tests: A New Conceptual Framework. *Jama*. 1998;280(23):2020-2027. doi:10.1001/jama.280.23.2020.

47. Paes BA, Modi A, Dunmore R. Changing physicians' behavior using combined strategies and an evidence-based protocol. *Arch Pediatr Adolesc Med*. 1994;148(12):1277-1280. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med3&NEWS=N&AN=7951806.

48. Hendryx MS, Fieselmann JF, Bock MJ, Wakefield DS, Helms CM, Bentler SE. Outreach Education to Improve Quality of Rural ICU Care: Results of a Randomized Trial. *Am J Respir Crit Care Med*. 1998;158(2):418-423.

49. Merlani P, Garnerin P, Diby M, Ferring M, Ricou B. Quality Improvement Report: Linking Guideline To Regular Feedback To Increase Appropriate Requests For Clinical Tests: Blood Gas Analysis In Intensive Care. *BMJ*. 2001;323(7313):620-624. doi:10.1136/bmj.323.7313.620.

50. Diby M, Merlani P, Garnerin P, Ricou B. Harmonization of practice among different groups of caregivers: a guideline on arterial blood gas utilization. *J Nurs Care Qual*. 2005;20(4):327-334. http://search.ebscohost.com/login.aspx?direct=true&db=cin20&AN=106540547&site=ehost-live.

51. Beland D, D'Angelo C, Vinci D. Reducing unnecessary blood work in the neurosurgical ICU. *J Neurosci Nurs*. 2003;35(3):149-152. http://search.ebscohost.com/login.aspx?direct=true&db=cin20&AN=106867699&site=ehost-live.

52. Wisser D, Van Ackern K, Knoll E, Wisser H, Bertsch T. Blood Loss from Laboratory Tests. *Clin Chem*. 2003;49(10):1651-1655.

53. Petäjä J, Andersson S, Syrjälä M. A simple automatized audit system for following and managing practices of platelet and plasma transfusions in a neonatal intensive care unit. *Transfus Med*. 2004;14(4):281-288. doi:10.1111/j.0958-7578.2004.00515.x.

54. Calderon-Margalit R, Mor-Yosef S, Mayer M, Adler B, Shapira SC. An administrative intervention to improve the utilization of laboratory tests within a university hospital. *Int J Qual Heal Care*. 2005;17(3):243-248. doi:10.1093/intqhc/mzi025.

55. Schramm GE, Kashyap R, Mullon JJ, Gajic O, Afessa B. Septic shock: A multidisciplinary response team and weekly feedback to clinicians improve the process of care and mortality. *Crit Care Med*. 2011;39(2):252-258. doi:10.1097/CCM.0b013e3181ffde08.

56. Masud F, Larson-Pollock K, Leveque C, Vykoukal D. Establishing a Culture of Blood Management Through Education: A Quality Initiative Study of Postoperative Blood Use in CABG Patients at Methodist DeBakey Heart & Vascular Center. *Am J Med Qual*. 2011;26(5):349-356.

http://search.ebscohost.com/login.aspx?direct=true&db=cin20&AN=104577658&site=ehost-live.

57. Arnold DM, Lauzier F, Whittingham H, et al. A multifaceted strategy to reduce inappropriate use of frozen plasma transfusions in the intensive care unit. *J Crit Care*. 2011;26(6):636.e7-636.e13. doi:10.1016/j.jcrc.2011.02.005.

58. Beaty CA, Haggerty KA, Moser MG, et al. Disclosure of Physician-Specific Behavior Improves Blood Utilization Protocol Adherence in Cardiac Surgery. *Ann Thorac Surg*. 2013;96(6):2168-2174. doi:10.1016/j.athoracsur.2013.06.080.

59. Gutsche JT, Kornfield ZN, Speck RM, Patel PA, Atluri P, Augoustides JG. Impact of Guideline Implementation on Transfusion Practices in a Surgical Intensive Care Unit. *J Cardiothorac Vasc Anesth*. 2013;27(6):1189-1193. doi:10.1053/j.jvca.2013.05.040.

60. Yeh DD, Naraghi L, Larentzakis A, et al. Peer-to-peer physician feedback improves adherence to blood transfusion guidelines in the surgical intensive care unit. *J Trauma Acute Care Surg*. 2015;79(1):65-70. doi:10.1097/TA.0000000000000683.

61. Murphy DJ, Lyu PF, Gregg SR, et al. Using Incentives to Improve Resource Utilization: A Quazi-Experimental Evaluation of an ICU Quality Improvement Program. *Crit Care Med*. 2016;44(1):162-170. doi:10.1097/ccm.0000000000001395.

62. Arnold DM, Lauzier F, Whittingham H, et al. A multifaceted strategy to reduce inappropriate use of frozen plasma transfusions in the intensive care unit. *J Crit Care*. 2011;26(6):636.e7-636.e13. doi:10.1016/j.jcrc.2011.02.005.

63. Solomon RR, Clifford JS, Gutman SI. The Use of Laboratory Intervention to Stem the Flow of Fresh-Frozen Plasma. *Am J Clin Pathol*. 1988;89(4):518-521.

64. Pendry K. The use of big data in transfusion medicine. *Transfus Med*. 2015;25(3):129-137. doi:10.1111/tme.12223.

65. Grol R, Grimshaw J. From best evidence to best practice: Effective implementation of change in patients' care. *Lancet*. 2003;362(9391):1225-1230. doi:10.1016/S0140-6736(03)14546-1.

66. Greco PJ, Eisenberg JM. Changing Physicians' Practices. *N Engl J Med*. 1993;329(17):1271-1274.

67. Grol R. Implementing guidelines in general practice care. *Qual Heal Care*. 1992;1(3):184-191. doi:10.1136/qshc.1.3.184.

68. Ivers NM, Grimshaw JM. Reducing research waste with implementation laboratories. *Lancet*. 2016;388(10044):547-548. doi:10.1016/S0140-6736(16)31256-9.

69. Ivers NM, Grimshaw JM, Jamtvedt G, et al. Growing Literature, Stagnant Science? Systematic Review, Meta-Regression and Cumulative Analysis of Audit and Feedback Interventions in Health Care. *J Gen Intern Med*. 2014;29(11):1534-1541. doi:10.1007/s11606-014-2913-y.

70. Tate DF, Lytle LA, Sherwood NE, et al. Deconstructing interventions: approaches to studying behavior change techniques across obesity interventions. *Transl Behav Med*. 2016;6(2):236-243. doi:10.1007/s13142-015-0369-1.

71. Craig P, Dieppe P, Macintyre S, Mitchie S, Nazareth I, Petticrew M. Developing and evaluating complex interventions: The new Medical Research Council guidance. *BMJ*. 2008;337(a1655). doi:10.1136/bmj.a1655.

72. Bjurling-Sjöberg P, Wadensten B, Pöder U, Jansson I, Nordgren L. Balancing intertwined responsibilities: A grounded theory study of teamwork in everyday intensive care unit practice. *J Interprof Care*. 2017;31(2):233-244. doi:10.1080/13561820.2016.1255184.

73. Alameddine M, Dainty KN, Deber R, Sibbald WJ (Bill). The intensive care unit work environment: Current challenges and recommendations for the future. *J Crit Care*. 2009;24(2):243-248. doi:10.1016/j.jcrc.2008.03.038.

74. Goodacre S. Uncontrolled before-after studies: Discouraged by Cochrane and the EMJ. *Emerg Med J*. 2015;32(7):507-508. https://doi.org/10.1136/emermed-2015-204761.

75. Eccles M, Grimshaw J, Campbell M, Ramsay C. Research designs for studies evaluating the effectiveness of change and improvement strategies. *Qual Saf Heal Care*. 2003;12(1):47-52. http://qualitysafety.bmj.com/content/qhc/12/1/47.full.pdf%0Ahttp://www.ncbi.nlm.nih.gov/pubmed/12571345.

76. Verstappen WHJM, Van Der Weijden T, Ter Riet G, Grimshaw J, Winkens R, Grol RPTM. Block design allowed for control of the Hawthorne effect in a randomized controlled trial of test ordering. *J Clin Epidemiol*. 2004;57(11):1119-1123. doi:10.1016/j.jclinepi.2004.03.009.

77. Effective Practice and Organisation of Care (EPOC). EPOC Taxonomy; 2015. Available at: https://epoc.cochrane.org/epoc-taxonomy.

78. Powell BJ, Waltz TJ, Chinman MJ, et al. A refined compilation of implementation strategies: Results from the Expert Recommendations for Implementing Change (ERIC) project. *Implement Sci*. 2015;10(21). doi:10.1186/s13012-015-0209-1.

79. Hoffmann TC, Glasziou PP, Boutron I, et al. Better reporting of interventions: Template for intervention description and replication (TIDieR) checklist and guide. *BMJ*. 2014;348:g1687. doi:10.1136/bmj.g1687.

80. Kanter M, Courneya PT. Perspective on Publishing Quality Improvement Efforts. *Perm J*. 2017;21:17-140. doi:10.7812/TPP/17-140.

81. Hempel S, Rubenstein L V., Shanman RM, et al. Identifying quality improvement intervention publications - A comparison of electronic search strategies. *Implement Sci*. 2011;6(85). doi:10.1186/1748-5908-6-85.

82. Sterne J, Egger M, Moher D (editors). Chapter 10: Addressing reporting biases. In: Higgins J, Green S, eds. *Cochrane Handbook for Systematic Reviews of Intervention. Version 5.1.0 (Updated March 2011)*. The Cochrane Collaboration; 2011. Available from www.handbook.cochrane.org.

83. Gude WT, Roos-Blom M-J, van der Veer SN, et al. Health professionals' perceptions about their clinical performance and the influence of audit and feedback on their intentions to improve practice: a theory-based study in Dutch intensive care units. *Implement Sci*. 2018;13(1):33. doi:10.1186/s13012-018-0727-8.

84. The Nordic Cochrane Centre- The Cochrane Collaboration. Review Manager (RevMan). 2014.

85. BMJ Clinical Evidence. What is GRADE? The BMJ Clinical Evidence Website. http://clinicalevidence.bmj.com/x/set/static/ebm/learn/665072.html. Published 2012. Accessed October 19, 2016.

86. Thomas RE, Croal BL, Ramsay C, Eccles M, Grimshaw J. Effect of Enhanced Feedback and Brief

Educational Reminder Messages on Laboratory Test Requesting in Primary Care: a Cluster Randomised Trial. *Lancet*. 2006;367(9527):1990-1996. doi:10.1016/S0140-6736(06)68888-0.

87. Bates DW, Kuperman GJ, Rittenberg E, et al. A Randomized Trial of a Computer-based Intervention to Reduce Utilization of Redundant Laboratory Tests. *Am J Med*. 1999;106(2):144-150. doi:10.1016/S0002-9343(98)00410-0.

88. Gortmaker SL, Bickford AF, Mathewson HO, Dumbaugh K, Tirrell PC. A Successful Experiment to Reduce Unnecessary Laboratory Use in a Community Hospital. *Med Care*. 1988;26(6):631-642.

89. Chen P, Tanasijevic MJ, Schoenenberger RA, Fiskio J, Kuperman GJ, Bates DW. A Computer-Based Intervention for Improving the Appropriateness of Antiepileptic Drug Level Monitoring. *Am J Clin Pathol*. 2003;119(3):432-438. doi:10.1309/A96xu9yku298hb2r.

## Chapter 3 (Draft Component Manuscript 2):

## How Well Do Critical Care Audit & Feedback Interventions Adhere to Best Practice? Development and Application of an Evaluation Tool

### PREFACE

The study described in Chapter 3 addresses two main objectives:

1. The development and piloting of an evaluation tool based on recent guidance for best practice in the design of A&F interventions.[1]
2. Application of this tool to a sample of A&F intervention studies identified through systematic review, to assess whether existing A&F interventions in the critical care setting comply with these suggestions for best practice.

**Contributions: MF** and **KC** conducted screening for the systematic review, to identify the sample of A&F interventions. **JCB** and **JP** provided guidance throughout this process and resolved conflicts between reviewers. All members of the study team were involved in the development of the evaluation tool (**JCB, JP, KC, MP, MF**). **MF** and **MP** piloted the evaluation tool. Formal assessment of the identified studies was completed by **MF** with guidance from **JCB** and **JP**. **MF** drafted the manuscript and **JCB** and **JP** edited and made contributions to the manuscript.

**Appendices:** Appendices for documented changes to the protocol, the PRISMA flow diagram (reprint[2]), the evaluation tool criteria items, response scales and anchors, and the Research Ethics Board approval letter can be found at the end of this chapter.

**Ethics Approval**: This study was approved by the Ottawa Health Science Network Research Ethics Board (OHSN-REB; Protocol ID: 20160951-01H).

**Publication Citation:**
Foster M, Presseau J, Carroll K, Papoulias M, Brehaut J. "How Well Do A&F Interventions Adhere to Best Practice? An Evaluation Tool." (In preparation).

**How Well Do Critical Care Audit & Feedback Interventions Adhere to Best Practice? An Evaluation Tool**

Madison Foster, MSc (candidate)[1,2*], Justin Presseau, PhD[1,2,3], Kelly Carroll, MA,[2] Maria Papoulias,[2] Lauralyn McIntyre[1,4], Jamie Brehaut, PhD[1,2]

[1]School of Epidemiology and Public Health, University of Ottawa, 451 Smyth Road, Ottawa, Ontario, K1H 8M5, Canada

[2]Ottawa Hospital Research Institute, Clinical Epidemiology Program, The Ottawa Hospital, General Campus, 501 Smyth Road, Centre for Practice Changing Research, Box 201B, Ottawa, Ontario, K1H 8L6, Canada

[3] School of Psychology, University of Ottawa, 136 Jean-Jacques Lussier, Vanier Hall, Ottawa, Ontario, K1N 6N5, Canada

[4]Department of Critical Care Medicine, The Ottawa Hospital, General Campus, 501 Smyth Road, Ottawa, Ontario, K1H 8L6, Canada

***Corresponding Author:** Madison Foster (madfoster@ohri.ca), C/O Kelly Carroll, The Ottawa Hospital, General Campus, 501 Smyth Road, Centre for Practice Changing Research, Box 201B, Ottawa, Ontario, K1H 8L6, Canada

**ABSTRACT**

**Background:** Healthcare Audit and Feedback (A&F) interventions have been shown to be an effective means of changing healthcare professional behaviour. However, further work is required to optimize this class of intervention, as efforts to date have not resulted in improved A&F. Recent published suggestions have provided best practices for a series of modifiable A&F elements, which may help to increase intervention effectiveness. We aimed to develop a generalizable evaluation tool that can be used to assess whether A&F interventions conform to these suggestions for best practice and to apply the tool to a sample of critical care A&F studies.

**Methods:** The 15 suggestions for improved feedback interventions published by Brehaut et al. were deconstructed into 52 individual rateable items. Items were developed through iterative consensus meetings among researchers. These items were then piloted on 12 A&F reports (three rounds, four interventions per round). After each round of consensus, items were modified to improve clarity and specificity, to help increase the reliability between coders, and the coding manual was updated accordingly. One reviewer then used the tool to rate critical care A&F interventions to assess conformity to best practices. This sample of interventions was obtained from a previously completed systematic review, which had identified studies evaluating the effect of A&F interventions on provider ordering of laboratory tests and transfusions as compared to other interventions or standard care, in the critical care setting.

**Results:** The evaluation tool was applied to 17 A&F interventions implemented in critical care settings. Several interventions only provided feedback once. Studies rarely piloted feedback forms, involved stakeholders during development, or conducted barrier or engagement assessments. Reporting of the majority of our key intervention details was limited.

**Conclusions:** Many of the theory-informed best practice suggestions were not consistently applied in this sample. Standardized reporting and posting of detailed intervention descriptions and feedback templates may help to further advance research in this field. Future research to test whether the potentially

underutilized suggestions may improve intervention effectiveness, as well as to study their mechanisms of action, will help to optimize A&F in the critical care setting.

**Registration Number:** Not applicable.

**Keywords:** Audit, Feedback, A&F, optimization, critical care

## BACKGROUND

Auditing healthcare provider behaviour and then feeding these data back to individuals or groups (often with comparators and guidance on appropriate behaviour) is a common intervention.[3] Systematic reviews suggest that optimization of these audit and feedback interventions (A&F) is required.[3,4] Despite clear evidence that A&F usually has a positive effect on behaviour (1.3% [continuous outcomes]; 4.3% [dichotomous outcomes]), effect sizes across trials of such intervention are variable, ranging from relatively large to null or even negative effects.[3] This variation has important implications. In some cases, it is possible for A&F to result in more harm than benefit; if A&F is not executed in an effective way, providers' performance (and thus the care received by patients) may be negatively impacted. Moreover, resources used to conduct the audit and develop and disseminate the feedback can be wasted. On the other hand, A&F can produce quite large improvements in care. However, previous analysis has shown feedback interventions to date have not improved.[4]

To provide feedback developers with practical guidance, a recent publication summarized suggestions for the optimization of A&F. This work compiled lessons from interviews with experts in A&F theory, as well as research team experience, to produce 15 theory-informed suggestions for high quality A&F interventions (**Table 1**).[1] These suggestions focus on easily modifiable elements of A&F (target action, data, design and delivery factors) proposed to improve effectiveness. The extent to which these suggestions are already followed in the literature is unclear; we propose to explore the degree to which these suggestions have been applied in a sample of existing A&F interventions.

A&F may be a particularly well-suited intervention to change behaviour in complex environments such as the intensive care unit (ICU). A&F's flexibility allows for provision of feedback to individuals and/or groups, through a variety of modalities, which may be useful in addressing the team-based and multidisciplinary nature[5,6] of the critical care setting. Furthermore, behaviours within critical care (such as frequent test ordering) may become routine (i.e. a "behaviour [that] is repeated in the same context").[7] Providing performance data on routinized behaviours may help to highlight the frequency with which these orders are placed and flag them as an area requiring improvement.

We recently conducted a systematic review[2] of the use and effectiveness of A&F interventions in critical care. In the current study, we sought to assess the extent to which identified interventions were consistent with Brehaut et al.'s 15 suggestions.[1] As the suggestions were designed to provide general guidance to feedback developers, rather than serve as prescriptive instructions, each one encompasses multiple concepts and may be applied in a variety of different ways; it would thus be difficult to reliably rate consistency with these multidimensional suggestions. To better assess how existing A&F interventions may be 'consistent' with this guidance, we created an evaluation tool by deconstructing each of the 15 suggestions into specific, rateable items. Here we report the development and application of this evaluation tool to a sample of

intensive care unit A&F interventions, to identify priority areas for future A&F research in the critical care setting.

**Table 1. Brehaut and colleagues' 15 suggestions for improved audit and feedback interventions**[1]

| Nature of the Desired Action | Nature of the Data Available for Feedback |
|---|---|
| 1. Recommend actions that are consistent with established goals and priorities.<br><br>2. Recommend actions that can improve and are under the recipient's control.<br><br>3. Recommend specific actions. | 4. Provide multiple instances of feedback.<br><br>5. Provide feedback as soon as possible and at a frequency informed by the number of new patient cases.<br><br>6. Provide individual rather than general data.<br><br>7. Choose comparators that reinforce desired behaviour change. |
| **Feedback Display** | **Delivering the Feedback Intervention** |
| 8. Closely link the visual display and summary message.<br><br>9. Provide feedback in more than one way.<br><br>10. Minimize extraneous cognitive load for feedback recipients. | 11. Address barriers to feedback use.<br><br>12. Provide short, actionable messages followed by optional detail.<br><br>13. Address credibility of the information.<br><br>14. Prevent defensive reactions to feedback.<br><br>15.Construct feedback through social interaction. |

**Objectives:**

The main objectives of our study were to:

1) Develop and pilot an evaluation tool based on Brehaut and colleagues' 15 suggestions for improving feedback interventions.[1]

2) Assess consistency of existing critical care A&F interventions with this guidance[1] for best practice.

**METHODS**

**Identification and Collection of Study Materials**

Studies evaluating A&F interventions, in the context of improving laboratory test and transfusion (red blood cell, platelet, plasma, cryoprecipitate) ordering in an intensive care setting were previously identified through systematic review.[2] The review was conducted to summarize the current evidence on the use of A&F for quality improvement of lab test and transfusion ordering decisions in critical care. The review was

registered with the International Prospective Register of Systematic Reviews (PROSPERO),[8] and the methods were reported in detail according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA)[9] in a separate publication (in preparation).[2] In brief, any empirical studies involving quantitative data were included. Sixteen studies (17 publications) were identified; six of these studies aimed to improve lab test ordering,[10–16] eight aimed to improve transfusion ordering,[17–24] and two assessed both behaviours.[25,26] Corresponding authors were contacted by email. We requested an example of the feedback form used (de-identified) as well as any other pertinent information regarding the study. If no response was received within two weeks, a follow-up e-mail was sent.

**Item Development**

As the guidance published by Brehaut et al. aims to provide general suggestions for feedback developers, several suggestions encompass multiple concepts. Furthermore, the suggestions may be operationalized in different ways. To better assess how existing A&F interventions adhere to the suggestions, we deconstructed each one into specific, ratable items. These items were designed to capture details regarding elements of A&F interventions hypothesized to affect effectiveness.

Items were developed by the research team (JCB, JP, MF, KC, MP). First, suggestions encompassing more than one distinct concept were split into multiple items to improve ease and reliability of evaluation. Next, suggestions were operationalized and re-phrased into items that could allow rating of individual A&F intervention studies. Items were deliberately phrased to be generalizable across a range of different A&F interventions. This was an iterative process, whereby items were discussed amongst members of the team until consensus was reached that the items fully represented the key components of each suggestion. Items were then de-duplicated, and any items that were judged unrateable were removed. Response scales (i.e. Yes/No/Unclear) were also developed with individualized anchors to increase inter-rater reliability.

**Evaluation Tool Piloting**

A sample of feedback interventions was selected from the 2012 A&F Cochrane Review[3] (non-ICU, and not necessarily focussed on test or transfusion ordering) to pilot the evaluation criteria. Each sample intervention was independently rated with the pilot criteria by two raters (MF and MP). Consensus meetings were held to compare the results between raters, discuss discrepancies and modify items as necessary to improve their clarity, rateability and mutual exclusivity. This was an iterative process requiring multiple consensus meetings. Three rounds of consensus were completed, with four studies rated per round. Inter-rater reliability was measured by calculating Cohen's Kappa for the 39 items which used a 'Yes/No/Unclear/Not Applicable' scale.[27] The pilot study concluded once all ambiguities had been clarified and the research team agreed that the criteria items comprehensively covered all 15 suggestions.[1]

**Data Extraction**

After development and pilot testing, we used the evaluation tool to assess a sample of 17 A&F interventions (from 16 studies)[10–26] identified by our previous systematic review.[2] One reviewer (MF) extracted data both from published reports and, when provided by authors, the sample feedback forms.

### Analysis

Simple counts and percentages describing the number of articles consistent with each item were determined. Due to the relatively small sample size (n=17 feedback interventions), originally planned analyses to determine whether consistency with each criteria item varies according to different study characteristics, were not completed. Gaps in the current literature were identified and discussed narratively.

### Ethics

This study was approved by the Ottawa Health Science Network Research Ethics Board (OHSN-REB; Protocol ID: 20160951-01H). Any changes from the original study protocol were minor and have been outlined in **Appendix A**.

## RESULTS

### *Identification of Studies and Collection of Study Materials*

Sixteen studies (17 publications) describing 17 feedback interventions (one trial compared two different types of feedback) were identified by systematic review.[2] **Appendix B** provides a flow diagram of the study selection. Corresponding authors from the 17 publications were contacted to request feedback form examples. Four authors were able to provide an example of the feedback form utilized in the trial and two responded but were unable to provide further details (response rate=35%; feedback forms=5/17).[A] Two additional studies provided an example of the feedback form in the publication (total number of feedback forms=7/17).

### *Sample of A&F Interventions*

**Table 2** (reproduced from the systematic review[2]) describes the sample of A&F interventions assessed in this study. The review[2] identified a heterogeneous sample of multicomponent quality improvement interventions involving A&F; eight aimed to improve lab test ordering and ten aimed to improve transfusion ordering (two aimed to improve both practices). Fifteen of the 17 interventions incorporated one or more additional components, such as education, guidelines, opinion leaders, financial incentives, checklists or administrative interventions. The plurality of interventions reported providing feedback more than once (53%), in only a written format (41%), with data aggregated at the group level only (41%). Feedback was most often provided to multiple groups of healthcare providers (29%) or physicians only (24%). Heterogeneity of the outcomes precluded meta-analysis, however, overall the majority of interventions reported statistically significant behaviour changes in the hypothesized direction. Most studies were judged to be of high risk of bias, due to use of an uncontrolled before/after design, lack of time series analysis and poor reporting of intervention details, hindering replication.

### *Development of the Evaluation Tool*

After several consensus meetings, the 15 suggestions were deconstructed and operationalized into 52 criteria items (Nature of the Desired Action= 9 items; Nature of the Data Available for Feedback= 17 items;

[A]One of the authors provided examples for the trial with two types of A&F

Feedback Display= 9 items; Delivering the Feedback Intervention=17 items). The full evaluation tool, including response scales and anchors, is described in **Appendix C**. During piloting of the evaluation tool, a Cohen's Kappa of 0.58 was computed for the 39 items using a 'Yes/No/Unclear/Not applicable' scale. This represents "moderate agreement" as per Landis and Koch, but is below Krippendorff's cut-off "suggesting that conclusions should be discounted".[27] This relatively low agreement score was partially driven by discrepancies in determining whether the item was not present ('no') versus 'not applicable' or 'unclear'. We therefore proceeded to use our primary approach of establishing consensus for each piloted study.

***Application of the Evaluation Tool***

***Nature of the Desired Action***

**Figure 1** describes consistency with the nine items operationalized from three suggestions focused on the nature of the feedback's desired action.[B] **Suggestion 1** ("Recommend actions that are consistent with established goals and priorities"[1]) was deconstructed into four ratable items: *presence of an internal goal-setting component* (yes: 1/17); *feedback presented in the context of an external, explicit priority* (yes: 13/17); *a description of the priority level* (institutional: 4, departmental: 2, unit: 2, national: 1, healthcare system: 1, unclear: 3); and *whether the feedback directly addressed one or more of the goals or priorities* (yes: 13/14). Four ratable items were developed to assess consistency with **Suggestion 2** ("Recommend actions that can improve and are under the recipient's control"[1]): *whether the feedback provided data on previous performance* (yes: 6/17); *if a clear discrepancy was described or shown between that of the recipient's behaviour and a goal/benchmark/target* (yes: 9/17; other (potential for implied target of 100% compliance): 1/17); *if it was reasonable that the recipients could be responsible for the change in behaviour*

[B] To enhance clarity, only items with a 'Yes/No/Not Reported/Unclear' scale have been displayed. Descriptive items (e.g. 'description of the priority level') have been described in the text.

**Table 2. Summary of Study Characteristics**[a, d]

| Clinical Behaviour Targeted | Number (%) of studies (n=16) |
|---|---|
| **Laboratory Test Ordering** | **8 (50.0%)** |
| Multiple, miscellaneous or unspecified tests | 3 (18.8%) |
| ABG | 2 (6.3%) |
| Lactate & Blood cultures | 1 (6.3%) |
| Superficial cultures | 1 (6.3%) |
| Blood work | 1 (6.3%) |
| **Transfusion Ordering** | **10 (62.5%)** |
| RBCs | 6 (37.5%) |
| FP/FFP | 3 (18.8%) |
| All (RBC, FFP, platelets, cryoprecipitate) | 1 (6.3%) |
| **Study Design** | |
| Uncontrolled Before After | 13 (81.3%) |
| Cluster Randomized Controlled Trial | 1 (6.3%) |
| Controlled Clinical Trial[b] | 1 (6.3%) |
| Controlled Before After | 1 (6.3%) |
| **Data Collection** | |
| Prospective | 5 (31.3%) |
| Retrospective | 4 (25.0%) |
| Mixed | 3 (18.8%) |
| Unclear | 4 (25%) |
| **Funding** | |
| Not Reported | 8 (50%) |
| Government Grant | 4 (25%) |
| Institutional[c] and Non-profit Grants | 2 (12.5%) |
| Institutional[c] | 1 (6.3%) |
| No Funding | 1 (6.3%) |

| Country | Number (%) of studies (n=16) |
|---|---|
| USA | 9 (56.3%) |
| Canada | 2 (12.5%) |
| Finland | 1 (6.3%) |
| Germany | 1 (6.3%) |
| Israel | 1 (6.3%) |
| The Netherlands | 1 (6.3%) |
| Switzerland | 1 (6.3%) |
| **Number of Sites** | |
| Single centre, single ICU study | 9 (56.3%) |
| Single centre, multi-ICU study | 4 (25.0%) |
| Multicentre study | 2 (12.5%) |
| Single centre, # ICUs unclear | 1 (6.3%) |
| **Hospital Type** | |
| Teaching | 11 (68.8%) |
| Not Reported | 3 (18.8%) |
| Other: Veteran's Administration Medical Centre | 1 (6.3%) |
| **ICU Type** | |
| Surgical | 2 (12.5%) |
| Neonatal | 2 (12.5%) |
| Cardiac Surgery | 3 (18.8%) |
| Neurosurgical | 1 (6.3%) |
| Medical | 1 (6.3%) |
| Mixed Patient Population | 2 (12.5%) |
| Multiple Types of ICUs | 3 (18.8%) |
| Not Specified | 2 (12.5%) |

**a:** Proportions were calculated for the 16 studies, rather than the 17 publications. Totals may be slightly greater or less than 100% due to rounding. **b:** The control group was another type of A&F. **c:** 'Institution' refers to both hospitals and academic institutions. **d:** Reprint.[2] *Abbreviations:* ABG=Arterial Blood Gas; CP= cryoprecipitate; FP= Frozen Plasma; FFP= Fresh Frozen Plasma; ICU= Intensive Care Unit; RBC= Red Blood Cell.

(yes:16/17, unclear: 1/17); and *whether the feedback provided data on behaviours* (yes: 17/17),[C] *outcomes* (yes: 5/17), *both* (yes: 5/17). Consistency with **Suggestion 3** ("Recommend specific actions"[1]) was evaluated by one ratable item: *whether corrective actions to help support behaviour change were given* (i.e. an action plan, coping strategy or menu of options) (yes: 3/17).

***Nature of the Data Available for Feedback***

**Figure 2** describes the sample's consistency with the 17 items operationalized from four suggestions relating to the nature of the data available for feedback.[D] **Suggestion 4** ("Provide multiple instances of feedback"[1]) was evaluated by two items: *whether feedback on performance of a given behaviour was provided more than once* (yes: 10/17; no: 4/17; other (unclear/variable): 3/17); and *whether feedback was sustained beyond the study* (yes: 3/17; unclear: 5/17). Four items were extracted to assess consistency with **Suggestion 5** ("Provide feedback as soon as possible and at a frequency informed by the number of new patient cases"[1]): *whether a justification for the interval between feedback reports was provided* (yes: 3/13); *whether this justification was related to the number of patient cases* (yes: 2/3); *the average age of the feedback data* (hours: 2/17; days: 2/17; weeks: 3/17; months: 2/17; days + weeks: 3/17; weeks + months + unclear: 1/17; unclear: 4/17), and *when feedback was provided more than once, the interval between reports* (one day: 1/13, one week: 2/13, one week + one month: 1/13, one month: 3/13, variable + one month: 2/13; variable + 1-3 months: 1/13, unclear/variable: 3/13). **Suggestion 6** ("Provide individual rather than general data"[1]) was deconstructed into six ratable items: *was a justification for the specificity of the feedback data provided* (yes: 5/17); *did the feedback include data about the individual's own performance* (yes: 5/17); *did the feedback include group performance data for which the recipient is a member* (yes: 11/17); *if yes, what was the level of the group* (unit: 4/11; team (multiple units): 2/11; ward + institution: 1/11, unit + department: 1/11, institution (ICUs) + potentially provider groups: 1/11; not reported: 2/11); *did the feedback include patient-level data* (yes: 6/17); and *did the feedback include aggregated patient data* (yes: 11/17). Consistency with **Suggestion 7** ("Choose comparators that reinforce desired behaviour change"[1]) was assessed by 5 items: *do the authors provide a justification for comparators* (no: 9/9); *does the feedback provide any comparators* (yes: 9/17), *how many comparators are provided* (one: 1/9; two: 4/9; unclear, but at least one: 3/9; unclear, but at least two: 1/9); *a description of the comparators* (own/group previous performance: 6/14 comparators; other's performance 4/14 comparators; target: 2/14 comparators; other (group average over time): 1/14 comparators; other (peers' individual performance): 1/14 comparators); and *does the feedback include one or more aspirational comparators* (yes: 2/17; other (no, unless 100% compliance implied): 2/17).

[C] This finding is likely skewed due to the inclusion criteria of the review (reporting on test or transfusion orders)
[D] To enhance clarity, only items with a 'Yes/No/Not Reported/Unclear' scale have been displayed. Descriptive items (e.g. 'average age of the feedback data') have been described in the text.

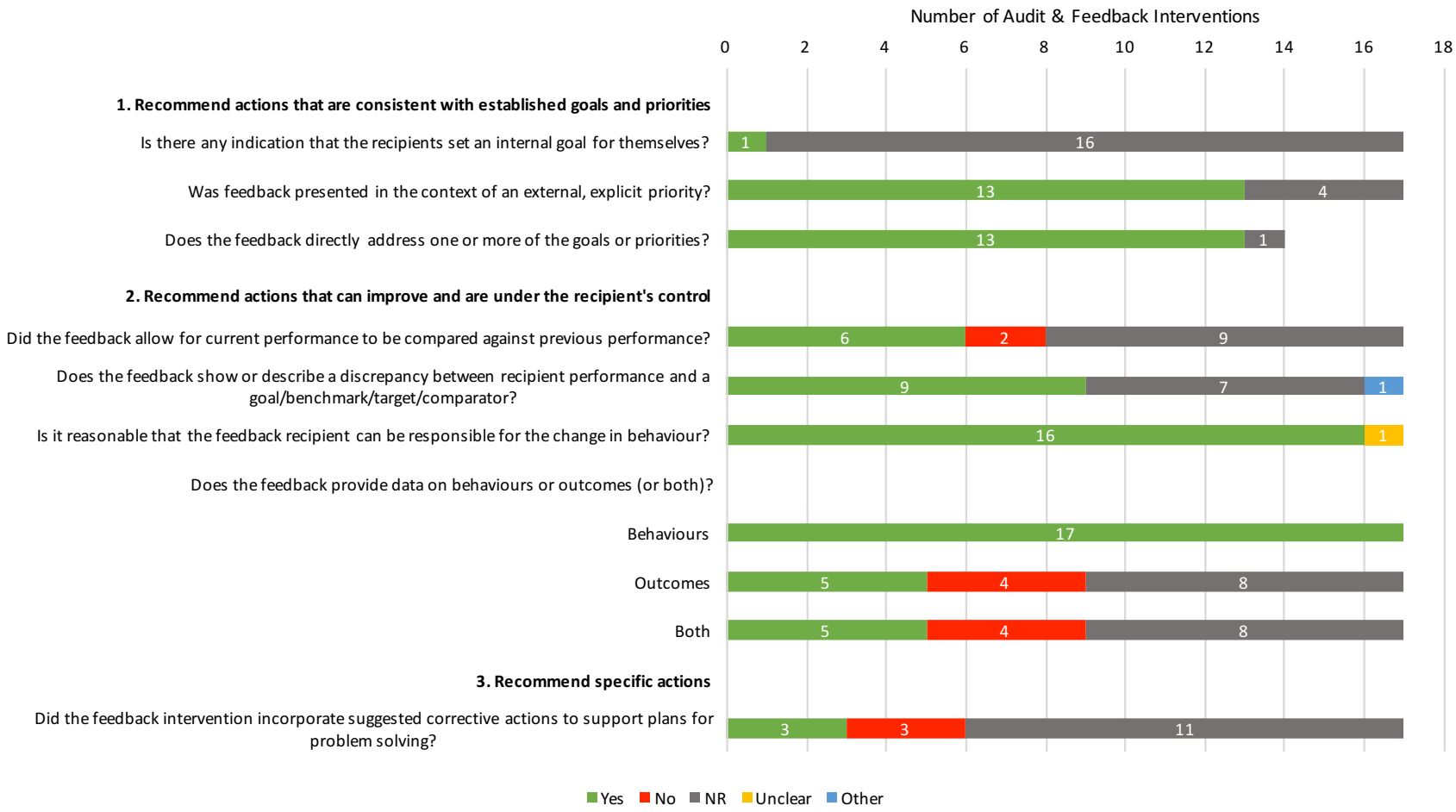



**Figure 1. Description of feedback interventions according to the 'Nature of the Desired Action' items (n=17 feedback interventions).**

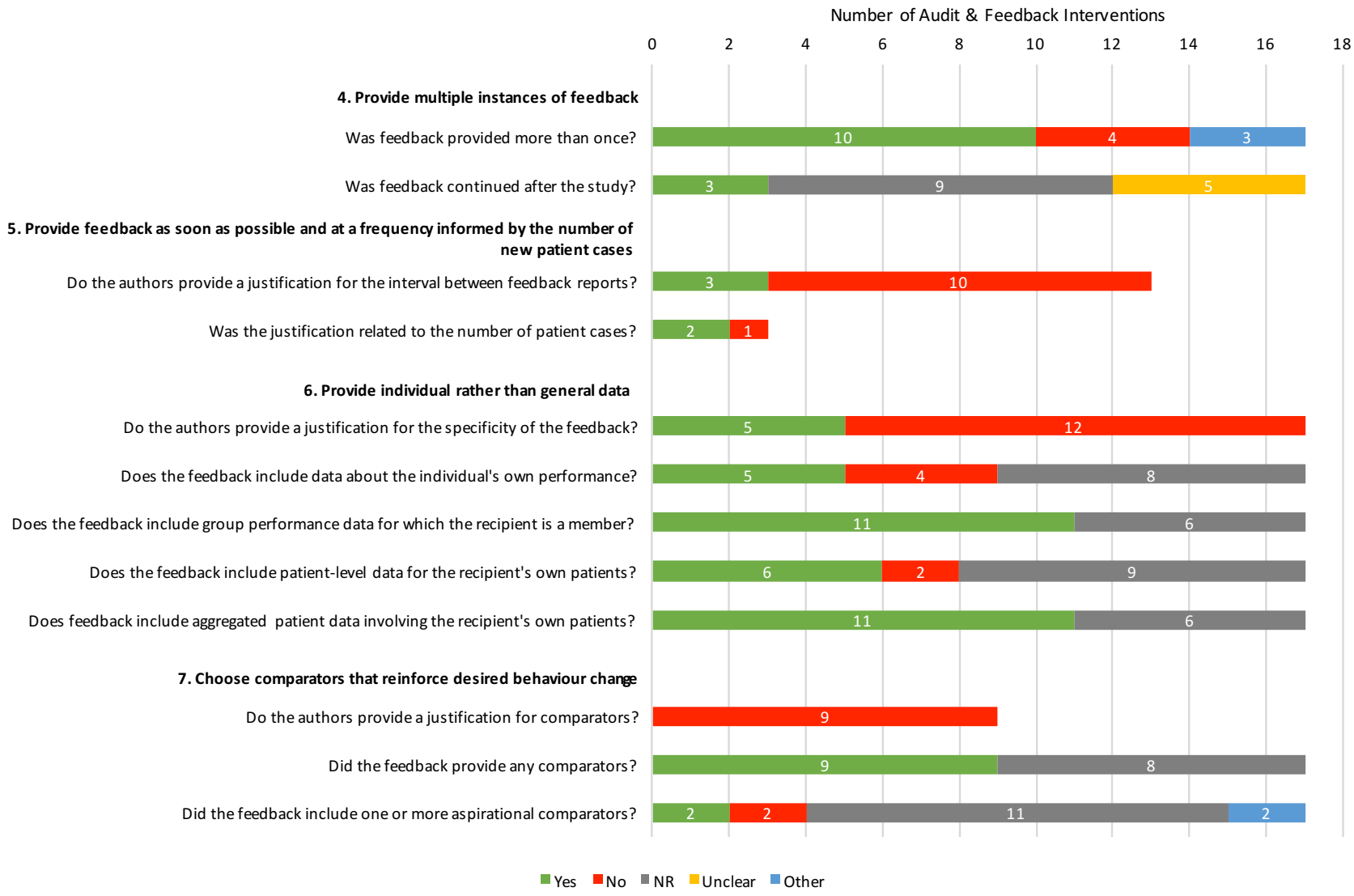



**Figure 2. Description of feedback interventions according to the 'Nature of the Data Available for Feedback' items (n=17 feedback interventions).**

***Feedback Display***

**Figure 3** describes the sample's consistency with the 9 items operationalized from the 3 suggestions relating to the feedback's display.[E] **Suggestion 8** ("Closely link the visual display and summary message"[1]) was evaluated by one criteria item: *are the visual display and summary message presented in visual proximity of each other* (yes: 1/15; not reported: 14/15; N/A: 2/17). **Suggestion 9** ("Provide feedback in more than 1 way"[1]) was assessed by two criteria items: *was feedback provided in more than one way* (yes: 8/17; not reported: 9/17); and *a description of the format used* (verbal: 8/17, text: 5/17, numerical information: 7/17, graphs or tables: 6/17, summary message: 3/17, other important elements: 2/17 (colour coding)). To gauge consistency with **Suggestion 10** ("Minimize extraneous cognitive load for feedback recipients"[1]) we assessed six factors: *was the feedback intervention pilot-tested* (unclear: 1/17); *length of the feedback report* (one page: 1/11; two reports, one page each: 1/11; not reported: 9/11; N/A (i.e. e-mail, presentation, poster): 6/17); *the number of behaviours addressed by the feedback* (at least one: 7/17; two: 1/17; at least two: 2/17; at least three: 1/17; three to five: 2/17; five to nine: 3/17; at least 7, up to 44: 1/17); *the number of clinical variables addressed by the feedback* (none: 3/17; at least one: 1/17; at least four: 2/17; at least six: 1/17; not reported: 10/17), *the number of graphs or tables used* (none: 1/17; one graph: 1/17; two graphs: 1/17; two graphs + one table: 1/17; three graphs + one table: 1/17; unclear, at least one graph: 1/17; unclear, at least two graphs: 1/17; not reported: 10/17), and *whether the feedback used graphical elements that may have led to misinterpretation* (yes: 2/16).

***Delivering the Feedback Intervention***

**Figure 4** describes the sample's consistency with 17 items operationalized from the five suggestions centred around the delivery of the feedback. **Suggestion 11** ("Address barriers to feedback use"[1]) was deconstructed into four criteria items: *were potential drivers and barriers to recipient engagement assessed* (no: 3/17; not reported: 14/17); *was the assessment informed by theory* (not applicable: 17/17); *was there an assessment of whether the recipients engaged with the feedback* (yes: 2/17, other (no, but feedback had a verbal component): 1/17; other (not reported, but feedback had a verbal component): 5/17); and *was the assessment informed by theory* (no: 1/2; not reported 1/2). **Suggestion 12** ("Provide short, actionable messages followed by optional detail"[1]) was assessed by two criteria items: *are the summary messages actionable* (no: 3/17, not reported: 14/17); and *is there additional, more detailed feedback provided alongside the summary message* (yes: 2/17, unclear: 1/17). Consistency with **Suggestion 13** ("Address credibility of the information") was evaluated by four items: *does the feedback indicate who is providing the data* (yes (feedback provided verbally or through e-mail): 10/17); *does the feedback indicate the source of comparators* (yes: 6/9, other (yes, would be obvious): 1/9); *whether the feedback intervention was delivered by a supervisor or close colleague* (yes: 6/17); and *whether the feedback intervention was supported by a relevant professional organization* (unclear: 1/17). **Suggestion 14** ("Prevent defensive reactions to feedback") was assessed by one criteria item: *did the feedback intervention include reassurance that the intervention would not trigger punitive measures* (other (no, but other aspects aiming to reduce defensive reactions were reported): 6/17). For **Suggestion 15** ("Construct feedback through social interaction") we assessed six criteria items: *did development of the feedback intervention involve members of the target group* (not reported: 17/17), *whether feedback was explicitly designed to be received and*

[E]To enhance clarity, only items with a 'Yes/No/Not Reported/Unclear' scale have been displayed. Descriptive items (e.g. 'the number of behaviours addressed by the feedback') have been described in the text.

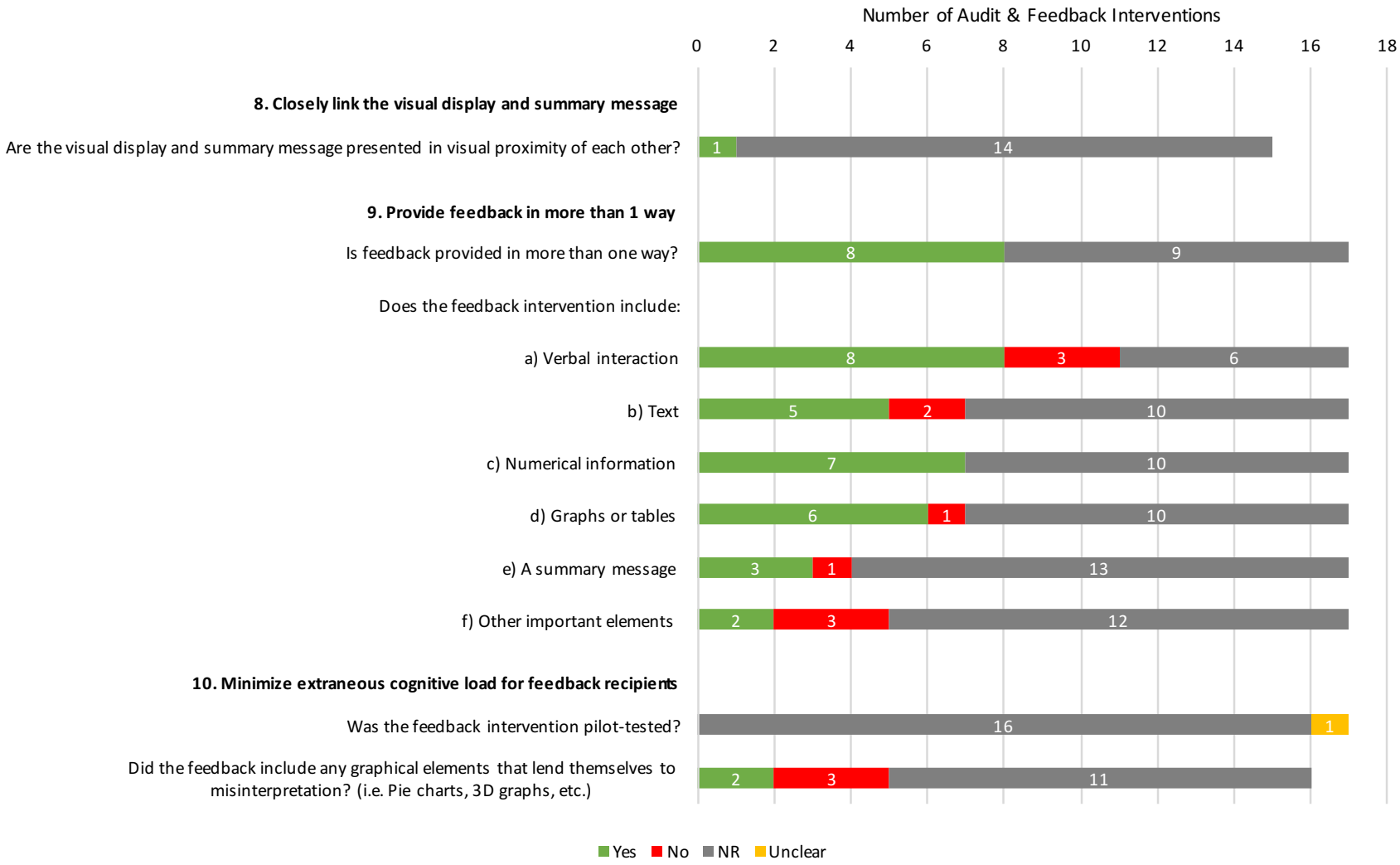



**Figure 3. Description of feedback interventions according to the 'Feedback Display' items (n=17 feedback interventions).**

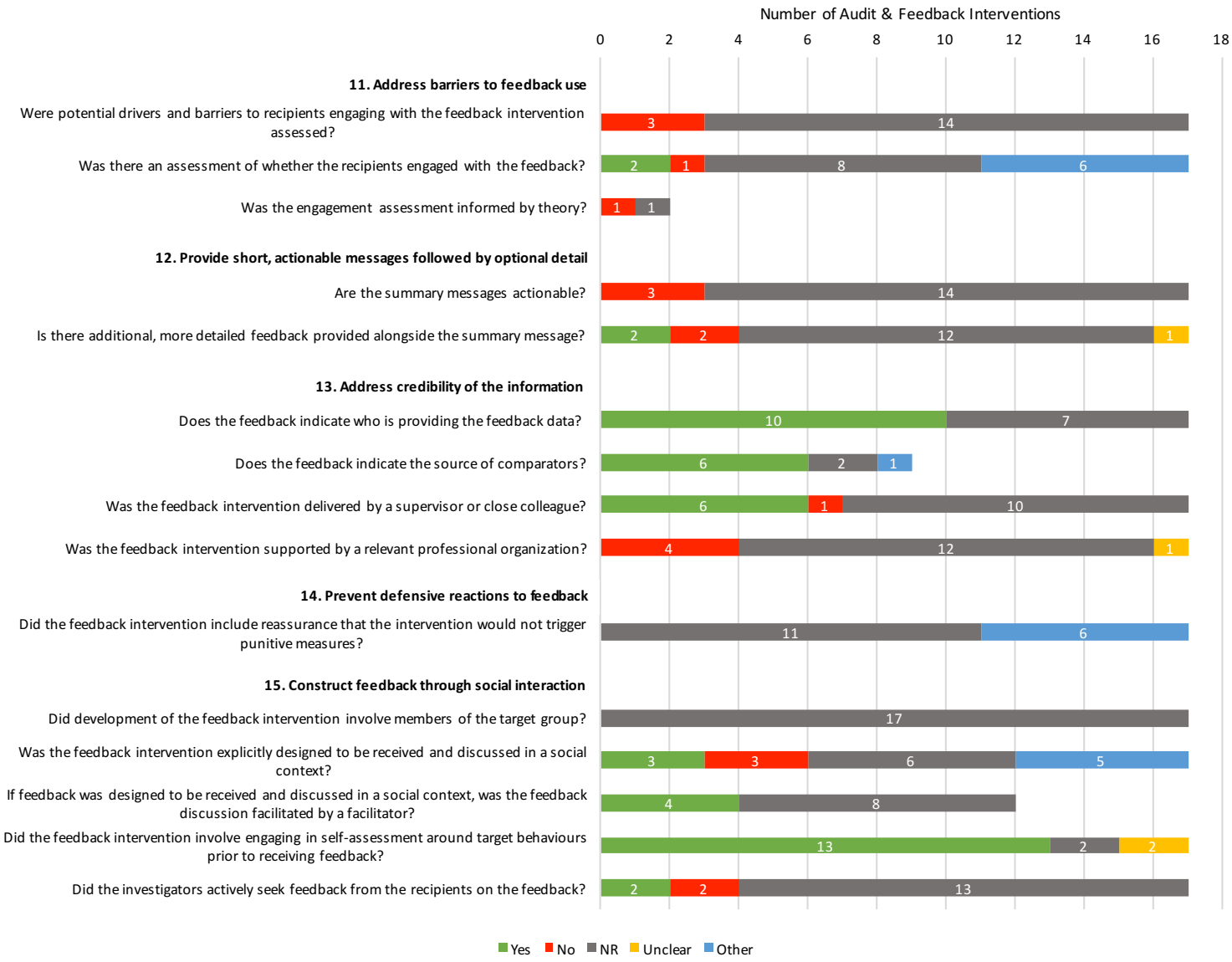



**Figure 4. Description of feedback interventions according to the 'Delivering the Feedback Intervention' items (n=17 feedback interventions).**

*discussed in a social context* (yes: 3/17; other (no explicit statement reported, but feedback was provided in a social context): 5/17), *if provided more than once, how often was feedback received and discussed in a social context* (every time: 4/6, variably (aimed for every time for one of two types of feedback): 1/6; unclear: 1/6; N/A: 11/17); *if feedback was designed to be received and discussed in a social context, was the feedback discussion facilitated by a facilitator* (yes: 4/12); *did the feedback intervention involve engaging in self-assessment around target behaviours prior to receiving feedback* (i.e. an educational session) (yes: 13/17; unclear: 2/17), *did the investigators actively seek feedback from the recipients of the feedback* (yes: 2/17).

## DISCUSSION:

Based on the 15 suggestions published by Brehaut and colleagues, we developed an evaluation tool to assess the extent to which existing A&F interventions already apply these practices. Application of this tool allowed us to describe the landscape of A&F methods used for a sample of critical care interventions, as well as to assess how they align with recent theory-informed guidance. Our work shows that most suggestions had not been consistently implemented across this sample of critical care studies. Of the 15 suggestions, only one was almost universally applied ("Recommend actions that can improve and are under the recipient's control"). This was not particularly surprising as all articles within the sample were published prior to or within the same year as the suggestions for best practice,[1] which were hypothesized to be relatively underutilized elements within the existing A&F literature. The results from this study suggest there may be considerable room for improvement in the development and delivery of A&F interventions for test and transfusion ordering in the critical care setting, and point to several theoretical considerations that warrant further study.

We also found that many details regarding feedback elements were not reported. As we were not always able to access an example of the template, it was difficult to judge overall consistency with the suggestions related to the design and delivery of the feedback. Our findings suggest a standardized method for reporting A&F intervention details would enable future studies to build upon the extant literature.

### *Use of Theory in A&F*

There is great interest in utilizing theory to improve the design, implementation and assessment of behaviour change interventions, as suggested in the Medical Research Council's guidance.[28,29] *A priori* predictions of mechanisms of action for complex interventions through consideration of relevant theories can facilitate a better understanding of why an intervention is or is not successful.[30–32] Recent analyses of the A&F literature (the 140 trials included in the Cochrane systematic review),[30] as well as the more general implementation literature (guideline implementation)[33] however, have revealed low rates of reported theory use. A review assessing the use of theory across Cochrane systematic review A&F interventions found theory was mentioned in only 20 of the articles (14%), and a mere 13 of these referenced theory in terms of A&F design.[30] Researchers may have difficulty selecting theories for application to their interventions and studies, due to the lack of consensus and, until recently, lack of guidance on how best to choose from numerous theories.[34–36] To synthesize theoretically informed guidance for A&F developers, Brehaut and colleagues conducted interviews with theory experts and drew from team experience and systematic reviews.[1] Our finding that many of these theoretically informed suggestions are underutilized in the existing

critical care A&F literature, is therefore in line with these previous studies. Below we've described several key suggestions for which our sample showed low consistency, to highlight priorities for future research.

***Underutilized Suggestions***

Counter to Brehaut et al.'s suggestion to "provide multiple instances of feedback",[1] some interventions only provided feedback once, some did not clearly report whether feedback was provided more than once, and others provided feedback variably (i.e. only if an order was placed inappropriately). Although many of the interventions did provide feedback more than once, our finding that this suggestion was not universally applied is an important one. Providing feedback more than once allows for a cyclical process whereby the recipient first receives feedback on their behaviour (and potentially suggestions on how to improve).[1] If allowed the chance to change their behaviour, upon receiving feedback again, individuals may gauge whether their efforts were successful or not. Without iterative feedback, recipients may not be able to determine their progress on whether their efforts were successful. As it is suggested that individuals are unable to accurately assess their own performance, this is an important part of the feedback loop.[1,3,37]

Another suggestion that was relatively underutilized in our sample was the provision of individual feedback data, as most interventions presented data aggregated at the group level. Brehaut et al.'s recent guidance suggests feedback provide data at the individual level, to avoid discounting of the data.[1] However, as noted in one study, it may be difficult to "assign" orders to individual healthcare providers if the decision is made by the team.[19] A previous meta-analysis also found a combination of group and individual data to result in a larger effect size than either type alone.[38] It may therefore be of interest to further assess whether providing both individual and group level data is more effective in team-based settings such as the ICU.

While none of the studies reported providing reassurance that the feedback intervention would not result in punitive measures, six studies did report the incorporation of other aspects (e.g. providing both positive and negative feedback, providing group data to avoid singling out individuals, using non-punitive wording, etc.) that were consistent with the suggestion to aim to "Prevent defensive reactions to feedback".[1] This is an important component, as previous qualitative work has identified that providers may have negative perceptions of feedback.[39,40] It is also imperative to note that our criteria item represents only one way of potentially preventing defensive reactions, the effectiveness of which should still be tested. Further work is required to elucidate methods on how best to avoid negative reactions to feedback. Another interesting consideration is that several interventions provided feedback only on inappropriate orders. Providing only negative feedback, repeatedly, may be counterproductive, in leading to decreased self-efficacy and motivation.[1,41]

Other significant findings from our study included the lack of reported use of elements such as piloting of the feedback form, involvement of key stakeholders, barrier and engagement assessments, goal setting, and action and coping plans. This lack of reporting could potentially represent that these steps are not being taken. If that is the case, it would be valuable to assess whether incorporation of these elements in A&F studies helps to improve efficacy in the critical care setting. Involving stakeholders in the development process may also help to identify priorities and appropriate modalities through which to provide feedback. Previous qualitative work has found that ICU specialists feel A&F to be a "fragmented or discontinuous communication", "often not actionable", and have noted that the audit process can "[lack] transparency and credibility"[39]. Engaging stakeholders throughout the development of the feedback may also help to ensure providers feel a part of the process and that the feedback provided is useful and positively received. Further,

as test and transfusion ordering are likely habitual behaviours, use of supports such as action or coping plans may be especially pertinent, because it's hypothesized that these plans can help to form new habits.[7,42]

***Strengths & Limitations***

The development of our evaluation tool represents an important step forward in our work to improve feedback interventions. A total of 52 criteria items were operationalized from the 15 suggestions for best practice.[1] To address the uncertainty surrounding the specifics of how best to apply each suggestion, we developed a comprehensive set of items which aimed to capture the various ways in which these suggestions could be employed. This tool allows for assessment of the extent to which A&F interventions adhere to recent guidance for best practice.[1] Future studies may apply this tool to assess how A&F interventions in various settings adhere to these suggestions, as well as whether adoption of these practices improve over time. Moreover, our tool may be used prospectively for the development of A&F interventions, to test the various hypotheses.

A limitation of our evaluation is the potential for lack of reporting on key details in the intervention descriptions available to us. As demonstrated by Colquhoun et al. the reporting quality of A&F intervention details varies.[43] Since the majority of studies in our sample used multiple intervention components, space limits may have especially inhibited reporting of such details. As varied reporting was anticipated, we aimed to counter this limitation by contacting the study authors to obtain further details and request an example of the feedback form used during the study; having access to the feedback form allows for more complete coding.

With examples of the feedback template for only seven of the 17 interventions, several of our assessments relied on the details reported within manuscripts. Poor reporting of intervention details and lack of access to the feedback template may have led to an under-representation of adherence (i.e. consistency with the suggestions may actually be higher than what is reported in the published literature). We therefore aimed to take a conservative approach and coded 'not reported' (or 'unclear' if the text provided partial but ambiguous details) whenever lack of access to the feedback template limited our ability to judge the criteria item. Assessment of criteria items relating to the feedback design and delivery especially suffered without access to the feedback form as these details were typically not reported on. Several of the feedback interventions also involved verbal summaries, rendering some items not possible to assess.

Due to the variation in reporting of intervention details and limited access to feedback templates, the current evaluation tool is not meant to generate a score representing overall consistency. Further psychometric testing will be required in future work to assess construct, content and criterion validity.[44] Furthermore, it is important to note that any one criteria item or suggestion should not be considered as a requirement; contextual factors may affect the appropriateness or even limit the ability to implement any given suggestion (i.e. type of data available).[1,29,45] We also had a relatively small sample upon which to apply the tool. Future studies could apply the tool to a larger sample of A&F interventions in the critical care setting, with more complete intervention details, to compare findings.

***Implications for Future Directions***

Studies of A&F interventions should explore whether the provision of both individual and group feedback may prove to be beneficial, particularly in team based settings like the ICU. Future studies may also wish

to test the effectiveness of including elements such as goal setting, aspirational comparators, support by a relevant professional organization, piloting of the feedback form, barrier and engagement assessments, action and coping plans, and engagement with key stakeholders. Assessment of strategies to mitigate negative reactions to the feedback, such as the proposed "feedforward" strategy, which involves having the recipient recall a past success and using this to identify an "internal standard of excellence" to compare to their performance,[1,41] would also be useful.

Future A&F studies should also aim to clearly report intervention details (i.e. when A&F is provided, how often, by whom) and provide an example of the feedback template. This will help to advance the field by allowing other researchers to assess what was done and build upon this research. More complete reporting will also help to improve chances that the intervention can be replicated by others. To address space limitations, we propose future studies provide an example of the feedback template and report a more detailed description of the intervention development and delivery in online, supplemental appendices.

## CONCLUSIONS

Application of our evaluation tool to a sample of critical care A&F interventions helped us to determine that features consistent with the 15 theoretically informed suggestions are rarely reported. Of the 15 suggestions, one was almost universally applied, one was well applied when we were able to assess for it, seven were rarely applied, and three were difficult to assess due to lack of access to the feedback template. If rare application simply represents a reporting issue, we believe this calls for the need to standardize the way in which feedback interventions are reported. We propose that examples of the feedback form and a detailed methods section be posted online as appendices. Reporting checklists and standard nomenclature taxonomies may also help to ensure comprehensive and clear reporting of interventions.

## DECLARATIONS

***Ethics approval and consent to participate:*** This study was approved by the Ottawa Health Science Network Research Ethics Board (OHSN-REB; Protocol ID: 20160951-01H). The corresponding authors of studies identified through systematic review were contacted to request the sharing of feedback forms. Individuals were informed that by sharing these study materials they were providing consent to participate in the study.

***Consent for publication:*** Not applicable.

***Availability of data and material:*** The authors may be contacted for data and material requests. However, the data and materials pertaining to individual participants will not be shared to protect privacy.

***Competing interests:*** The authors have no competing interests to declare.

***Funding:*** MF received a Queen Elizabeth II scholarship, a Graduate Scholarship from the University of Ottawa, and a studentship stipend from the Ottawa Hospital Research Institute.

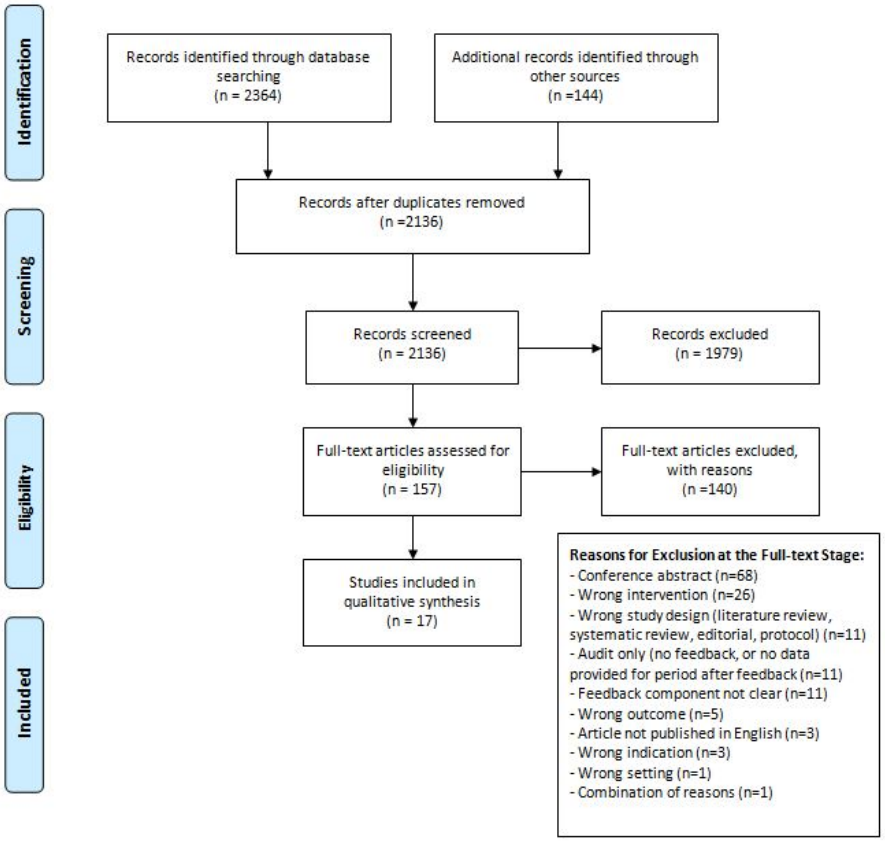
***Authors' contributions:*** **MF** and **KC** conducted screening for the systematic review, to identify the sample of A&F interventions. **JCB** and **JP** provided guidance throughout this process and resolved conflicts between reviewers. All members of the study team were involved in the development of the evaluation tool (**JCB, JP, KC, MP, MF**). **MF** and **MP** piloted the evaluation tool. Formal assessment of the identified studies was completed by **MF** with guidance from **JCB** and **JP**. **MF** drafted the manuscript and **JCB** and **JP** edited and made contributions to the manuscript. The guarantor of this report is **MF**.

***Authors' information (optional):*** Madison Foster (madfoster@ohri.ca), Jamie Brehaut (jbrehaut@ohri.ca), Justin Presseau (jpresseau@ohri.ca).

## Appendix A: Changes to Protocol

| Section | Change |
|---|---|
| Methods | We had originally planned to continue the piloting phase until an IRR of 80% was reached. However, during the piloting phase the study team decided that the formal assessment would involve the true validation of the evaluation tool. The IRR was still recorded throughout the piloting process, however, we no longer required an IRR of 80% to proceed to the formal evaluation of studies. Instead, we used the results from the piloting study to further clarify the criteria items, and add any relevant examples to the guidance section for coders. We also originally planned to compute kappa scores for each item after each meeting; this was changed to calculation of an overall kappa for the 39 items with a 'Yes/No/Unclear/Not applicable' scale for each round and at the end of the piloting phase. |
| Methods | Data was to be extracted from the feedback forms and articles in the same manner as described for the pilot phase. And if coders were unable to come to an agreement by consensus, a third member of the study team was to be asked to resolve the disagreement. Data extraction was instead completed by one researcher, due to limited resources. |
| Methods | The proportion of articles in compliance (%) and the inter-rater agreement (%) for each criterion was to be described. Descriptive statistics were to summarize study characteristics such as the type of laboratory test or blood component targeted, patient condition, type of ICU, intervention target (i.e. healthcare professional type), study type, study country, type of funding, and approach (i.e. single mechanism versus multifaceted). Chi-square tests (categorical variables) and t-tests (continuous variables) were to be used to determine whether compliance with each criteria item varied according to different study characteristics. A p-value less than 0.05 was to be considered significant.<br><br>The number of articles rather than percentage was used, and inter-rater agreement for each criterion will be calculated after duplicate extraction is completed for publication. The originally planned analyses were not completed due to a small sample size. A narrative analysis was instead completed. |

## Appendix B: PRISMA Flow Diagram[9] outlining the selection of articles (reprint[2])



**Appendix C. Evaluation Tool Criteria Items, Response Scales and Anchors Developed from Brehaut et al.'s 15 Suggestions**[1]

| Description | Criteria Item | Response Scale[A] | Anchor |
|---|---|---|---|
| **Suggestion #1: Recommend actions that are consistent with established goals and priorities** | | | |
| 1 a) Goal | Is there any indication that the recipients set an internal goal for themselves (i.e. a specific, numerical target/threshold)? | Yes, No, Not Reported, Unclear | Code as **'Yes'** if the FB intervention involved the recipients setting an internal goal for themselves. **Example:** The FB intervention included peer discussions where individual goals were set and discussed. Code as **'No'** if the FB intervention did not involve the recipients setting an internal goal (and the feedback form or additional details are available to confirm this). Code as **'Not reported'** if there are no details in the text regarding whether the FB intervention involved internal goal setting by the recipients (and the feedback form/additional details are not available). Code as **'Unclear'** if the text contains an ambiguous statement and the feedback form/additional details are not available to confirm. |
| 1 b) Priority | Was feedback presented in the context of an external, explicit priority? | Yes, No, Not Reported, Unclear | Code as **'Yes'** if the FB is given in the context of a guideline or initiative. **Note:** This guideline or initiative should be made clear to the FB recipients (i.e. is discussed on the FB form, OR the article describes that the recipients were aware of the guideline/initiative). Code as **'No'** if the FB was not given in the context of a guideline or initiative (and the feedback form or additional details are available to confirm this). **Example:** The authors generally discuss guidelines or the need to change a behaviour in the introduction, but there is no indication of this priority on the FB form. Code as **'Not reported'** if there are no details in the text regarding whether the FB intervention involved an external, explicit priority (and the feedback form/additional details are not available). **Example:** The article discusses providing FB to physicians in light of a provincial guideline, however it is unclear whether the FB indicated this to the physicians. Code as **'Unclear'** if the text contains an ambiguous statement and the feedback form/additional details are not available to confirm. **Note:** There could potentially be more than one (i.e. A provincial guideline and a hospital initiative). |

[A] An 'Other' option was added as required.

| | | | |
|---|---|---|---|
| 1 c) Priority Level | If yes, at what level was the priority set? | National/Federal, Provincial/State, Municipal, Institutional, Departmental, Individual-healthcare providers, Individual-researchers, Other (Please specify), N/A (No discussion of an explicit priority), Unclear, No form | |
| 1 d) Link to FB | Does the feedback directly address one or more of the goals or priorities? | Yes, No, Not Reported, Unclear, N/A (no goal/priority) | Code as **'Yes'** if the FB corresponds with either: one of the goals set by recipients or one of the external priorities presented. **Example:** FB is presented in the context of a hospital initiative to reduce inappropriate transfusions and the FB provides information on the number of transfusions which were inappropriate, along with suggestions on how to reduce inappropriate orders. Code as **'No'** if the feedback does not correspond with one of the goals set by recipients or one of the external priorities presented. **Example:** FB is presented in the context of a provincial guideline to reduce inappropriate laboratory tests, but the FB to recipients only provides information on their laboratory expenditure (i.e. there is no information on the appropriateness of the lab tests). Code as **'Not reported'** if there are no details in the text and the feedback form/additional details are not available. Code as **'Unclear'** if the text contains an ambiguous statement and the feedback form/additional details are not available to confirm. Code as **'N/A (no goal/priority)'** if the FB recipients do not set internal goals for themselves and the FB is not presented in the context of a priority. |
| **Suggestion #2: Recommend actions that can improve and are under the recipient's control** | | | |
| 2 a) Previous performance | Was feedback on performance provided to allow current performance to be compared against previous performance? | Yes, No, Not Reported, Unclear | Code as **'Yes'** if the FB shows an individual's performance over time (i.e. performance data from at least two time points). Code as **'No'** if the FB only shows performance from one time point. Code as **'Not reported'** if there are no details in the text and the feedback form/additional details are not available. Code as **'Unclear'** if the text contains an ambiguous statement and the feedback form/additional details are not available to confirm. |
| 2 b) Discrepancy | Does the feedback (or the description of the FB) describe or show a discrepancy between recipient | Yes, No, Not Reported, Unclear | Code as **'Yes'** if the FB included at least one comparator (including 'own previous performance') and the discrepancy between these two performance levels was made clear (i.e. the difference is noted, a bar graph is displayed etc.). Code as **'No'** if no comparators were provided, or if a comparator was provided and the discrepancy was not made clear (i.e. the performance of the individual is displayed on one page, and the performance of the comparator is displayed on another). Code as **'Not reported'** if there are no details in the text and the feedback form/additional details |

| | | | |
|---|---|---|---|
| | performance and the goal/ benchmark/ target/comparator? | | are not available (i.e. the article does not include an example of the FB form and from the description of the intervention it is not clear whether the performance levels were displayed on the same graph or not). Code as **'Unclear'** if the text contains an ambiguous statement and the feedback form/additional details are not available to confirm. |
| 2 c) Control | Is it reasonable that the feedback recipient can be responsible for the change in behaviour? | Yes, No, Not Reported, Unclear | **Note:** The recipient does not need to be directly responsible. Code as **'Yes'** if the recipient can make a difference in the behaviour being fed back, either directly (i.e. it is their behaviour that needs to change) or indirectly (i.e. they are to discuss with/train the individuals whose behaviour needs to change). Code as **'No'** if the FB recipient cannot be responsible for the change in behaviour. **Example:** FB on physician test ordering is provided to the CEOs of the hospital. Code as **'Not reported'** if there are no details in the text and the feedback form/additional details are not available. Code as **'Unclear'** if the text contains an ambiguous statement and the feedback form/additional details are not available to confirm whether the recipient can make a difference in the behaviour being fed back. |
| 2 d) Outcome Type | Does the feedback provide data on behaviours, outcomes, or both? | For each, answer: Yes, No, Not Reported, Unclear | Behavioural data includes process measures such as prescribing habits, test ordering, etc. Outcome data includes patient outcomes such as patient blood pressure, mortality, length of stay etc.<br><br>Code as **'Yes'** if clearly present. Code as **'No'** if clearly absent. Code as **'Not reported'** if there are no details in the text and the feedback form/additional details are not available. Code as **'Unclear'** if the text contains an ambiguous statement and the feedback form/additional details are not available to confirm. |
| **Suggestion #3: Recommend specific actions** | | | |
| 3 a) Corrective Actions | Did the FB intervention incorporate suggested corrective actions to support plans for problem solving? For example: action plans, coping strategies, a menu of options, etc. | Yes, No, Not Reported, Unclear | **Note:** Actions may be mandated or developed in a group setting. Code as **'Yes'** if the FB intervention provided an action plan, a coping strategy, a menu of options, or facilitated recipients in developing a plan. **Example:** The FB indicated that a physician's actions were non-compliant with the guidelines and provided a menu of options to proceed [i.e. "(1) rescheduling the patient sooner, (2) marking the clinic encounter form... to suggest the physician perform the indicated preventative care on the patient's next scheduled visit, (3) indicating the protocol was not applicable for this patient (but the physician agrees, in principle, with the protocol), (4) stopping this reminder (the physician disagrees with the protocol), and (5) pulling the patient's chart for review"].[46] Code as **'No'** if the FB intervention did not provide corrective actions to support plans for problem solving. Code as **'Not reported'** if there are no details in the text and the feedback form/additional details are not available. Code as **'Unclear'** if the text contains an ambiguous statement and the feedback form/additional details are not available to confirm whether the FB intervention included corrective actions to support plans for problem solving. |

| | | | |
|---|---|---|---|
| **Suggestion #4: Provide multiple instances of feedback** | | | |
| 4 a) # of FB Reports | Was feedback (for a given behaviour) provided more than once? | Yes, No, Not Reported, Unclear | **Note:** One FB report with multiple time-points constitutes receiving FB once. Three FB reports, each on a different clinical behaviour[47] constitutes receiving FB once for a given behaviour.<br><br>Code as **'Not reported'** if there are no details in the text and the feedback form/additional details are not available. Code as **'Unclear'** if the text contains an ambiguous statement and the feedback form/additional details are not available to confirm. |
| 4 b) Sustained? | Did recipients continue to receive feedback on their performance after the study was completed? | Yes, No, Not Reported, Unclear | **Note: This is a reporting issue.** Code as **'Yes'** if the authors explicitly state that feedback continued after completion of the study. Code as **'No'** if the authors explicitly state that feedback was only provided for the duration of the study. Code as **'Not reported'** if there are no details in the text. Code as **'Unclear'** if the text contains an ambiguous statement and it is unclear whether the feedback initiative was continued beyond the study. **Example:** The authors state that a larger scale program was being prepared, but it is unclear whether or not this program was actually put in place.[47] |
| **Suggestion #5: Provide feedback as soon as possible and at a frequency informed by the number of new patient cases** | | | |
| 5 a) Age of the data | What is the average age of the data (i.e. interval between the clinical encounter and delivery of the feedback)? | Describe (i.e. Days, Weeks, Months, Years, Not Reported, Unclear) | **Example:** If feedback reports are provided weekly, the average age of the data is 'days'. If feedback reports are provided monthly, the average age of the data is 'weeks'. If feedback reports are provided annually, the average age of the data is 'months', etc. Code as **'Not reported'** if there are no details in the text and the feedback form/additional details are not available. Code as **'Unclear'** if the text contains an ambiguous statement and the feedback form/additional details are not available to confirm. |
| 5 b) Interval | If feedback was provided more than once, what was the time interval between the receipt of feedback reports? | Describe (i.e. Hours, Days, Weeks, Months, Years, Variable, Not Reported, Unclear, N/A) | **Example:** An interval of 1 week ='days', an interval of 1 month = 'weeks', an interval of 1 year= 'months' etc. |
| 5 c) Justification-Interval | Do the authors of the study provide a justification for the interval between feedback reports? | Yes, No, Unclear, N/A | **Note: This is a reporting issue.** Code as **'Yes'** if the authors state a reason for choosing the interval between feedback reports. This may include a statement in the introduction stating that a certain interval has been effective in the past. Code as **'No'** if the authors do not provide a reason for choosing the interval between feedback reports. Code as **'Unclear'** if it is not clear whether the authors have provided a sufficient justification. Code as **'N/A'** if feedback was only provided once. |

| | | | |
|---|---|---|---|
| 5 d) Justification related to patient #? | Was the justification for the interval between feedback reports related to the number of patient cases? | Yes, No, Unclear, N/A | Code as **'Yes'** if the justification is related to the number of patient cases. **Example:** To avoid alert fatigue, the authors explain that feedback was only provided quarterly, as physicians only see a few of these patients every month.[1] Code as **'No'** if the justification is not related to the number of patient cases. Code as **'Unclear'** if it is unclear whether the justification is related to the number of patient cases. Code as **'N/A'** if no justification was given. |
| **Suggestion #6: Provide individual rather than general data** | | | |
| 6 a) Individual performance | Was feedback given about the individual's own performance? | Yes, No, Not Reported, Unclear | **Note:** This question does not apply to aggregate level data for a group of which the recipient is a member of; see next question. Code as **'Yes'** if FB is about an individual's own performance. **Example:** Dr. X receives feedback on his own performance (i.e. The number of B12 tests he ordered). Code as **'No'** if only group performance data is given. Code as **'Not reported'** if there are no details in the text and the feedback form/additional details are not available. Code as **'Unclear'** if the text contains an ambiguous statement and the feedback form/additional details are not available to confirm whether individual performance data was given. |
| 6 b) Group performance | Was FB about the performance of a group of which the recipient is a member? | Yes, No, Not Reported, Unclear | Code as **'Yes'** if FB provides performance data for a group of which the recipient is a member of. **Example:** Dr. X receives feedback on the number of B12 tests ordered by his entire practice. Code as **'No'** if no group performance data is given. Code as **'Not reported'** if there are no details in the text and the feedback form/additional details are not available. Code as **'Unclear'** if the text contains an ambiguous statement and the feedback form/additional details are not available to confirm whether group level performance data was provided. |
| 6 c) Group level | What is the level of the group? | Describe: Unit, Department, Practice, Hospital, Region, Province, Other (please specify), N/A, Not Reported | |
| 6 d) Individual data | Did feedback include patient-level data for the recipient's own patients? | Yes, No, Not Reported, Unclear | Code as **'Yes'** if FB provided individual patient data. **Example:** Dr. Y receives feedback on the number of blood tests she ordered for each individual patient over the past week. Code as **'No'** if FB does not provide patient-level data. Code as **'Not reported'** if there are no details in the text and the feedback form/additional details are not available. Code as **'Unclear'** if the text contains an ambiguous statement and the feedback form/additional details are not available to confirm whether patient level data was provided. |
| 6 e) Aggregated data | Did feedback include aggregated patient data | Yes, No, Not Reported, Unclear | Code as **'Yes'** if FB provides aggregated patient data. **Example:** Dr. Y receives feedback on the average number of blood tests she ordered per patient over the past month. OR, Dr.Y receives feedback on the average number of blood tests ordered by her practice over the past month. Code |

| | | | |
|---|---|---|---|
| | involving recipient's own patients? | | as **'No'** if only individual patient level data was provided. Code as **'Not reported'** if there are no details in the text and the feedback form/additional details are not available. Code as **'Unclear'** if the text contains an ambiguous statement and the feedback form/additional details are not available to confirm whether aggregated or individual patient level data was provided. |
| 6 f) Justification-Specificity | Do the authors give a justification for the specificity of the feedback (or a reason for the level of data presented)? | Yes, No, Unclear | **Note: This is a reporting issue.** Code as **'Yes'** if the authors provide a justification for either the specificity of the feedback or feedback data. **Example:** We chose to provide individual performance data, as this has been shown to be more effective than group performance data, as per a recent systematic review. Code as **'No'** if no justification is given for either the specificity of the feedback or the feedback data. Code as **'Unclear'** if it is unclear whether the authors have provided a reason or not. |
| **Suggestion #7: Choose comparators that reinforce desired behaviour change** | | | |
| 7 a) Comparator? | Did the feedback provide any comparators? | Yes, No, Not Reported, Unclear | Code as **'Yes'** if recipients were given at least one clear comparator. Code as **'No'** if recipients were not given a clear comparator. Code as **'Not reported'** if there are no details in the text and the feedback form/additional details are not available. Code as **'Unclear'** if the text contains an ambiguous statement and the feedback form/additional details are not available to confirm. |
| 7 b) # of Comparators | How many comparators were provided? | 1, 2, 3, More than 3, None, Not Reported, Unclear | **Note:** A comparator may be a target, benchmark, peer performance or self-comparator if there is a direct comparison to self at different time points. |
| 7 c) Describe | Describe all that apply. | Own previous performance, Other's performance, Benchmark/ Standardized Guideline/Target, Other, Not Reported, Unclear | If 'Other', describe. |
| 7 d) Aspirational Comparator | Does feedback include one or more aspirational comparators (as opposed to average performance comparators)? | Yes, No, Not Reported, Unclear | **Note: Mean/average level comparators do not qualify.** Code as **'Yes'** if the FB provided a comparator or target based on a high percentile (i.e. the performance of the top 10%) or a standard. Code as **'No'** if the FB did not provide a comparator, or if the FB provided a mean/average level target. Code as **'Not reported'** if there are no details in the text and the feedback form/additional details are not available. Code as **'Unclear'** if the text contains an ambiguous statement and the feedback form/additional details are not available to confirm if the FB provided a comparator/target and it is unclear whether it is based on a high percentile. |

| | | | |
|---|---|---|---|
| 7 e) Justification-Comparators | Do the authors provide a justification for which comparators were used? | Yes, No, Unclear, N/A | **Note: This is a reporting issue.** Code as **'Yes'** if the authors provide a reason for choosing a certain comparator. This may include a statement in the introduction noting the effectiveness of a comparator. **Example:** "Performance feedback, especially with *comparison to peers*, is another effective method of improving compliance with clinical practice guidelines. Although not enough is known about the optimal characteristics of feedback, *comparison with an 'achievable benchmark of care'* appears promising."[48] Code as **'No'** if no clear reason is given for choosing a comparator. Code as **'Unclear'** if it is unclear whether the authors provided a reason or not. Code as **'N/A'** if no comparator is provided. |
| **Suggestion #8: Closely link the visual display and summary message** | | | |
| 8 a) Linking of FB elements | Are the visual display and the summary message presented in visual proximity of each other? | Yes, No, Not Reported, Unclear, N/A | Code as **'Yes'** if there are any obvious issues with how the FB elements are linked. **Example:** 1) The summary message for a graph is shown on a different page than the graph.[1] **2)** The summary message and visual display give conflicting information i.e. the summary message indicates the physician needs to increase their number of referrals, but the visual display shows that they have more referrals than the top 10% benchmark.[1] Code as **'No'** if there are no obvious issues with how the FB elements are linked. Code as **'Not reported'** if there are no details in the text and the feedback form/additional details are not available. Code as **'Unclear'** if the text contains an ambiguous statement and the feedback form/additional details are not available to confirm. Code as N/A if either the summary message or visual display is missing. |
| **Suggestion #9: Provide feedback in more than one way** | | | |
| 9 a) >1 way? | Was feedback provided in more than one way? | Yes, No, Not Reported, Unclear | Based on factors assessed in the next item (9b) was feedback provided in more than one way? |
| 9 b) Format | Does the feedback intervention include: 1) Verbal interaction; 2) Text; 3) Numerical information; 4) Graphs or tables; 5) A summary message; 6) Other important elements (please specify)? | Provide an answer for each item: Yes, No, Not Reported, Unclear | Code as **'Yes'** if clearly present. Code as **'No'** if clearly absent. Code as **'Not reported'** if there are no details in the text and the feedback form/additional details are not available. Code as **'Unclear'** if the text contains an ambiguous statement and the feedback form/additional details are not available to confirm. |

| Suggestion #10: Minimize extraneous cognitive load for feedback recipients | | | |
|---|---|---|---|
| 10 a) Pilot | Was the feedback intervention pilot-tested? | Yes (With target population), Yes (With non-target population), No, Not Reported, Unclear | **Note: This is a reporting issue.** Code as **'Yes'** if the article indicates that the FB intervention was pilot-tested. Code as **'No'** if the article indicates that the FB intervention was not pilot-tested. Code as **'Not reported'** if there are no details in the text and additional details from the author were not available. Code as **'Unclear'** if the text contains an ambiguous statement and it is unclear whether the FB intervention was pilot-tested. |
| 10 b) # of Pages | How long is the feedback report? | Describe. | **Note:** Extract the number of pages for reports. Note whether pages are single-sided or double-sided (if able). If the length is not given (and the form is not available), code as **'Not Reported'**. Code as 'N/A' if the feedback was provided verbally, involved an e-mail or poster, etc. |
| 10 c) # of Behaviours | How many behaviours does the feedback address? | 1, 2, 3, 4, 5-9, more than 9, Not Reported, Unclear, N/A | **Note:** Must be a provider behaviour (a behaviour the provider is expected to change). 'N/A' applies to non-provider behaviours.<br>**For the purpose of this question 'behaviour' refers to:**<br>- A task, procedure or action completed by the provider to address a specific concern, condition or clinical issue.<br>- If multiple tasks, procedures or actions are considered (i.e. ordering of multiple tests) and the specifics are provided (i.e. the names of the specific tests) this should be coded as multiple behaviours.[49]<br>- If multiple tasks, procedures or actions are considered (i.e. prescribing of different antibiotics), however the specifics are not provided (i.e. the article focuses generally on prescription of antibiotics) this should be coded as one behaviour (for another example see Herbert et al.[50]).<br>- If a single task, procedure or action is considered (i.e. ordering of benzodiazepines) in multiple contexts (i.e. in combination with other pharmaceuticals, length of the prescription, and whether the benzodiazepine is long-acting),[51] still code as ONE behaviour.<br>- If the feedback addresses a guideline which involves multiple tasks, procedures or actions, follow the guidance outlined above.<br>Note: If >9 behaviours are addressed, please specify whether they are of the same general behaviour (i.e. test ordering), or a variety of different behaviours (i.e. different guidelines).<br><br>Code as **'Not reported'** if there are no details in the text and the feedback form/additional details are not available. Code as **'Unclear'** if the text contains an ambiguous statement and the feedback form/additional details are not available to confirm.<br><br>**Examples:**<br>1) Referral of smoking patients to a quit line[52] would count as 1 behaviour.<br>2) Assessing whether physicians 'ask, 'advise', 'assess' and 'assist' their smoking patients[48] would count as 4 behaviours.<br>3) Prescription of antibiotics for pediatric Upper Respiratory Infections, acute bronchitis and |

| | | | |
|---|---|---|---|
| | | | purulent rhinitis would count as 3 behaviours.<br>4) The ordering of Carcino-embryonic antigen, CA-125 and Follicle Stimulating Hormone tests[49] would count as 3 behaviours.<br>5) Ordering 2 laboratory tests per clinical issue, for 3 clinical issues would count as 6 behaviours in total (example adapted from Verstappen et al).[47]<br><br>**Note:** In some cases a proxy for behaviour may be used. For instance the number of prescriptions filled (or cost of prescriptions)[53] may be used as a proxy for prescription behaviour. Try to only code physician behaviours. However, if you include a proxy as a 'behaviour' ensure that it is clearly associated with physician behaviour. |
| 10 d) Describe | Please specify the behaviour. | Describe, Not Reported, Unclear, N/A | In a couple of words, describe the behaviour. If the feedback addresses more than one behaviour, categorize and qualitatively describe the behaviours. **Example:** Ordering of 12 lab tests (3 clinical topics, 4 lab tests per clinical topic) (example adapted from Verstappen et al.).[47]<br><br>Code as **'Not reported'** if there are no details in the text and the feedback form/additional details are not available. Code as **'Unclear'** if the text contains an ambiguous statement and the feedback form/additional details are not available to confirm. |
| 10 e) Clinical Variables | How many clinical variables were fed back to the recipients? | Describe, Not Reported, Unclear | Includes variables descriptive of the patient (i.e. LOS, mortality). Does not include behaviours.<br><br>Code as **'Not reported'** if there are no details in the text and the feedback form/additional details are not available. Code as **'Unclear'** if the text contains an ambiguous statement and the feedback form/additional details are not available to confirm. |
| 10 f) # of Graphs or Tables | How many graphs or tables are used? | Describe, Not Reported, Unclear | Code as **'Not reported'** if there are no details in the text and the feedback form/additional details are not available. Can also code if only a partial example of a FB form is given.<br>Code as **'Unclear'** if the text contains an ambiguous statement and the feedback form/additional details are not available to confirm. |
| 10 g) Graphical elements | Did the FB include any graphical elements that lend themselves to misinterpretation? (Ex. Pie charts, 3D graphs, etc.) | Yes, No, Not Reported, Unclear, N/A | Code as **'Yes'** if pie charts, 3D graphs, shadow effects etc. are used. Code as **'No'** if there are no obvious issues with the presentation of the graphs. Code as **'Not reported'** if there are no details in the text and the feedback form/additional details are not available. Code as **'Not Reported'** if the description of the intervention indicates/suggests graphs were used, but no example is provided. Code as **'Unclear'** if the text contains an ambiguous statement and the feedback form/additional details are not available to confirm. Code as **'N/A'** if graphs were not used. |

| **Suggestion #11: Address barriers to feedback use** | | | |
|---|---|---|---|
| 11 a) Barriers assessment | Were potential drivers and barriers to recipients engaging with the FB intervention assessed? | Yes, No, Not Reported, Unclear | Code as **'Yes'** if the authors explicitly state that they completed an assessment of potential drivers and barriers. **Example:** The authors report that they completed an *a priori* focus group with physicians to identify potential barriers to the recipients engaging with the feedback. Code as **'No'** if the authors do not report an assessment of potential drivers and barriers. Code as **'Not reported'** if there are no details in the text and the feedback form/additional details are not available. Code as **'Unclear'** if the text contains an ambiguous statement and the feedback form/additional details are not available to confirm whether the authors completed a drivers/barriers assessment (i.e. the authors discuss barriers identified, but it is unclear whether these barriers were identified before or after FB was given). |
| 11 b) Barriers assessment-Theory | Was the assessment informed by theory? | Yes, No, Not Reported, Unclear, N/A | Code as **'Yes'** if the authors explicitly mention that the assessment was informed by theory. Code as **'No'** if the authors indicate that the assessment was not informed by theory. Code as **'Not reported'** if there are no details in the text and the feedback form/additional details are not available. Code as **'Unclear'** if the text contains an ambiguous statement and the feedback form/additional details are not available to confirm whether the assessment was informed by theory. Code as **'N/A'** if there was no drivers/barriers assessment. |
| 11 c) Engagement assessment | Was there an assessment of whether the recipients engaged with the feedback? | Yes, No, Not Reported, Unclear | Code as **'Yes'** if:<br>• The authors clearly assessed whether the recipients engaged with the FB. For example, if the recipients were given a questionnaire asking their understanding of the FB, or if the authors were able to assess how many people clicked on a link to the FB.<br>• The authors sent a survey at the same time as the feedback and calculated a response rate.<br><br>Code as **'No'** if:<br>• The authors discuss assessing only a few of the recipient's engagement with the FB (i.e. only the team leaders).[54]<br>• The authors report a response rate from a survey, however the survey was completed at a later time.[52]<br><br>Code as **'Not reported'** if:<br>• The authors do not discuss assessing recipients' engagement with the FB; there are no details in the text and the feedback form/additional details are not available.<br><br>Code as **'Unclear'** if:<br>• There is an ambiguous statement suggesting the authors assessed recipients' engagement with the FB, however it is not entirely clear.<br>• If the feedback intervention involved a meeting component, but the authors do not explicitly state whether the engagement of recipients was assessed. |

| | | | |
|---|---|---|---|
| | | | • The authors report a response rate from a survey, however it is unclear whether the survey was sent at the same time as the FB.[52] |
| 11 d) Engagement assessment-Theory | Was the assessment informed by theory? | Yes, No, Not Reported, Unclear, N/A | Code as **'Yes'** if the authors explicitly mention that the assessment was informed by theory. Code as **'No'** if the authors indicate that the assessment was not informed by theory, or it is clear that theory was not applied (example: attendance was taken to assess engagement). Code as **'Not reported'** if there are no details in the text and the feedback form/additional details are not available. Code as **'Unclear'** if the text contains an ambiguous statement and the feedback form/additional details are not available to confirm whether the assessment was informed by theory. Code as **'N/A'** if there was no engagement assessment. |
| **Suggestion #12: Provide short, actionable messages followed by optional detail** | | | |
| 12 a) Actionable messages | Are the summary messages actionable (or described as actionable)? | Yes, No, Not Reported, Unclear, N/A | **Note: The summary message should be a summary of the data.**<br>**Code as 'Yes' if:**<br>• The FB included a summary message(s) which directed the recipient on how to proceed (i.e. the message includes a verb).<br>**Example:** 1) "You [need to **increase** your referrals by 9] to achieve the benchmark for this quarter" (adapted from Wadland et al.).[52]<br>2) Feedback included a manual which "described a number of ways to use the results including discussion with colleagues and assistants..."[55]<br>3) "A succinct evidence-based message to guide future prescribing."[50]<br>4) "CA125 should not be used to screen, diagnose, or exclude malignancy."[49]<br><br>**Code as 'No' if:**<br>• The summary message(s) did not direct recipients on how to proceed (i.e. the message did not contain a verb).<br>**Example:** 1) "You had 3 referrals for the quarter."[52] The benchmark is 9 referrals (adapted from Wadland et al.).[52]<br>2) The FB included a letter explaining that "these conditions were being evaluated because little evidence supported antibiotics for their treatment".[56]<br><br>**Code as 'Not reported' if:**<br>• There are no details in the text and the feedback form/additional details are not available.<br>**Example:** The article describes that the FB included a cover letter or a meeting discussing the FB, and it is unclear whether actionable summary messages were given.<br><br>**Code as 'Unclear' if:**<br>• The text contains an ambiguous statement and the feedback form/additional details are not available to confirm whether the summary message(s) directed the recipient on how to proceed. |

| | | | |
|---|---|---|---|
| 12 b) Additional information | Is there additional, more detailed feedback provided alongside the summary message? | Yes, No, Not Reported, Unclear | Code as **'Yes'** if the FB intervention provided additional information that was easy to navigate. **Example:** The feedback e-mail contains a hyperlink to additional information. Or, the key message of a feedback report contains the appendix page number where additional information can be found.<br><br>Code as **'No'** if the FB intervention did not provide additional information, or if the additional information is not easy to navigate. **Example:** An intervention which provides 15 pages of additional information on guidelines.<br><br>Code as **'Not reported'** if there are no details in the text and the feedback form/additional details are not available. Code as **'Unclear'** if the text contains an ambiguous statement and the feedback form/additional details are not available to confirm whether additional information was provided, or if it is unclear whether the additional information was easy to navigate. **Note:** The example in the publication by Thomas et al.[49] does not count. |
| **Suggestion #13: Address credibility of the information** | | | |
| 13 a) FB data source | Does the feedback (or the description of the FB) indicate who is providing the FB data? | Yes, No, Not Reported, Unclear | Code as **'Yes'** if:<br>• The FB form clearly indicates who is providing the data.[57]<br>• The article explicitly states that the FB recipients gave informed consent for data retrieval.<br>• The data utilized came from the hospital pharmacy or hospital computer system and it is highly likely that the physicians were aware this is where the data came from (i.e. they received print outs from the computer system).[46,53]<br>• The FB was mailed and the authors state that the FB showed who was providing the data.<br>• Feedback was provided verbally or was e-mailed.<br><br>Code as **'No'** if:<br>• The FB form does not indicate who is providing the data.<br><br>Code as **'Not reported'** if:<br>• There is no example of the feedback form available, and the article does not explain who provided the data.<br>• No example of the feedback form is provided and it is unclear if the FB form relayed who provided the data.<br>• The FB was mailed and it is unclear whether the FB form relayed who provided the data.<br><br>Code as **'Unclear'** if:<br>• The text contains an ambiguous statement and the feedback form/additional details are not available to confirm.<br><br>**Note:** For studies utilizing data from the hospital computer system, you may assume that it was clear to recipients where the data was coming from. |

| | | | |
|---|---|---|---|
| 13 b) Comparators source | Does the FB intervention indicate the source of comparators? | Yes, No, Not Reported, Unclear | Code as **'Yes'** if the FB form (or the description of the intervention) indicates all sources of comparators. **Example:** The feedback form displays "Benchmark is based on the top 10% of practices referring patients to Quit the Nic and is recalculated for each quarter."[52] Code as **'No'** if the source for one or more of the comparators is missing. For example, if the FB form shows a 'peer's' comparison, but there is no definition of whom a 'peer' is (i.e. practice, region, all participants in the study). Code as **'Not reported'** if there are no details in the text and the feedback form/additional details are not available. Code as **'Unclear'** if the text contains an ambiguous statement and the feedback form/additional details are not available to confirm whether the source for at least one comparator was provided. |
| 13 c) FB Delivery-Supervisor | Was the FB intervention delivered by a supervisor or close colleague? | Yes, No, Not Reported, Unclear | Code as **'Yes'** if the FB intervention was delivered by a supervisor or close colleague. **Example:** An intervention delivered to interns by their respective chief residents. Code as **'No'** if the FB intervention was not delivered by a supervisor or close colleague. **Example:** FB is delivered by investigators, or by an outside professional organization. Code as **'Not reported'** if there are no details in the text and the feedback form/additional details are not available. Code as **'Unclear'** if the text contains an ambiguous statement and the feedback form/additional details are not available to confirm whether the intervention was delivered by a supervisor or close colleague. |
| 13 d) FB Delivery-Organization | Was the FB intervention supported by a relevant professional organization? | Yes, No, Not Reported, Unclear | Code as **'Yes'** if the FB intervention was supported by a relevant professional organization (i.e. The FB form identifies the organization's logo,[52] or if no example of the FB form is given, the organization's support is clearly discussed in the article). **Examples:** The Department of Family Medicine - Michigan State University, Blue Cross Blue Shield of Michigan (Quit the Nic Quit Line), etc.[52]<br><br>Code as **'No'** if no professional organizations are indicated on the FB form. **Example:** The diagnostic center which provided the data facilitated the social discussions.[47] **Note:** The organization should be pinned to a recommendation.<br><br>Code as **'Not reported'** if there are no details in the text and the feedback form/additional details are not available. Code as **'Unclear'** if the text contains an ambiguous statement and the feedback form/additional details are not available to confirm whether the FB intervention was supported by a professional organization. **Example:** If the intervention included an educational component which clearly involved support from a professional organization (i.e. a brochure produced by the organization), however it is unclear whether they also provided support for the FB portion of the intervention.[56] |

| | | | |
|---|---|---|---|
| **Suggestion #14: Prevent defensive reactions to feedback** | | | |
| 14 a) Defensive reactions | Did the feedback intervention include reassurance that the intervention would not trigger punitive measures? | Yes, No, Not Reported, Unclear | Code as **'Yes'** if the FB intervention included reassurance that the intervention would not trigger punitive measures. **Example** (Yes): 1) Physicians were assured that this intervention would not involve any punitive measures. 2) In the Thomas et al. paper[49] the feedback includes the statement "It must be stressed that these requesting trend comparisons are not an exact science, hence a requesting level above or below the Grampian average does not necessarily imply appropriate or inappropriate test use."[57]<br><br>Code as **'Not reported'** if there are no details in the text and the feedback form/additional details are not available. Code as **'Unclear'** if the text contains an ambiguous statement and the feedback form/additional details are not available to confirm. |
| **Suggestion #15: Construct feedback through social interaction** | | | |
| 15 a) Social aspect- Development | Did development of the FB intervention involve members of the target group? | Yes, No, Not Reported, Unclear | Code as **'Yes'** if the development of the FB intervention involved members of the target group (**Note:** The target group refers to the FB recipients in the study- i.e. developers and recipients were co-constructing the meaning of the feedback together). **Example:** The authors completed *a priori* focus groups to determine what types of behaviour physicians wanted feedback on.<br><br>Code as **'No'** if the development of the FB intervention did not involve members of the target group.<br><br>Code as **'Not reported'** if there are no details in the text and the feedback form/additional details are not available.<br><br>Code as **'Unclear'** if the text contains an ambiguous statement and the feedback form/additional details are not available to confirm whether members of the target group were involved; or for multifaceted interventions, if it is unclear whether target members were involved in the development of the FB portion specifically. |
| 15 b) Social Context | Was the FB explicitly designed to be received and discussed in a social context? | Yes, No, Not Reported, Unclear | **Code as 'Yes' if:**<br>• There is an explicit statement that the FB intervention was designed to be received and discussed in a social context. **Note:** This may include virtual settings, if two-way communication occurs.<br><br>**Code as 'No' if:**<br>• The FB intervention was clearly not designed to be received and discussed in a social context.<br><br>**Code as 'Not reported' if:**<br>• There are no details in the text and the feedback form/additional details are not available. |

| | | | |
|---|---|---|---|
| | | | **<u>Code as 'Unclear' if:</u>**<br>• The text contains an ambiguous statement and the feedback form/additional details are not available to confirm whether the feedback was explicitly designed to be received and discussed in a social context.<br><br>**Examples**<br>**YES:** The feedback was distributed and discussed during staff meetings, and a statement describing this is included in the methods section.<br>**NO:** Feedback was mailed to recipients.<br>**UNCLEAR:** The feedback report was mailed but included materials, such as a power-point presentation, to aid in dissemination of the FB, potentially in a social context.[58]<br>"Reports were delivered individually to providers 52% of the time, delivered to the group during a provider meeting 30% of the time, and placed in the provider's mailbox for independent review 18% of the time."[48] It was unclear whether the intervention was intended to be delivered in a social context/ with a social discussion. |
| 15 c) How often? | If feedback was provided more than once, how often was feedback received and discussed in a social context? | Every time FB was provided, Only once, Variably, Not Reported, Unclear, Other (please specify), N/A | |
| 15 d) Facilitation | If FB was designed to be received and discussed in a social context, was the FB discussion facilitated by a facilitator? | Yes, No, Not Reported, Unclear, N/A | **<u>Code as 'Yes' if:</u>**<br>• The FB intervention involved at least one instance of a facilitated social discussion about the FB.<br>**Note:** A facilitator may be either internal to the group/institution, or from an external party.<br>**<u>Code as 'No' if:</u>**<br>• The FB intervention involved one or more instances of a social discussion (about the FB), but none of the sessions involved a facilitator.<br>• The FB intervention involved another component (i.e. education) involving one or more instances of a facilitated social discussion, however the discussion was not about the FB.<br>**<u>Code as 'Not Reported' if:</u>**<br>• There are no details in the text and the feedback form/additional details are not available.<br>**<u>Code as 'Unclear' if:</u>**<br>• The text contains an ambiguous statement and the feedback form/additional details are not available to confirm whether a facilitator was present, or it is unclear whether the discussion was about the FB.<br>**<u>Code as 'N/A' if:</u>**<br>• The FB intervention was not designed to be received and discussed in a social context. |

| | | | |
|---|---|---|---|
| 15 e) Self-assessment | Did the FB intervention involve engaging in self-assessment around target behaviours prior to receiving FB? | Yes, No, Not Reported, Unclear | **Code as 'Yes' if:**<br>• The article discusses a session/survey, which occurred prior to the dissemination of FB and the authors explicitly state that it involved recipient self-assessment of the target behaviours.<br>• The intervention involved an educational component.<br>**Code as 'No' if:**<br>• The article discusses a session/survey which involved recipient self-assessment of the target behaviours, however the session/survey took place after the dissemination of FB.<br>**Code as 'Not reported' if:**<br>• There are no details in the text and the feedback form/additional details are not available.<br>• The article discusses a session/survey, which occurred prior to the dissemination of FB, however there is no indication that it involved recipient self-assessment of the target behaviours.<br>**Code as 'Unclear' if:**<br>• The text contains an ambiguous statement and the feedback form/additional details are not available to confirm.<br>• The article discusses a session/survey which occurred prior to the dissemination of the FB, but it is unclear whether the session/survey involved recipient self-assessment of the target behaviours.<br>• The article discusses a session/survey that involved recipient self-assessment of the target behaviours, however it is unclear whether this session/survey took place prior to the dissemination of FB.<br><br>**Example:**<br>**YES** 1) "A pretest was completed by all 50 medical residents prior to distribution of the first newsletter... three questions were asked about individual prescribing charges, five concerning specific prescriptions, and six regarding laboratory testing."[53] |
| 15 f) FB on FB | Did the investigators actively seek FB from the recipients on the FB? | Yes, No, Not Reported, Unclear | **Code as 'Yes' if:**<br>• The recipients of the FB intervention were able to provide feedback on the feedback intervention, to the developers (i.e. through a survey, interview, etc.).<br>• The authors of the article discuss a pilot feedback intervention (not necessarily with the FB recipients from the current study) and explicitly state that the recipients received their own data.<br>• The FB report requires a response from the recipient, i.e. whether they agree with the FB.[46]<br>**Code as 'No' if:**<br>• There is no indication that recipients of the FB intervention were able to provide feedback on the feedback intervention to the developers; and there is no indication that a pilot study was completed using the participants' own data.<br>• The authors of the article discuss a pilot feedback intervention, however the FB did not use the recipient's own data (i.e. a sample form was used).<br>**Code as 'Not reported' if:**<br>• There are no details in the text and the feedback form/additional details are not available. |

| | | | |
|---|---|---|---|
| | | | **<u>Code as 'Unclear' if:</u>**<br>• The text contains an ambiguous statement and the feedback form/additional details are not available to confirm.<br>• The authors discuss completing a post-study interview/survey with the FB recipients, however it is unclear whether they were specifically asked about the FB form.[48]<br>• The authors of the article discuss a pilot feedback intervention, however they do not explicitly state whether the recipients received their own data.<br><br>**Examples:**<br>**YES**<br>1) "Participants who received portraits [feedback] individually completed a brief questionnaire concerning their understanding of the portraits [feedback] and any suggestions for improvement."[50]<br>2) "The physician was required to respond to each item on the report, indicating what action, if any, should be taken. These responses were limited to (1) rescheduling the patient sooner, (2) marking the clinic encounter form... to suggest the physician perform the indicated preventative care on the patient's next scheduled visit, (3) indicating the protocol was not applicable for this patient (but the physician agrees, in principle, with the protocol), (4) stopping this reminder (the physician disagrees with the protocol), and (5) pulling the patient's chart for review."[46]<br><br>**UNCLEAR**<br>1) "Efforts were made to educate providers on the new quitline numbers; however, despite this retraining, **<u>post-study interviews</u>** revealed that many providers were not aware that this service remained available to all clinic patients."[48]<br>2) "Two hundred physicians in 28 PBSGs [Practice-Based Small Groups] joined the trial, with an additional PBSG serving an advisory role to pre-test the materials";[50] Not clear whether it was their data or a sample form. |

**Appendix D: Pilot Inter-Rater Reliability**

| Overall Agreement | | Madison | | | | |
|---|---|---|---|---|---|---|
| | | Yes | No | Unclear | N/A | Total |
| Maria | Yes | 108 | 9 | 16 | 4 | 137 |
| | No | 21 | 183 | 19 | 22 | 245 |
| | Unclear | 9 | 20 | 22 | 2 | 53 |
| | N/A | 0 | 2 | 3 | 24 | 29 |
| | Total | 138 | 214 | 60 | 52 | 464 |

**Total Agreement (n) =** 108+ 183+22+24= 337

**Total Agreement (%) =** (337/464) x 100 = 72.6%

| Code | Calculation | Expected Frequency |
|---|---|---|
| Yes | (137 x 138)/464 | 40.7 |
| No | (214 x 245)/464 | 113.0 |
| Unclear | (60 x 53)/464 | 6.9 |
| N/A | (52 x 29)/464 | 3.25 |
| Sum | 40.7+ 113.0 + 6.9 +3.25 | 163.85 |

**Cohen's Kappa**

K = (Σ agreement -Σ expected frequency)/(N-Σ expected frequency)

K= (337-163.85)/(464-163.85)

K= 0.58

## Appendix E: Research Ethics Board Approval Letter

Ottawa Hospital Research Institute
Institut de recherche de l'Hôpital d'Ottawa

uOttawa

UNIVERSITY OF OTTAWA HEART INSTITUTE
INSTITUT DE CARDIOLOGIE DE L'UNIVERSITÉ D'OTTAWA

*Ottawa Health Science Network Research Ethics Board/ Conseil d'éthique de la recherche du Réseau de science de la santé d'Ottawa*
Civic Box 411 725 Parkdale Avenue, Ottawa, Ontario K1Y 4E9 613-798-5555 ext. 14902 Fax : 613-761-4311
http://www.ohri.ca/ohsn-reb

April 19, 2017

Dr. Jamie Brehaut
Ottawa Hospital - General Campus
Clinical Epidemiology Program
Centre for Practice- Changing Research (CPCR)
501 Smyth Rd, Room L1284, Box 201 B
Ottawa, ON K1H 8L6

Dear Dr. Brehaut:

| | | |
|---|---|---|
| **Re: Protocol #** | **20160951-01H** | **Evaluation of Feedback Interventions for the Reduction of Inappropriate Laboratory Tests and Transfusions in Intensive Care Units** |
| **Protocol approval valid until -** | **April 18, 2018** | |

I am pleased to inform you that this protocol underwent expedited review by the Ottawa Health Science Network Research Ethics Board (OHSN-REB) and is approved. No changes, amendments or addenda may be made to the protocol or the consent form without the OHSN-REB's review and approval.

Approval is for the following documents:
- Protocol version #1 dated December 6, 2016
- English recruitment e-mail version #2 dated March 15, 2017
- English reminder recruitment e-mail version #1 dated March 15, 2017

The REB no longer requires a 'valid until' date at the bottom of all approved informed consent forms. The consent forms currently approved for use by the REB are listed above.

If the study is to continue beyond the expiry date noted above, a Renewal Form should be submitted to the REB approximately six weeks prior to the current expiry date. If the study has been completed by this date, a Termination Report should be submitted.

OHSN-REB complies with the membership requirements and operates in compliance with the Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans; the International Conference on Harmonization - Good Clinical Practice: Consolidated Guideline; and the provisions of the Personal Health Information Protection Act 2004.

## References


1. Brehaut JC, Colquhoun HL, Eva KW, et al. Practice Feedback Interventions: 15 Suggestions for Optimizing Effectiveness. *Ann Intern Med*. 2016;164(6):435-441. doi:10.7326/M15-2248.

2. Foster M, Presseau J, McCleary N, Carroll K, Brehaut J. The Use of Audit and Feedback Interventions for the Improvement of Laboratory Test and Transfusion Ordering in Intensive Care Units: A Systematic Review. *(In Prep*.

3. Ivers N, Jamtvedt G, Flottorp S, et al. Audit and Feedback : Effects on Professional Practice and Healthcare Outcomes ( Review ). *Cochrane Database Syst Rev*. 2012;(6):1-227. doi:10.1002/14651858.CD000259.pub3.

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. *J Gen Intern Med*. 2014;29(11):1534-1541. doi:10.1007/s11606-014-2913-y.

5. Bjurling-Sjöberg P, Wadensten B, Pöder U, Jansson I, Nordgren L. Balancing intertwined responsibilities: A grounded theory study of teamwork in everyday intensive care unit practice. *J Interprof Care*. 2017;31(2):233-244. doi:10.1080/13561820.2016.1255184.

6. Alameddine M, Dainty KN, Deber R, Sibbald WJ (Bill). The intensive care unit work environment: Current challenges and recommendations for the future. *J Crit Care*. 2009;24(2):243-248. doi:10.1016/j.jcrc.2008.03.038.

7. Nilsen P, Roback K, Broström A, Ellstrøm P-E. Creatures of habit: accounting for the role of habit in implementation research on clinical behavior change. *Implement Sci*. 2012;7:53. doi:10.1186/1748-5908-7-53.

8. Centre for Reviews and Dissemination- University of York. PROSPERO International Prospoective Register for Systematic Reviews- Guidance notes for registering a systematic review protocol with PROSPERO. 2016:1-23. www.crd.york.ac.uk/prospero.

9. Moher D, Liberati A, Tetzlaff J, Altman D, The PRISMA Group. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLoS Med*. 2009;6(7):e1000097. doi:10.1371/journal.pmed1000097.

10. Merlani P, Garnerin P, Diby M, Ferring M, Ricou B. Quality Improvement Report: Linking Guideline To Regular Feedback To Increase Appropriate Requests For Clinical Tests: Blood Gas Analysis In Intensive Care. *BMJ*. 2001;323(7313):620-624. doi:10.1136/bmj.323.7313.620.

11. Beland D, D'Angelo C, Vinci D. Reducing unnecessary blood work in the neurosurgical ICU. *J Neurosci Nurs*. 2003;35(3):149-152. http://search.ebscohost.com/login.aspx?direct=true&db=cin20&AN=106867699&site=ehost-live.

12. Hendryx MS, Fieselmann JF, Bock MJ, Wakefield DS, Helms CM, Bentler SE. Outreach Education to Improve Quality of Rural ICU Care: Results of a Randomized Trial. *Am J Respir Crit Care Med*. 1998;158(2):418-423.

13. Paes BA, Modi A, Dunmore R. Changing physicians' behavior using combined strategies and an evidence-based protocol. *Arch Pediatr Adolesc Med*. 1994;148(12):1277-1280. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med3&NEWS=N&AN=7951806.

14. Wisser D, Van Ackern K, Knoll E, Wisser H, Bertsch T. Blood Loss from Laboratory Tests. *Clin Chem*. 2003;49(10):1651-1655.

15. Diby M, Merlani P, Garnerin P, Ricou B. Harmonization of practice among different groups of caregivers: a guideline on arterial blood gas utilization. *J Nurs Care Qual*. 2005;20(4):327-334. http://search.ebscohost.com/login.aspx?direct=true&db=cin20&AN=106540547&site=ehost-live.

16. Calderon-Margalit R, Mor-Yosef S, Mayer M, Adler B, Shapira SC. An administrative intervention to improve the utilization of laboratory tests within a university hospital. *Int J Qual Heal Care*. 2005;17(3):243-248. doi:10.1093/intqhc/mzi025.

17. Petäjä J, Andersson S, Syrjälä M. A simple automatized audit system for following and managing practices of platelet and plasma transfusions in a neonatal intensive care unit. *Transfus Med*. 2004;14(4):281-288. doi:10.1111/j.0958-7578.2004.00515.x.

18. Gutsche JT, Kornfield ZN, Speck RM, Patel PA, Atluri P, Augoustides JG. Impact of Guideline Implementation on Transfusion Practices in a Surgical Intensive Care Unit. *J Cardiothorac Vasc Anesth*. 2013;27(6):1189-1193. doi:10.1053/j.jvca.2013.05.040.

19. Yeh DD, Naraghi L, Larentzakis A, et al. Peer-to-peer physician feedback improves adherence to blood transfusion guidelines in the surgical intensive care unit. *J Trauma Acute Care Surg*. 2015;79(1):65-70. doi:10.1097/TA.0000000000000683.

20. Borgert M, Binnekade J, Paulus F, Goossens A, Vroom M, Dongelmans D. Timely individual audit and feedback significantly improves transfusion bundle compliance—a comparative study. *Int J Qual Heal Care*. 2016;28(5):601-607. doi:10.1093/intqhc/mzw071.

21. Masud F, Larson-Pollock K, Leveque C, Vykoukal D. Establishing a Culture of Blood Management Through Education: A Quality Initiative Study of Postoperative Blood Use in CABG Patients at Methodist DeBakey Heart & Vascular Center. *Am J Med Qual*. 2011;26(5):349-356. http://search.ebscohost.com/login.aspx?direct=true&db=cin20&AN=104577658&site=ehost-live.

22. Beaty CA, Haggerty KA, Moser MG, et al. Disclosure of Physician-Specific Behavior Improves Blood Utilization Protocol Adherence in Cardiac Surgery. *Ann Thorac Surg*. 2013;96(6):2168-2174. doi:10.1016/j.athoracsur.2013.06.080.

23. Arnold DM, Lauzier F, Whittingham H, et al. A multifaceted strategy to reduce inappropriate use of frozen plasma transfusions in the intensive care unit. *J Crit Care*. 2011;26(6):636.e7-636.e13. doi:10.1016/j.jcrc.2011.02.005.

24. Solomon RR, Clifford JS, Gutman SI. The Use of Laboratory Intervention to Stem the Flow of Fresh-Frozen Plasma. *Am J Clin Pathol*. 1988;89(4):518-521.

25. Murphy DJ, Lyu PF, Gregg SR, et al. Using Incentives to Improve Resource Utilization: A Quazi-Experimental Evaluation of an ICU Quality Improvement Program. *Crit Care Med*. 2016;44(1):162-170. doi:10.1097/ccm.0000000000001395.

26. Schramm GE, Kashyap R, Mullon JJ, Gajic O, Afessa B. Septic shock: A multidisciplinary response team and weekly feedback to clinicians improve the process of care and mortality. *Crit Care Med*. 2011;39(2):252-258. doi:10.1097/CCM.0b013e3181ffde08.

27. Hallgren KA. Computing Inter-Rater Reliability for Observational Data: An Overview and Tutorial. *Tutor Quant Methods Psychol*. 2012;8(1):23-34. doi:10.20982/tqmp.08.1.p023.

28. Eccles M, Grimshaw J, Walker A, Johnston M, Pitts N. Changing the Behavior of Healthcare Professionals: The Use of Theory in Promoting the Uptake of Research Findings. *J Clin Epidemiol*. 2005;58(2):107-112. doi:10.1016/j.jclinepi.2004.09.002.

29. Craig P, Dieppe P, Macintyre S, Mitchie S, Nazareth I, Petticrew M. Developing and evaluating complex interventions: The new Medical Research Council guidance. *BMJ*. 2008;337(a1655). doi:10.1136/bmj.a1655.

30. Colquhoun HL, Brehaut JC, Sales A, et al. A systematic review of the use of theory in randomized controlled trials of audit and feedback. *Implement Sci*. 2013;8:66. doi:10.1186/1748-5908-8-66.

31. Atkins L, Francis J, Islam R, et al. A guide to using the Theoretical Domains Framework of behaviour change to investigate implementation problems. *Implement Sci*. 2017;12(1):77. doi:10.1186/s13012-017-0605-9.

32. Grol RP, Bosch MC, Hulscher ME, Eccles M, Wensing M. Planning and Studying Improvement in Patient Care: The Use of Theoretical Perspectives. *Milbank Q*. 2007;85(1):93-138.

33. Davies P, Walker A, Grimshaw J. 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. *Implement Sci*. 2010;5:14.

34. French SD, Green SE, O'Connor DA, et al. Developing theory-informed behaviour change interventions to implement evidence into practice: a systematic approach using the Theoretical Domains Framework. *Implement Sci*. 2012;7(1):38. doi:10.1186/1748-5908-7-38.

35. Birken SA, Powell BJ, Shea CM, et al. Criteria for selecting implementation science theories and frameworks: results from an international survey. *Implement Sci*. 2017;12(1):124. doi:10.1186/s13012-017-0656-y.

36. Birken SA, Rohweder CL, Powell BJ, et al. T-CaST: an implementation theory comparison and selection tool. *Implement Sci*. 2018;13:143.

37. Davis D a, Mazmanian PE, Fordis M, Harrison R Van, Thorpe KE, Perrier L. Accuracy of Physician Self-assessment Compared With Observed Measures of Competence. *JAMA*. 2006;296(9):1094-1102. doi:10.1001/jama.296.9.1094.

38. Hysong ST. Meta-Analysis: Audit and Feedback Features Impact Effectiveness on Care Quality. *Med Care*. 2009;47(3):356-363.

39. Sinuff T, Muscedere J, Rozmovits L, Dale CM, Scales DC. A Qualitative Study of the Variable Effects of Audit and Feedback in the ICU. *BMJ Qual Saf*. 2015;24(6):393-399. doi:10.1136/bmjqs-2015-003978.

40. Webster F, Patel J, Rice K, et al. How to Make Feedback More Effective? Qualitative Findings from Pilot Testing of an Audit and Feedback Report for Endoscopists. *Can J Gastroenterol Hepatol*. 2016;2016:4983790. doi:10.1155/2016/4983790.

41. Kluger AN, Van Dijk D. Feedback, the various tasks of the doctor, and the feedforward alternative. *Med Educ*. 2010;44(12):1166-1174. doi:10.1111/j.1365-2923.2010.03849.x.

42. Potthoff S, Presseau J, Sniehotta FF, Johnston M, Elovainio M, Avery L. Planning to be routine : habit as a mediator of the planning-behaviour relationship in healthcare professionals. *Implement Sci*. 2017;12(1):24. doi:10.1186/s13012-017-0551-6.

43. Colquhoun H, Michie S, Sales A, et al. Reporting and design elements of audit and feedback interventions: A secondary review [published online Jan 25 2016]. *BMJ Qual Saf*. 2016. doi:10.1136/bmjqs-2015-005004.

44. DeVon HA, Block ME, Moyle-Wright P, et al. A Psychometric Toolbox for Testing Validity and Reliability. *J Nurs Scholarsh*. 2007;39(2):155-164.

45. Rycroft-Malone J, Harvey G, Seers K, Kitson A, McCormack B, Titchen A. An exploration of the factors that influence the implementation of evidence into practice. *J Clin Nurs*. 2004;13(8):913-924. doi:10.1111/j.1365-2702.2004.01007.x.

46. Tierney WM, Hui SL, McDonald CJ. Delayed Feedback of Physician Performance Versus Immediate Reminders to Perform Preventative Care: Effects on Physician Compliance. *Med Care*. 1986;24(8):659-666.

47. Verstappen WHJM, Van Der Weijden T, Dubois WI, et al. Improving Test Ordering in Primary Care: The Added Value of a Small-Group Quality Improvement Strategy Compared With Classic Feedback Only. *Ann Fam Med*. 2004;2(6):569-575. doi:10.1370/afm.244.

48. Bentz CJ, Bayley KB, Bonin KE, et al. Provider feedback to improve 5A's tobacco cessation in primary care: A cluster randomized clinical trial. *Nicotine Tob Res*. 2007;9(3):341-349. doi:10.1080/14622200701188828.

49. Thomas RE, Croal BL, Ramsay C, Eccles M, Grimshaw J. Effect of enhanced feedback and brief educational reminder messages on laboratory test requesting in primary care: a cluster randomised trial. *Lancet*. 2006;367(9527):1990-1996. doi:10.1016/S0140-6736(06)68888-0.

50. Herbert CP, Wright JM, Maclure M, et al. Better Prescribing Project: a randomized controlled trial of the impact of case-based educational modules and personal prescribing feedback on prescribing for hypertension in primary care. *Fam Pract*. 2004;21(5):575-581. doi:10.1093/fampra/cmh515.

51. Pimlott NJG, Hux JE, Wilson LM, Kahan M, Li C, Rosser WW. Educating Physicians to Reduce Benzodiazepine Use by Elderly Patients: A Randomized Controlled Trial. *CMAJ*. 2003;168(7):835-839.

52. Wadland WC, Holtrop JS, Weismantel D, Pathak PK, Fadel H, Powell J. Practice-Based Referrals to a Tobacco Cessation Quit Line: Assessing the Impact of Comparative Feedback vs General Reminders. *Ann Fam Med*. 2007;5(2):135-142. doi:10.1370/afm.650.

53. Hershey CO, Goldberg HI, Cohen DI. The Effect of Computerized Feedback Coupled with a Newsletter upon Outpatient Prescribing Charges: A Randomized Controlled Trial. *Med Care*. 1988;26(1):88-94.

54. Cohen DI, Jones P, Littenberg B, Neuhauser D. Does Cost Information Availability Reduce Physician Test Usage?: A Randomized Clinical Trial with Unexpected Findings. *Med Care*. 1982;20(3):286-292. doi:10.2307/3764297.

55. Vingerhoets E, Wensing M, Grol R. Feedback of patients' evaluations of general practice care: a randomised trial. *Qual Heal Care*. 2001;10:224-228. doi:10.1136/qhc.0100224..

56. Mainous AG, Hueston WJ, Love MM, Evans ME, Finger R. An Evaluation of Statewide Strategies to Reduce Antibiotic Overuse. *Fam Med*. 2000;32(1):22-29.

57. Thomas RE, Croal BL, Ramsay C, Eccles M, Grimshaw J. Example Feedback Report from "Effect

of enhanced feedback and brief educational reminder messages on laboratory test requesting in primary care: a cluster randomised trial." *Obtained from authors*. 2002.

58. Beck CA, Richard H, Tu J V, Pilote L. Administrative Data Feedback for Effective Cardiac Treatment: AFFECT, A Cluster Randomized Trial. *JAMA*. 2005;294(3):309-317. doi:10.1001/jama.294.3.309.

## Chapter 4: Integration and Discussion of Results

We hypothesized that A&F could be a useful but potentially underutilized behaviour change intervention in the critical care setting, to modify test and transfusion ordering. To assess the various ways in which this class of intervention had been applied, and to summarize the evidence on A&F's effectiveness in this setting, we conducted a systematic review of the literature. Our systematic review revealed that a relatively small number of published studies focus on testing A&F for changing critical care test and transfusion ordering. The studies identified showed that A&F is often combined with additional quality improvement interventions, usually having an overall modest, positive effect on the behaviour of interest (however, similar to that of the Cochrane review,[131] effects ranged from substantially positive to, in one case, negative effects). To take a more in-depth look at the A&F methods applied, and to assess how these compared to recently published guidance, we developed an evaluation tool to assess the extent to which existing A&F interventions in ICU adhered to recommended best practice.[135] We found that many of the suggested best practices[135] were under-applied in this setting, suggesting these A&F methods warrant further investigation.

We chose to focus on the improvement of two commonly ordered and inter-related resources in the critical care setting; laboratory tests and transfusions.[30,50,54,88,121,124,161,162] While A&F is often a relatively feasible and sustainable behaviour change intervention due to availability of electronic data,[135,138] our systematic review identified only a small number of studies. This may be partially due to the nature of quality improvement interventions; this work is often considered part of usual care and does not always require formal Research Ethics Board (REB) oversight.[163] These projects may not be published as often, or may be more difficult to identify in electronic databases.[164,165] As with any setting, challenges to implementing changes in critical care have been previously flagged, including limited resources, heavy workload, lack of time, and the critical condition of the patient population.[143,145] This setting may also present unique challenges due to the vulnerability of the patient population and the considerable amount of data clinicians are already required to register.[116]

Our relatively small yield of articles may also represent the difficulty in defining and addressing the issue of inappropriate ordering. It appears that many implementation interventions in the ICU tend to involve use of a protocol, clinical practice guideline (CPG), bundle and/or checklist and have most commonly focused on improving processes or outcomes for ventilator associated pneumonia, catheter related blood stream infections and sedation practices.[144] For ordering of laboratory tests, it may be difficult to develop uniform guidelines or protocols, and in turn to determine consistency with such guidelines, as 'appropriateness' can be dependent on many different factors.[27,166,167] Meanwhile, for transfusion ordering, a previous qualitative study has noted that among other factors, some Canadian ICU physicians feel further evidence to support a restrictive strategy is needed; thus, there appears to be a degree of equipoise surrounding this practice.[168]

Most of the studies in our sample reported modest, positive effects on test and transfusion ordering (improvement in the desired direction). Our findings align with the Cochrane review's overall positive results (dichotomous= 4.3%, IQR= 0.5-16%, continuous= 1.3%, IQR=1.3-28.9%).[131] Our findings are also consistent with Kobewka et al.'s review, which assessed use of various behaviour change interventions for improvement of test ordering, and found interventions with an A&F component to have a 22.0% (IQR= 8.6-34.6) median relative reduction on test use ("exclusively" A&F= 18.4%, IQR= 2.1-24.8).[82]

**Optimization of A&F and Underutilized Suggestions**[A]

As effect sizes in trials of A&F have stagnated, a greater understanding of the active mechanisms contributing to effective feedback may help to optimize this class of intervention.[135,169] There have been calls in implementation science, and for A&F specifically, to better incorporate the use of theory to do so.[128,152,170] Theory-informed interventions and study designs can help to better understand why an intervention does or does not work, and to allow for generalizations to be tested in other contexts.[128,150,152,154,171,172] The 15 suggestions for optimized feedback were developed as recommendations thought to be potentially underutilized in the current literature, and were informed by theory through qualitative interviews with theorists.[135] They target easily modifiable elements of A&F that may be associated with improved effectiveness. Development of an evaluation tool from the 15 suggestions for optimized feedback[135] allowed for a more in-depth assessment of the critical care A&F intervention sample. By assessing whether these hypothesized best practices have been applied in the critical care setting, we aimed to characterize theoretically relevant gaps in the literature. Application of the tool revealed that many of the suggestions have not been consistently applied in this context, identifying priorities for future research.

Counter to Brehaut et al.'s suggestion to "provide multiple instances of feedback", nearly a quarter of interventions (n=4) only provided feedback once, and in another 18% (n=3) of interventions it was unclear, or the feedback was only provided variably (i.e. when inappropriate orders occurred). Our findings are similar to that of a secondary analysis of the Cochrane systematic review sample of A&F trials conducted across diverse healthcare settings, which found that feedback was often only provided once (24%), or that it was not clearly stated how often the feedback was provided (24%).[160] This was an important finding, as providing multiple instances of feedback is crucial to completing the 'feedback loop'. Feedback should be iterative to allow recipients the opportunity to assess whether their efforts to change behaviour were effective.[135] Furthermore, providing feedback more than once may be helpful in bringing the behaviour back to the recipients' attention and to enable recipients to remember the suggested practice improvements.[135,173,174]

Most studies in our sample provided aggregate, unit level data, despite Brehaut et al.'s recent guidance that suggests individual data be provided whenever possible, as recipients may discount feedback providing data at the group level.[135] One study noted the difficulty in providing accurate individual data (i.e. "assigning a transfusion event to a single provider") when patients are treated by a team, and suggest that provision of group feedback may be less threatening.[175] However, two studies in our sample compared feedback with different levels of data; one study found public, individual feedback to have a greater effect than group feedback.[176] The other found group and individual feedback to be more effective than group feedback alone.[83] This is consistent with a previous meta-analysis, which reported a larger effect size for feedback providing both group and individual data, as compared to either level of data alone.[138] Based on these considerations, and the team-based nature of the ICU,[112,115] it may be of interest to further assess whether providing both individual and group data is of benefit in the context of critical care.

Another key finding was that the majority of feedback interventions did not report the incorporation of goal setting. Goal theory points to goal setting as a core means of motivating behaviour change.[137,157] Goals

[A] Note: Results from Chapter 3 are discussed at an aggregate level to maintain confidentiality of the authors who provided additional materials and details.

provide a target for recipients to work towards, and feedback provides data on performance for recipients to gauge progress in reaching their goal.[137] The salience of a goal can enhance an individual's commitment to achieving it.[157] So long as a justification is presented, assigned and self-set goals can result in similar performance effects.[157] However, care should be taken to ensure assigned goals do not require actions contradictory to existing personal goals.[157] Thirteen of the interventions in our sample did present feedback in the context of an explicit, external priority (usually at the institutional level), which could potentially help to increase perceived importance of the feedback. Further work in the healthcare and critical care setting could thus focus on elucidating differences in efficacy of feedback that incorporates internal goal setting versus externally-assigned goals, and feedback aligning with provider-identified priorities versus presenting feedback in the context of external priorities.

Only three interventions explicitly reported providing corrective actions (e.g. counseling, suggestions, making providers aware of a tool to help identify potentially inappropriate orders), in line with Brehaut et al.'s third suggestion, "Recommend specific actions".[135] Our finding that this suggestion was relatively underutilized is consistent with a previous interview study of ICU staff, which found that often times feedback does not provide appropriate guidance on how to make corrective actions.[177] Future A&F studies in this setting may wish to incorporate development of action plans, as there is robust empirical evidence to support the use of these plans for changing behaviour.[178] While relatively simple to formulate, action plans (also known as implementation intentions, or 'if-then' plans) specify what actions an individual intends to take and in what specific contexts, allowing recipients to easily recognize appropriate situations. These plans also help to make specified actions automatic.[178] Combining these plans with "explicit targets" in an A&F intervention is supported by the meta-regression results from the most recent Cochrane review.[131] Coping strategies similarly involve "linking situational cues contingently associated with undesired behaviours (if-condition) with cognitive or behavioural coping responses aimed at inhibiting the undesired, or prioritizing the desired, response".[179] A systematic review assessing the use of coping strategies to address various healthy and unhealthy behaviours found these plans could be helpful when assistance was provided to prepare the plans.[180] Results from this review also suggest coping plans enhance the effect of action plans.[180] Furthermore, providing action and coping plans to support corrective actions may be especially important in the context of targeting habitual behaviours (i.e. frequent test or transfusion ordering), as creating these plans may help to develop new, preferred habits.[80,181]

Counter to Brehaut et al.'s suggestion to "address barriers to feedback use",[135] no publications reported the use of a barriers assessment, and only two of the interventions involved assessment of recipient engagement. Completion of barrier and engagement assessments may help to increase the efficiency with which the research is conducted, by identifying issues that can be addressed when developing the A&F, or in subsequent iterations of the intervention. These assessments may also be theory-based, using frameworks such as the Theoretical Domains Framework, to ensure the process is comprehensive and systematic, and to aid in the selection of potential solutions.[182] Barrier assessments can identify unforeseen challenges,[183] and engagement assessments are important tools, as A&F is unlikely to have the intended effect if recipients do not look at, read, or engage with the feedback and associated activities (e.g. an action planning exercise).[135,184] This should be reflected in the development, evaluation and reporting of A&F interventions.

As most of the interventions were multi-component and involved an educational aspect, there was often an opportunity for recipients to engage in self-assessment around the targeted behaviour. However, none of the publications reported involving the target recipients in the development of the feedback, and details

regarding the general feedback development were rarely reported. Furthermore, few of the interventions were explicitly designed to be provided and discussed in a social setting, had a facilitator, or actively obtained feedback from the recipients. A key area for future research is thus the incorporation of social interaction in the feedback intervention, as it is suggested a more active approach may better encourage recipient learning and improve sustainability of the feedback.[135,185]

**Consistency with Remaining Suggestions**

Only one of the 15 suggestions was almost universally applied across all interventions ("Recommend actions that can improve and are under the recipients' control"[135]). Feedback was provided to the individuals responsible for ordering the tests or transfusions (or their direct supervisor), as opposed to individuals unlikely to have control over test or transfusion ordering (i.e. hospital administration). However, this may have been partially driven by the inclusion criteria used to identify the sample (requirement to report on lab test or transfusion ordering, i.e. a provider behaviour). We also note that the suggestion "Provide feedback in more than 1 way" was applied in all cases where we were able to assess this item. It is hypothesized that this practice may allow individuals to select their preferred format from the various options, and enhance formation of a mental model.[135] It may also increase the chances that a recipient pays attention to and remembers the feedback.[135] In ten of the interventions feedback was provided verbally or through e-mail, and in six cases was delivered by a supervisor or colleague, and so it was likely clear who was providing the data. This may have lent credibility to the data, however previous qualitative work has shown credibility of the audit to be a concern of ICU staff.[177] It would therefore be valuable to assess how A&F interventions can improve perceived credibility of the data. While no studies reported including reassurance that the intervention would not result in punitive measures, six reported or communicated other aspects which aimed to reduce negative reactions to the feedback (e.g. providing group level data or confidential feedback so as not to single out specific providers, providing both positive and negative feedback, framing feedback in a non-punitive way, or acknowledging that the feedback was well-received and not viewed as punitive due to a team-oriented institutional culture). However, further work is needed to identify the effectiveness of such mitigation strategies.

Our evaluation of the sample of critical care A&F interventions found that of those providing feedback more than once, the majority provided feedback monthly (n=3), monthly and variably (n=2) or weekly (n=2); in another three cases the interval was unclear/variable. The average age of the data was most commonly a combination of days and weeks (n=3) or weeks (n=3). Previous literature suggests that providing feedback soon after auditing ("timely"[186] or "frequent"[138] feedback) may be beneficial; however, a balance must be struck to ensure recipients are not burdened by too frequent feedback.[135] A process evaluation of a multi-component intervention involving A&F for ICU quality improvement, indicated most recipients (77% of survey respondents) were content with the "intensity" of the feedback (monthly and quarterly), though during a focus group several revealed that this was too frequent.[148] Conversely, an interview study which asked ICU staff about their experience with feedback to date, found "timeliness" of feedback to be a key issue among providers; one intensive care physician questioned the usefulness of delayed feedback if it does not allow for corrective actions to be taken right away.[177] ICU staff also noted the importance of providing adequate data on previous performance over time, as a cross-sectional snapshot may not provide a representative look at practice.[177] As most of the literature directly assessing the effect of timing on feedback effectiveness comes from outside of the healthcare discipline, it would be of interest to study whether tailoring this characteristic, such as based on caseload, helps to optimize feedback.[135]

There was variation in terms of whether a comparator was provided. Just over half (n=9) of the interventions clearly provided (or reported providing) a comparator. Comparators can provide perspective in terms of the distance between a recipient's performance and that of a desired level of performance (standard, goal, others' performance). Highlighting this distance is an important aspect, as the theory behind feedback predicts that the motivation to change behaviour comes from the need to reduce this distance.[137] Thus, if this distance is not clearly shown in the feedback, through use of a comparator or other means, this may indicate a critical flaw in the design. In line with our finding that goal setting was rarely incorporated, an aspirational comparator was only used in two interventions (though for two others 100% compliance may have been implied). Future studies could assess whether the use of an aspirational comparator (likely to depict a larger gap in practice as opposed to the mean) helps to improve performance.[187] Another interesting area of study would be whether explicit or implied comparators are more effective (i.e. provision of feedback over time, or an implied standard of 100% compliance).

**Challenges in Assessing Consistency with Suggestions**

Assessing consistency with criteria items related to the design and delivery of feedback proved difficult to judge, as we did not have access to the feedback form in most cases. Furthermore, several of the feedback interventions involved verbal summaries, rendering some items not possible to assess. For the most part, we were not able to determine whether the summary message was in close proximity to the visual display, or whether the messages were short, actionable and followed by optional detail. Our conclusions regarding items related to cognitive overload (factors increasing the recipient's effort to process the information) were also somewhat limited. However, none of the interventions were reported to have been pilot tested (in one case it was unclear if piloting was completed). As many of the remaining display items were difficult to determine without access to the form, this is a key finding. If not already in practice, piloting of feedback forms could help to identify problems, ensure recipients are not overwhelmed by the amount of information and display, and to increase usability of the feedback.[188–190] Conversely, if piloting of feedback forms is practiced but not reported, this signals the need for more complete or standardized reporting.

**Considerations for Future Research**

Clearly future research in this area should focus on conducting randomized, head-to-head trials.[170] Our systematic review assessing the use of A&F interventions for the improvement of laboratory test and transfusion ordering identified only one randomized, controlled trial and one head-to-head comparison study. Investigators have proposed that since a substantial, cumulative evidence base supports the positive overall effect of A&F in healthcare settings, future research should focus on assessing various approaches to operationalizing A&F by completing head-to-head comparisons (versus controlled studies) to identify the most successful methods.[150,151] Use of more rigorous study designs will help to create a strong evidence base. As randomization within centres may not always be feasible for this implementation issue, due to the potential for contamination, cluster randomized trials may also be considered.[170,191]

When conducting studies with multicomponent interventions, future studies may also consider incorporating process evaluations and study designs that allow for a better understanding of which components are contributing to the success of an intervention.[170,192] Almost all of the A&F interventions identified in our review (88%) involved one or more additional intervention components such as education, guidelines, opinion leaders, financial incentives, checklists or administrative interventions. However, study designs did not typically allow for individual components to be assessed, or provide concurrent

comparisons. It is therefore challenging to build upon studies when it is unclear which intervention components are effective and which are not. Comparison studies directly assessing different combinations of intervention components would therefore help to provide causal evidence of which combinations are best.

Moving forward, collecting patient outcomes and cost data will also be important to ensure interventions do not have any unintended consequences. Pertinent patient outcomes such as length of stay, infection and mortality were not consistently reported across the sample of critical care A&F interventions. Cost data was also not often reported, and difficult to compare across studies due to differences in how this outcome was measured (i.e. savings versus cost). Reporting of these outcomes is imperative to comparing intervention safety, feasibility and sustainability. To allow for more straight-forward comparisons in the future, studies should aim to collect and provide information on core patient outcomes and cost data. Since there are differences in how laboratory test or transfusion costs may be quantified (i.e. healthcare provider time, laboratory reagents, equipment and operating costs),[27,28,193] transparent reporting of cost calculations would also be valuable.

**The Evaluation Tool**

Assessment of whether the sample of A&F interventions had applied the suggestions published by Brehaut et al.[135] required that we deconstruct and operationalize each of the general concepts encompassed into specific, ratable items. This tool can now provide researchers and feedback developers with a resource to identify how their A&F interventions may be further optimized. As the 15 suggestions were developed to provide general guidance (as opposed to serving as prescriptive instructions), our evaluation tool does not provide an overall consistency score. Application of this tool instead provides a descriptive overview of the elements of existing A&F interventions and provides a foundation on which to conduct further research, by identifying gaps in the literature that may be tested in future trials. To fully develop the evaluation tool, future work will also focus on psychometric testing (such as assessment of construct, content and criterion validity).[194]

**Implications for Practice**

The idea for this project originated from the Eastern Ontario Laboratory Association's (EORLA) interest in providing feedback to providers. EORLA supplies laboratory services to The Ottawa Hospital (TOH), as well as other member sites throughout Eastern Ontario.[195] With the most recent meta-analysis of healthcare audit and feedback (A&F) interventions showing an overall positive effect, there is evidence that A&F can be a potentially effective behaviour change intervention.[131] However, variation in effect size across studies, from largely positive, to no change, or even negative effects demonstrates that A&F can in some cases hurt performance.[173] Thus, while there is easily accessible data for EORLA to provide feedback, there is interest in first identifying evidence and theory-informed methods to design and deliver the feedback, to increase the chances of effectiveness. The underutilized suggestions identified in our work provide priorities for future research, which will help to confirm best practices that may be applied to practice.

**Limitations & Considerations**

Most of the studies reported an absolute (albeit modest) change in the hypothesized direction. While this is in line with the overall results from the most recent Cochrane review on A&F,[131] there are a variety of factors that prevent us from making strong conclusions about the effectiveness of A&F in this setting. The majority of studies identified in our systematic review (13/16) used an uncontrolled before/after study design. This finding was consistent with previous systematic reviews, which assessed behaviour change interventions to improve laboratory test or transfusion ordering.[82,100,101] The inherent bias associated with this type of study design limits our ability to draw causal conclusions on the effectiveness of A&F in this setting. Without a consecutive control group or time series analysis, apparent benefits from the interventions may be coincidental, simply due to secular trends or other confounding factors.[191,196,197] Nevertheless, the positive signal observed in the majority (almost all) of the studies suggests further investigation is warranted.

The fact that the majority of interventions showed beneficial effects could also indicate potential publication bias. Although our systematic review of the literature was comprehensive in that we assessed five databases as well as two registries, it is likely that many quality improvement interventions are never published, regardless of whether the results are positive or negative.[164] Future work building on this study could involve a search of the grey literature to identify unpublished quality improvement projects involving A&F. Typically, a funnel plot can be used to assess the potential for publication bias.[198] However, due to heterogeneity in the reported outcomes (i.e. total number of tests, number of appropriate transfusions, percent compliance, etc.) a funnel plot for each outcome category would be needed. As Cochrane recommends at minimum data from ten studies in order to conduct asymmetry statistical tests,[198] we have not currently used funnel plots, however this could be completed in the future as more studies are conducted, identified and categorized by outcome type.

A limitation of our evaluation tool is that it relies on the details reported in manuscripts, and thus in many cases we were unable to definitively determine whether or not a criteria item had been applied. Journal word limits may restrict the amount of information provided in the methods section, and this may be an issue for quality improvement studies in particular, as these interventions typically involve multiple components requiring description. We believe a potential solution to this problem could be to provide access to the feedback template, or supplemental methods in online appendices. Reporting guidelines such as the Template for Intervention Description and Replication (TIDieR) checklist could also be used to ensure standardized reporting of intervention elements.[199] This would allow for a more in-depth study of various A&F elements in future research, and better allow for others to reproduce and build upon these interventions.

The inconsistency in terms used to describe A&F and quality improvement intervention components was another difficulty encountered throughout our review. In many cases the intervention components were poorly defined, making it challenging to identify included studies during the screening phase. Variability in the terms used to describe various quality improvement intervention components has been previously described in the literature.[200] Future A&F studies for the improvement of test and transfusion ordering in the critical care setting should aim to use standardized intervention component terms, such as those described by the Cochrane Effective Practice of Organization and Care (EPOC) taxonomy,[130] the Expert

Recommendations for Implementing Change (ERIC) project,[200] and the Behaviour Change Taxonomy (BCT).[201,202] This will help to ensure that interventions can be consistently identified and compared.

**Conclusions**

Our work suggests A&F is a potentially effective intervention for the improvement of test and transfusion ordering in the critical care setting. Moving forward, use of more rigorous study designs and efforts to improve reporting of interventions will help to advance and build upon research in this field. Development of the evaluation tool has helped to operationalize Brehaut et al.'s 15 suggestions[135] into criteria items that can be applied to A&F intervention studies, to assess compliance with best practices. By applying this tool to a sample of critical care A&F interventions, we have identified several areas of focus for future A&F research; our results may suggest that the lessons from Brehaut and colleagues are generally underutilized in this critical care A&F sample. It would be of interest to study whether these elements help to improve the effectiveness of A&F in the critical care setting, as well as to further elucidate the mechanisms by which A&F with these elements work. Our evaluation tool also provides many opportunities to conduct further research; to assess whether adherence with the suggestions improves over time, as well as application to other settings. Future work to optimize A&F interventions in the critical care setting will help to identify more effective strategies by which to improve provider laboratory test and transfusion ordering, impacting both patient care and resource use.

**References**


1. Zhi M, Ding EL, Theisen-Toupal J, Whelan J, Arnaout R. The Landscape of Inappropriate Laboratory Testing: A 15-year Meta-Analysis. *PLoS One*. 2013;8(11):1-8. doi:10.1371/journal.pone.0078962.

2. Yeh DD. Laboratory Utilization. In: Lewandrowski K, Sluss PM, eds. *Utilization Management in the Clinical Laboratory and Other Ancillary Services*. Cham: Springer International Publishing; 2017:203-210. doi:10.1007/978-3-319-34199-6_18.

3. Aschengrau A, Seage III GR. *Essentials of Epidemiology in Public Health*. 3rd ed. Burlington, MA: Jones & Bartlett Learning; 2014.

4. Manegold C. Treatment algorithm in 2014 for advanced non-small cell lung cancer: Therapy selection by tumour histology and molecular biology. *Adv Med Sci*. 2014;59(2):308-313. doi:10.1016/j.advms.2014.08.008.

5. Bassetti M, Poulakou G, Ruppe E, Bouza E, van Hal SJ, Brink A. Antimicrobial resistance in the next 30 years, humankind, bugs and drugs: a visionary approach. *Intensive Care Med*. 2017:1-12. doi:10.1007/s00134-017-4878-x.

6. Lord Carter of Coles. *Report of the Review of NHS Pathlogy Services in England*.; 2006. http://webarchive.nationalarchives.gov.uk/20130107105354/http:/www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/@dh/@en/documents/digitalasset/dh_091984.pdf.

7. Busby J, Schroeder K, Woltersdorf W, et al. Temporal growth and geographic variation in the use of laboratory tests by NHS general practices: Using routine data to identify research priorities. *Br J Gen Pract*. 2013;63(609):256-266. doi:10.3399/bjgp13X665224.

8. Thomas RE, Vaska M, Naugler C, Turin TC. Interventions at the laboratory level to reduce laboratory test ordering by family physicians: Systematic review. *Clin Biochem*. 2015;48(18):1358-1365. doi:10.1016/j.clinbiochem.2015.09.014.

9. Plebani M. Errors in Clinical Laboratories or Errors in Laboratory Medicine? *Clin Chem Lab Med*. 2006;44(6):750-759. doi:10.1515/CCLM.2006.123.

10. Emson HE. The Growth of a Laboratory. *Can Med Assoc J*. 1965;93:211-218.

11. Naugler C. Laboratory Utilization Management in Canada. In: Lewandrowski K, Sluss P, eds. *Utilization Management in the Clinical Laboratory and Other Ancillary Services*. Springer, Cham; 2017.

12. Naugler C. A Perspective on Laboratory Utilization Management from Canada. *Clin Chim Acta*. 2014;427:142-144. doi:10.1016/j.cca.2013.09.022.

13. van Walraven C, Naylor CD. Do We Know What Inappropriate Laboratory Utilization Is? A Systematic Review of Laboratory Clinical Audits. *JAMA*. 1998;280(6):550-558. doi:10.1309/AJCPX1HIEM4KLGNU.

14. The Lewin Group, Wolcott J, Schwartz A, Goodman C. *Laboratory Medicine: A National Status Report*.; 2008. http://wwwn.cdc.gov/futurelabmedicine/pdfs/2007%5Cnstatus%5Cnreport%5Cnlaboratory_medicine_-_a_national_status_report_from_the_lewin_group_updated_2008-9.pdf.

15. Eastern Ontario Regional Laboratory Association. *Annual Report 2015-16: DRAFT*.

16. Bayne L, Lillian Bayne & Associates. *BC Laboratory Services Review*.; 2003.

17. Ministry of Health and Long-Term Care. *Laboratory Services Expert Panel*.; 2015.

18. Chassin M, Galvin R. The urgent need to improve health care quality: Institute of medicine national roundtable on health care quality. *Jama*. 1998;280(11):1000-1005. doi:10.1001/jama.280.11.1000.

19. Institute of Medicine (US) Roundtable on Evidence-Based Medicine. *The Healthcare Imperative: Lowering Costs and Improving Outcomes. Workshop Series Summary.* (Yong P, Saunders R, Olsen L, eds.). Washington, DC: National Academies Press; 2010. doi:10.17226/12750.

20. Lenzer J. Unnecessary care: are doctors in denial and is profit driven healthcare to blame? *Bmj*. 2012;345(oct02 3):e6230-e6230. doi:10.1136/bmj.e6230.

21. Lown Cardiovascular Research Foundation, New America Foundation. About The Conference. Avoiding Avoidable Care Web Page. http://avoidablecare.org/about-the-conference/. Accessed October 11, 2017.

22. Wintemute K, McDonald K, Huynh T, Pendrith C, Wilson L. Addressing overuse starts with physicians. Choosing Wisely Canada. *Can Fam Physician*. 2016;62(3):199-200. doi:62/3/199 [pii].

23. Cho HJ. Countering Medicine's Culture of More. *Am Med Assoc J Ethics*. 2015;17(11):1079-1081.

24. Canadian Institute for Health Information. Canadians have more than 1 million potentially unnecessary medical tests and treatments every year. Canadian Institute for Health Information Web Page. https://www.cihi.ca/en/canadians-have-more-than-1-million-potentially-unnecessary-medical-tests-and-treatments-every-year. Published 2017. Accessed October 11, 2017.

25. Berwick DM, Hackbarth AD. Eliminating Waste in US Health Care. *JAMA*. 2012;307(14):1513-1516. doi:10.1001/jama.2012.362.

26. Canadian Institute for Health Information. Unnecessary Care in Canada: Infographic. Canadian Institute for Health Information Web Page. https://www.cihi.ca/en/unnecessary-care-in-canada-infographic. Published 2017. Accessed October 11, 2017.

27. Lewandrowski K, Baron J, Dighe A. Utilization Management in the Clinical Laboratory: An Introduction and Overview. In: Lewandrowski K, Sluss P, eds. *Utilization Management in the Clinical Laboratory and Other Ancillary Services*. Springer, Cham; 2017.

28. Sluss P. Calculating Costs and Savings in Utilization Management. In: Lewandrowski K, Sluss P, eds. *Utilization Management in the Clinical Laboratory and Other Ancillary Services*. Springer, Cham; 2017.

29. Sedrak MS, Patel MS, Ziemba JB, et al. Residents' Self-Report on Why They Order Perceived Unnecessary Inpatient Laboratory Tests. *J Hosp Med*. 2016;11(12):869-872. doi:10.1002/jhm.2645.

30. Ezzie ME, Aberegg SK, O'Brien JM. Laboratory Testing in the Intensive Care Unit. *Crit Care Clin*. 2007;23(3):435-465. doi:10.1016/j.ccc.2007.07.005.

31. Garland A, Shaman Z, Baron J, Connors AF. Physician-attributable Differences in Intensive Care

Unit Costs: A Single-Center Study. *Am J Respir Crit Care Med*. 2006;174(11):1206-1210. doi:10.1164/rccm.200511-1810OC.

32. Mindemark M, Wernroth L, Larsson A. Costly regional variations in primary health care test utilization in Sweden. *Scand J Clin Lab Invest*. 2010;70(3):164-170. doi:10.3109/00365511003642519.

33. Verstappen WHJM, ter Riet G, Dubois WI, Winkens R, Grol RP, van der Weijden T. Variation in test ordering behaviour of GPs: Professional or context-related factors? *Fam Pract*. 2004;21(4):387-395. doi:10.1093/fampra/cmh408.

34. Cadogan SL, McHugh SM, Bradley CP, Browne JP, Cahill MR. General practitioner views on the determinants of test ordering: a theory-based qualitative approach to the development of an intervention to improve immunoglobulin requests in primary care. *Implement Sci*. 2016;11(1):102. doi:10.1186/s13012-016-0465-8.

35. Schroeder SA, Kenders K, Cooper J, Piemme TE. Use of Laboratory Tests and Pharmaceuticals: Variation Among Physicians and Effect of Cost Audit on Subsequent Use. *JAMA*. 1973;225(8):969-973. doi:10.1001/jama.1973.03220360029007.

36. Signorelli H, Straseski JA, Genzen JR, Walker BS, Jackson BR, Schmidt RL. Benchmarking to Identify Practice Variation in Test Ordering: A Potential Tool for Utilization Management. *Lab Med*. 2015;46(4):356-364. doi:10.1309/LM2JVTWX8TKCURMF.

37. Zimmerman JE, Seneff MG, Sun X, Wagner DP, Knaus WA. Evaluating laboratory usage in the intensive care unit: Patient and institutional characteristics that influence frequency of blood sampling. *Crit Care Med*. 1997;25(5):737-748. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=emed7&NEWS=N&AN=27249148.

38. Newman DB, Siontis KC, Chandrasekaran K, Jaffe AS, Kashiwagi DT. Intervention to Reduce Inappropriate Ionized Calcium Ordering Practices: A Quality-Improvement Project. *Perm J*. 2015;19(1):49-51. doi:10.7812/TPP/14-108.

39. Melanson SEF, Szymanski T, Rogers SO, et al. Utilization of Arterial Blood Gas Measurements in a Large Tertiary Care Hospital. *Am J Clin Pathol*. 2007;127(4):604-609. doi:10.1309/ELH5BPQ0T17RRK0M.

40. Morgen EK, Naugler C. Inappropriate Repeats of Six Common Tests in a Canadian City: A Population Cohort Study within a Laboratory Informatics Framework. *Am J Clin Pathol*. 2015;144(5):704-712. doi:10.1309/AJCPYXDAUS2F8XJY.

41. Gill J, Barakauskas VE, Thomas D, et al. Evaluation of thyroid test utilization through analysis of population-level data. *Clin Chem Lab Med*. 2017;[Epub ahea(0). doi:10.1515/cclm-2016-1049.

42. Weydert JA, Nobbs ND, Feld R, Kemp JD. A Simple, Focused, Computerized Query to Detect Overutilization of Laboratory Tests. *Arch Pathol Lab Med*. 2005;129(9):1141-1143. doi:10.1043/1543-2165(2005)129[1141:ASFCQT]2.0.CO;2.

43. Van Walraven C, Raymond M, for the Network of Eastern Ontario Medical Laboratories (NEO-MeL). Population-based Study of Repeat Laboratory Testing. *Clin Chem*. 2003;49(12):1997-2005.

44. Sood R, Sood A, Ghosh AK. Non-Evidence-Based Variables Affecting Physicians' Test-Ordering

Tendencies: A Systematic Review. *Neth J Med*. 2007;65(5):167-177. http://www.ncbi.nlm.nih.gov/pubmed/17519512.

45. Cadogan SL, Browne JP, Bradley CP, Cahill MR. The effectiveness of interventions to improve laboratory requesting patterns among primary care physicians: a systematic review. *Implement Sci*. 2015;10:167. doi:10.1186/s13012-015-0356-4.

46. Hindmarsh JT, Lyon AW. Strategies to Promote Rational Clinical Chemistry Test Utilization. *Clin Biochem*. 1996;29(4):291-299.

47. Brateanu A, Rothberg MB. Why do clinicians continue to order "routine preoperative tests" despite the evidence? *Cleve Clin J Med*. 2015;82(10):667-670. doi:10.3949/ccjm.82a.15118.

48. Miyakis S, Karamanof G, Liontos M, Mountokalakis TD. Factors contributing to inappropriate ordering of tests in an academic medical department and the effect of an educational feedback strategy. *Postgrad Med J*. 2006;82(974):823-829. doi:10.1136/pgmj.2006.049551.

49. Kahan NR, Waitman D-A, Vardy D. Curtailing Laboratory Test Ordering in a Managed Care Setting Through Redesign of a Computerized Order Form. *Am J Manag Care*. 2009;15(3):173-176. http://www.embase.com/search/results?subaction=viewrecord&from=export&id=L70015831%5Cnhttp://dx.doi.org/10.1111/j.1524-4733.2009.00537-2.x%5Cnhttp://sfx.library.uu.nl/utrecht?sid=EMBASE&issn=10983015&id=doi:10.1111/j.1524-4733.2009.00537-2.x&atitle=Curtai.

50. Raad S, Elliott R, Dickerson E, Khan B, Diab K. Reduction of Laboratory Utilization in the Intensive Care Unit. *J Intensive Care Med*. 2016:885066616651806. doi:10.1177/0885066616651806.

51. Civetta JM, Hudson-Civetta J a. Maintaining Quality of Care While Reducing Charges in the ICU: Ten ways. *Ann Surg*. 1985;202(4):524-532. doi:10.1097/00132586-198608000-00041.

52. Hickner J, Thompson PJ, Wilkinson T, et al. Primary Care Physicians' Challenges in Ordering Clinical Laboratory Tests and Interpreting Results. *J Am Board Fam Med*. 2014;27(2):268-274. doi:10.3122/jabfm.2014.02.130104.

53. Zimmerman JE, Seneff MG, Sun X, Wagner DP, Knaus WA. Evaluating Laboratory Usage in the Intensive Care Unit: Patient and Institutional Characteristics that Influence Frequency of Blood Sampling. *Crit Care Med*. 1997;25(5):737-748.

54. Cismondi F, Celi LA, Fialho AS, et al. Reducing Unnecessary Lab Testing in the ICU with Artificial Intelligence. *Int J Med Inform*. 2013;82(5):345-358. doi:10.1016/j.ijmedinf.2012.11.017.

55. Critical Care Societies Collaborative. Five Things Physicians and Patients Should Question. Choosing Wisely. http://www.choosingwisely.org/societies/critical-care-societies-collaborative-critical-care/. Published 2014. Accessed November 18, 2016.

56. Flabouris A, Bishop G, Williams L, Cunningham M. Routine Blood Test Ordering for Patients in Intensive Care. *Anaesth Intensive Care*. 2000;28(5):562-565.

57. Jakacka N, Snarski E, Mekuria S. Prevention of Iatrogenic Anemia in Critical and Neonatal Care. *Adv Clin Exp Med*. 2016;25(1):191-197. doi:10.17219/acem/32065.

58. Colla CH, Kinsella EA, Morden NE, Meyers DJ, Rosenthal MB, Sequist TD. Physician

Perceptions of Choosing Wisely and Drivers of Overuse. *Am J Manag Care*. 2016;22(5):337-343.

59. Axt-Adam P, van der Wouden JC, van der Does E. Influencing Behavior of Physicians Ordering Laboratory Tests : A Literature Study. *Med Care*. 2017;31(9):784-794.

60. Schattner A. The unbearable lightness of diagnostic testing: time to contain inappropriate test ordering. *Postgrad Med J*. 2008;84(998):618-621. doi:10.1136/pgmj.2008.072579.

61. Van Der Weijden T, Van Bokhoven MA, Dinant GJ, Van Hasselt CM, Grol RPTM. Understanding laboratory testing in diagnostic uncertainty: A qualitative study in general practice. *Br J Gen Pract*. 2002;52(485):974-980.

62. Levy MM, Dellinger RP, Townsend SR, et al. The Surviving Sepsis Campaign: Results of an international guideline-based performance improvement program targeting severe sepsis. *Intensive Care Med*. 2010;36(2):222-231. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med5&NEWS=N&AN=20069275.

63. Levy MM, Rhodes A, Phillips GS, et al. Surviving Sepsis Campaign: Association Between Performance Metrics and Outcomes in a 7.5-Year Study. *Crit Care Med*. 2015;43(1):3-12. doi:10.1097/CCM.0000000000000723.

64. Canadian Medical Association, University of Toronto. What is CWC? Choosing Wisely Canada Website. http://www.choosingwiselycanada.org/about/what-is-cwc/. Accessed March 22, 2017.

65. The American Board of Internal Medicine Foundation. About. Choosing Wisely Website. http://www.choosingwisely.org/about-us/. Published 2017. Accessed April 10, 2017.

66. Canadian Medical Association, University of Toronto. Choosing Wisely Canada- The Lists. Choosing Wisely Canada Website. http://www.choosingwiselycanada.org/recommendations/. Accessed April 10, 2017.

67. CMA's Forum on General and Family Practice Issues, College of Family Physicians of Canada. Eleven Things Physicians and Patients Should Question. Choosing Wisely Canada Website. http://www.choosingwiselycanada.org/recommendations/family-medicine/. Published 2014. Accessed April 10, 2017.

68. Canadian Anesthesiologists' Society. Five Things Physicians and Patients Should Question. Choosing Wisely Canada Website. http://www.choosingwiselycanada.org/recommendations/anesthesiology/. Published 2015. Accessed April 10, 2017.

69. Canadian Society of Endocrinology and Metabolism. Five Things Physicians and Patients Should Question. Choosing Wisely Canada Website. http://www.choosingwiselycanada.org/recommendations/endocrinology-and-metabolism/. Published 2014. Accessed April 10, 2017.

70. Canadian Federation of Medical Students, Fédération médicale étudiante du Québec. Six Things Medical Students and Trainees Should Question. Choosing Wisely Canada Website. http://www.choosingwiselycanada.org/recommendations/medical-students-and-trainees/. Published 2015. Accessed April 10, 2017.

71. Canadian Association of Pathologists. Five Things Physicians and Patients Should Question.

Choosing Wisely Canada Website. http://www.choosingwiselycanada.org/recommendations/pathology/. Published 2014. Accessed April 10, 2017.

72. National Institute for Health and Care Excellence. *Routine Preoperative Tests for Elective Surgery*.; 2016. https://www.nice.org.uk/guidance/ng45.

73. Verbrugghe S. Initiatives to Optimize the Utilization of Laboratory Tests. Ottawa: Canadian Agency for Drugs and Technolgies in Health. 2014;(Environmental Scan; issue 44).

74. Ontario Health Technology Advisory Committee. *Appropriateness Phase 1 OHTAC Recommendations: Annual Health Exams, Aspartate Aminotransferase Testing, Chloride Testing, Creatine Kinase Testing, Ferritin Testing, Folate Testing, and Vitamin B12 Testing*.; 2013. http://www.hqontario.ca/Portals/0/Documents/evidence/reports/recommendation-appropriateness-phase-1-130722-en.pdf.

75. Hong AS, Ross-Degnan D, Zhang F, Wharam JF. Small Decline In Low-Value Back Imaging Associated With The "Choosing Wisely" Campaign, 2012–14. *Health Aff*. 2017;36(4):671-679. doi:10.1377/hlthaff.2016.1263.

76. Rosenberg A, Agiro A, Gottlieb M, et al. Early Trends Among Seven Recommendations From the Choosing Wisely Campaign. *JAMA Intern Med*. 2015;175(12):1913-1920. doi:10.1001/jamainternmed.2015.5441.

77. Grover M, Abraham N, Chang Y-H, Tilburt J. Physician Cost Consciousness and Use of Low-Value Clinical Services. *J Am Board Fam Med*. 2016;29(6):785-792. doi:10.3122/jabfm.2016.06.160176.

78. Cabana MD, Rand CS, Powe NR, et al. Why Don't Physicians Follow Clinical Practice Guidelines? A Framework for Improvement. *JAMA*. 1999;282(15):1458-1465. doi:10.1001/jama.282.15.1458.

79. Zikmund-Fisher BJ, Kullgren JT, Fagerlin A, Klamerus ML, Bernstein SJ, Kerr EA. Perceived Barriers to Implementing Individual Choosing Wisely Recommendations in Two National Surveys of Primary Care Providers. *J Gen Intern Med*. 2017;32(2):210-217. Epub 2016. doi:10.1007/s11606-016-3853-5.

80. Nilsen P, Roback K, Broström A, Ellstrøm P-E. Creatures of habit: accounting for the role of habit in implementation research on clinical behavior change. *Implement Sci*. 2012;7:53. doi:10.1186/1748-5908-7-53.

81. Mostofian F, Ruban C, Simunovic N, Bhandari M. Changing Physician Behavior: What Works? *Am J Manag Care*. 2015;21(1):75-84.

82. Kobewka DM, Ronksley PE, McKay JA, Forster AJ, Van Walraven C. Influence of educational, audit and feedback, system based, and incentive and penalty interventions to reduce laboratory test utilization: A systematic review. *Clin Chem Lab Med*. 2015;53(2):157-183. doi:10.1515/cclm-2014-0778.

83. Borgert M, Binnekade J, Paulus F, Goossens A, Vroom M, Dongelmans D. Timely individual audit and feedback significantly improves transfusion bundle compliance—a comparative study. *Int J Qual Heal Care*. 2016;28(5):601-607. doi:10.1093/intqhc/mzw071.

84. Marik PE, Corwin HL. Efficacy of red blood cell transfusion in the critically ill: a systematic review of the literature. *Crit Care Med*. 2008;36(9):2667-2674. doi:10.1097/CCM.0b013e3181844677.

85. Seitz KP, Sevransky JE, Martin GS, Roback JD, Murphy DJ. Evaluation of RBC Transfusion Practice in Adult ICUs and the Effect of Restrictive Transfusion Protocols on Routine Care [published online Sept 14 2016]. *Crit Care Med*. 2016. doi:10.1097/CCM.0000000000002077.

86. Critical Care Societies Collaborative - Critical Care. Five Things Physicians and Patients Should Question. Choosing Wisely. http://www.choosingwisely.org/societies/critical-care-societies-collaborative-critical-care/. Published 2014. Accessed April 23, 2017.

87. Canadian Critical Care Society, Canadian Association of Critical Care Nurses, Canadian Society of Respiratory Therapists. Five Things Clinicians and Patients Should Question. Choosing Wisely Canada. http://www.choosingwiselycanada.org/recommendations/critical-care/. Published 2017. Accessed April 23, 2017.

88. McEvoy MT, Shander A. Anemia, Bleeding, and Blood Transfusion in the Intensive Care Unit: Causes, Risks, Costs, and New Strategies. *Am J Crit Care*. 2013;22(6 Suppl):eS1-13. doi:10.4037/ajcc2013729.

89. Carson JL, Stanworth SJ, Roubinian N, et al. Transfusion Thresholds and Other Strategies for Guiding Allogeneic Red Blood Cell Transfusion. *Cochrane Database Syst Rev*. 2016;(10):CD002042. doi:10.1002/14651858.CD002042.pub4.

90. Delaney M, Wendel S, Bercovitz RS, et al. Transfusion reactions: prevention, diagnosis, and treatment. *Lancet*. 2016;388(10061):2825-2836. doi:10.1016/S0140-6736(15)01313-6.

91. Thakkar RN, Lee KHK, Ness PM, et al. Relative Impact of a Patient Blood Management Program on Utilization of All Three Major Blood Components. *Transfusion*. 2016;56(9):2212-2220. doi:10.1111/trf.13718.

92. Zhu C, Gao Y, Li Z, et al. A Systematic Review and Meta-Analysis of the Clinical Appropriateness of Blood Transfusion in China. *Med*. 2015;94(50):e2164. doi:10.1097/MD.0000000000002164.

93. Murphy MF, Goodnough LT. The scientific basis for patient blood management. *Transfus Clin Biol*. 2015;22(3):90-96. doi:10.1016/j.tracli.2015.04.001.

94. Sadeghi A, Belali S, Asgari AA, Morovat Z, Malekzadeh R, Emadi A. Inappropriate Packed RBC Transfusion in a Tertiary Care Center. *Arch Iran Med*. 2017;20(2):83-85. doi:0172002/AIM.004.

95. Spradbrow J, Cohen R, Lin Y, et al. Evaluating appropriate red blood cell transfusions: A quality audit at 10 Ontario hospitals to determine the optimal measure for assessing appropriateness. *Transfusion*. 2016;[Epub ahea. doi:10.1111/trf.13737.

96. Holland LL, Brooks JP. Toward Rational Fresh Frozen Plasma Transfusion: The Effect of Plasma Transfusion on Coagulation Test Results. *Am J Clin Pathol*. 2006;126(1):133-139. doi:10.1309/NQXHUG7HND78LFFK.

97. Tinmouth A, Thompson T, Arnold DM, et al. Utilization of frozen plasma in Ontario: A provincewide audit reveals a high rate of inappropriate transfusions. *Transfusion*. 2013;53(10):2222-2229. doi:10.1111/trf.12231.

98. Alport EC, Callum JL, Nahirniak S, Eurich B, Hume HA. Cryoprecipitate use in 25 Canadian hospitals: Commonly used outside of the published guidelines. *Transfusion*. 2008;48(10):2122-2127. doi:10.1111/j.1537-2995.2008.01826.x.

99. Etchells M, Spradbrow J, Cohen R, Lin Y, Armali C, Lieberman L. Audit of appropriate use of platelet transfusions: validation of adjudication criteria. *Vox Sang*. 2018;113(1):40-50. doi:10.1111/vox.12550.

100. Wilson K, MacDougall L, Fergusson D, Graham I, Tinmouth A, Hébert PC. The effectiveness of interventions to reduce physician's levels of inappropriate transfusion: what can be learned from a systematic review of the literature. *Transfusion*. 2002;42(9):1224-1229. doi:10.1046/j.1537-2995.2002.00185.x.

101. Tinmouth A, Macdougall L, Fergusson D, et al. Reducing the Amount of Blood Transfused: A Systematic Review of Behavioral Interventions to Change Physicians' Transfusion Practices. *Arch Intern Med*. 2005;165(8):845-852. doi:10.1001/archinte.165.8.845.

102. American College of Obstetricians and Gynecologists. Ten Things Physicians and Patients Should Question. Choosing Wisely. http://www.choosingwisely.org/societies/american-college-of-obstetricians-and-gynecologists/. Accessed April 23, 2017.

103. American Association of Blood Banks. Five Things Physicians and Patients Should Question. Choosing Wisely. http://www.choosingwisely.org/societies/american-association-of-blood-banks/. Published 2014. Accessed April 23, 2017.

104. American Society of Hematology. Ten Things Physicians and Patients Should Question. Choosing Wisely. http://www.choosingwisely.org/societies/american-society-of-hematology/. Accessed April 23, 2017.

105. American Society of Anesthesiologists. Five Things Physicians and Patients Should Question. Choosing Wisely. http://www.choosingwisely.org/societies/american-society-of-anesthesiologists/. Published 2013. Accessed April 23, 2017.

106. Society of Hospital Medicine- Adult Hospital Medicine. Five Things Physicians and Patients Should Question. Choosing Wisely. http://www.choosingwisely.org/societies/society-of-hospital-medicine-adult/. Published 2013. Accessed April 23, 2017.

107. The Canadian Hematology Society, Société Canadienne d'Hématologie. Five Things Physicians and Patients Should Question. Choosing Wisely Canada. http://www.choosingwiselycanada.org/recommendations/hematology/. Published 2014. Accessed April 23, 2017.

108. Canadian Society of Palliative Care Physicians. Five Things Physicians and Patients Should Question. Choosing Wisely Canada. http://www.choosingwiselycanada.org/recommendations/palliative-care/. Published 2014. Accessed April 23, 2017.

109. Canadian Society for Transfusion Medicine, Société Canadienne de Médecine Transfusionnelle. Ten Things Physicians and Patients Should Question. Choosing Wisely Canada. http://www.choosingwiselycanada.org/recommendations/transfusion-medicine/. Published 2017. Accessed October 30, 2017.

110. Canadian Society of Internal Medicine. Five Things Physicians and Patients Should Question.

Choosing Wisely Canada. http://www.choosingwiselycanada.org/recommendations/internal-medicine/. Published 2014. Accessed April 23, 2017.

111. Donchin Y, Seagull FJ. The hostile environment of the intensive care unit. *Curr Opin Crit Care*. 2002;8(4):316-320. doi:10.1097/01.CCX.0000019846.92818.8B.

112. Bjurling-Sjöberg P, Wadensten B, Pöder U, Jansson I, Nordgren L. Balancing intertwined responsibilities: A grounded theory study of teamwork in everyday intensive care unit practice. *J Interprof Care*. 2017;31(2):233-244. doi:10.1080/13561820.2016.1255184.

113. Scales DC, Sibbald WJ. Medical technology in the intensive care unit. *Curr Opin Crit Care*. 2004;10(4):238-245. /Users/EWN/Documents/Arkiv_artikler/5100_5199/5174_med_techn_ICU.pdf.

114. Boev C. The Relationship Between Nurses' Perception of Work Environment and Patient Satisfaction in Adult Critical Care. *J Nurs Scholarsh*. 2012;44(4):368-375. doi:10.1111/j.1547-5069.2012.01466.x.

115. Alameddine M, Dainty KN, Deber R, Sibbald WJ (Bill). The intensive care unit work environment: Current challenges and recommendations for the future. *J Crit Care*. 2009;24(2):243-248. doi:10.1016/j.jcrc.2008.03.038.

116. Adhikari N, Lapinsky SE. Medical informatics in the Intensive Care Unit: Overview of Technology Assessment. *J Crit Care*. 2003;18(1):41-47. doi:10.1053/jcrc.2003.YJCRC9.

117. Reader TW, Cuthbertson BH, Decruyenaere J. Burnout in the ICU: Potential consequences for staff and patient well-being. *Intensive Care Med*. 2008;34(1):4-6. doi:10.1007/s00134-007-0908-4.

118. Verdon M, Merlani P, Perneger T, Ricou B. Burnout in a surgical ICU team. *Intensive Care Med*. 2008;34(1):152-156. doi:10.1007/s00134-007-0907-5.

119. Krinsley JS. Test-Ordering Strategy in the Intensive Care Unit. *J Intensive Care Med*. 2003;18(6):330-339. doi:10.1177/0885066603256344.

120. Seguin P, Bleichner J, Grolier J, Guillou Y, Mallédant Y. Effects of price information on test ordering in an intensive care unit. *Intensive Care Med*. 2002;28(3):332-335. doi:10.1007/s00134-002-1213-x.

121. Kleinpell RM, Farmer JC, Pastores SM. Reducing Unnecessary Testing in the Intensive Care Unit by Choosing Wisely. *Acute Crit Care*. 2018;33(1):1-6. doi:10.4266/acc.2018.00052.

122. Ullman AJ, Keogh S, Coyer F, Long DA, New K, Rickard CM. "True Blood" The Critical Care Story: An audit of blood sampling practice across three adult, paediatric and neonatal intensive care settings. *Aust Crit Care*. 2016;29(2):90-95. doi:10.1016/j.aucc.2015.06.002.

123. Carson JL, Duff A, Poses RM, et al. Effect of Anaemia and Cardiovascular Disease on Surgical Mortality and Morbidity. *Lancet*. 1996;348(9034):1055-1060. doi:10.1016/S0140-6736(96)04330-9.

124. Barie PS. Phlebotomy in the Intensive Care Unit: Strategies for Blood Conservation. *Crit Care*. 2004;8 Suppl 2:S34-6. doi:10.1186/cc2454.

125. Cohen ML. Epidemiology of Drug Resistance: Implications for a Post-Antimicrobial Era. *Science (80- )*. 1992;257(5073):1050-1055. doi:10.1126/science.257.5073.1050.

126. Jensen PR, Markewitz BA. Safe Reduction of Blood Volume in the Blood Gas Laboratory. *Lab Med*. 2016;47(4):326-329. doi:10.1093/labmed/lmw039.

127. Mikhaeil M, Day AG, Ilan R. Non-essential blood tests in the intensive care unit: a prospective observational study. *Can J Anesth Can d'Anesthésie*. 2017;64(3):290-295. doi:10.1007/s12630-016-0793-9.

128. Eccles M, Grimshaw J, Walker A, Johnston M, Pitts N. Changing the Behavior of Healthcare Professionals: The Use of Theory in Promoting the Uptake of Research Findings. *J Clin Epidemiol*. 2005;58(2):107-112. doi:10.1016/j.jclinepi.2004.09.002.

129. Straus SE, Tetroe J, Graham ID. Defining knowledge translation. *Can Med Assoc J*. 2009;181(3-4):165-168. doi:10.1503/cmaj.081229.

130. Effective Practice and Organisation of Care (EPOC). EPOC Taxonomy; 2015. Available at: https://epoc.cochrane.org/epoc-taxonomy.

131. Ivers N, Jamtvedt G, Flottorp S, et al. Audit and Feedback : Effects on Professional Practice and Healthcare Outcomes ( Review ). *Cochrane Database Syst Rev*. 2012;(6):1-227. doi:10.1002/14651858.CD000259.pub3.

132. Bennett JW, Glasziou PP. Computerised reminders and feedback in medication management: A systematic review of randomised controlled trials. *Med J Aust*. 2003;178(5):217-222. http://www.embase.com/search/results?subaction=viewrecord&from=export&id=L36329215%255Cnhttp://elvis.ubvu.vu.nl:9003/vulink?sid=EMBASE&issn=0025729X&id=doi:&atitle=Computerised+reminders+and+feedback+in+medication+management%253A+A+systematic+review+of+random.

133. Halpape K, Sulz L, Schuster B, Taylor R. Audit and Feedback-Focused approach to Evidence-based Care in Treating patients with pneumonia in hospital (AFFECT Study). *Can J Hosp Pharm*. 2014;67(1):17-27.

134. Pimlott NJG, Hux JE, Wilson LM, Kahan M, Li C, Rosser WW. Educating Physicians to Reduce Benzodiazepine Use by Elderly Patients: A Randomized Controlled Trial. *CMAJ*. 2003;168(7):835-839.

135. Brehaut JC, Colquhoun HL, Eva KW, et al. Practice Feedback Interventions: 15 Suggestions for Optimizing Effectiveness. *Ann Intern Med*. 2016;164(6):435-441. doi:10.7326/M15-2248.

136. Davis D a, Mazmanian PE, Fordis M, Harrison R Van, Thorpe KE, Perrier L. Accuracy of Physician Self-assessment Compared With Observed Measures of Competence. *JAMA*. 2006;296(9):1094-1102. doi:10.1001/jama.296.9.1094.

137. Kluger AN, DeNisi A. The Effects of Feedback Interventions on Performance: A Historical Review, a Meta-Analysis, and a Preliminary Feedback Intervention Theory. *Psychol Bull*. 1996;119(2):254-284. doi:10.1037/0033-2909.119.2.254.

138. Hysong ST. Meta-Analysis: Audit and Feedback Features Impact Effectiveness on Care Quality. *Med Care*. 2009;47(3):356-363.

139. Pendry K. The use of big data in transfusion medicine. *Transfus Med*. 2015;25(3):129-137. doi:10.1111/tme.12223.

140. Krein SL, Damschroder LJ, Kowalski CP, Forman J, Hofer TP, Saint S. The influence of

organizational context on quality improvement and patient safety efforts in infection prevention: A multi-center qualitative study. *Soc Sci Med*. 2010;71(9):1692-1701. doi:10.1016/j.socscimed.2010.07.041.

141. Rycroft-Malone J, Harvey G, Seers K, Kitson A, McCormack B, Titchen A. An exploration of the factors that influence the implementation of evidence into practice. *J Clin Nurs*. 2004;13(8):913-924. doi:10.1111/j.1365-2702.2004.01007.x.

142. Kaplan HC, Provost LP, Froehle CM, Margolis PA. The model for understanding success in quality (MUSIQ): Building a theory of context in healthcare quality improvement. *BMJ Qual Saf*. 2012;21(1):13-20. doi:10.1136/bmjqs-2011-000010.

143. de Vos ML, van der Veer SN, Graafmans WC, et al. Implementing quality indicators in intensive care units: exploring barriers to and facilitators of behaviour change. *Implement Sci*. 2010;5:52. doi:10.1186/1748-5908-5-52.

144. Sinuff T, Muscedere J, Adhikari NKJ, et al. Knowledge Translation Interventions for Critically Ill Patients: A Systematic Review. *Crit Care Med*. 2013;41(11):2627-2640. doi:10.1097/CCM.0b013e3182982b03.

145. Sinuff T, Cook D, Giacomini M, Heyland D, Dodek P. Facilitating Clinician Adherence to Guidelines in the Intensive Care Unit: A Multicenter, Qualitative Study. *Crit Care Med*. 2007;35(9):2083-2089. doi:10.1097/01.ccm.0000281446.15342.74.

146. Lapinsky SE, Holt D, Hallett D, Abdolell M, Adhikari NKJ. Survey of information technology in Intensive Care Units in Ontario, Canada. *BMC Med Inform Decis Mak*. 2008;8:1-6. doi:10.1186/1472-6947-8-5.

147. Vincent J-L, Fink MP, Marini JJ, et al. Intensive Care and Emergency Medicine: Progress Over the Past 25 Years. *Chest*. 2006;129(4):1061-1067. doi:10.1378/chest.129.4.1061.

148. de Vos MLG, van der Veer SN, Graafmans WC, et al. Process evaluation of a tailored multifaceted feedback program to improve the quality of intensive care by using quality indicators. *BMJ Qual Saf*. 2013;22(3):233-241. doi:10.1136/bmjqs-2012-001375.

149. Van De Klundert N, Holman R, Dongelmans DA, De Keizer NF. Data Resource Profile: The Dutch National Intensive Care Evaluation (NICE) Registry of Admissions to Adult Intensive Care Units. *Int J Epidemiol*. 2015;44(6):1850-1850h. doi:10.1093/ije/dyv291.

150. Ivers NM, Grimshaw JM, Jamtvedt G, et al. Growing Literature, Stagnant Science? Systematic Review, Meta-Regression and Cumulative Analysis of Audit and Feedback Interventions in Health Care. *J Gen Intern Med*. 2014;29(11):1534-1541. doi:10.1007/s11606-014-2913-y.

151. Ivers NM, Grimshaw JM. Reducing research waste with implementation laboratories. *Lancet*. 2016;388(10044):547-548. doi:10.1016/S0140-6736(16)31256-9.

152. Colquhoun HL, Brehaut JC, Sales A, et al. A systematic review of the use of theory in randomized controlled trials of audit and feedback. *Implement Sci*. 2013;8:66. doi:10.1186/1748-5908-8-66.

153. Birken SA, Powell BJ, Shea CM, et al. Criteria for selecting implementation science theories and frameworks: results from an international survey. *Implement Sci*. 2017;12(1):124. doi:10.1186/s13012-017-0656-y.

154. Grol RP, Bosch MC, Hulscher ME, Eccles M, Wensing M. Planning and Studying Improvement

in Patient Care: The Use of Theoretical Perspectives. *Milbank Q*. 2007;85(1):93-138.

155. Carver CS, Scheier MF. Control Theory: A Useful Conceptual Framework for Personality-Social, Clinical, and Health Psychology. *Psychol Bull*. 1982;92(1):111-135. doi:10.1037/0033-2909.92.1.111.

156. Bandura A. Social Cognitive Theory of Self-Regulation. *Organ Behav Hum Decis Process*. 1991;50(2):248-287. doi:10.1016/0749-5978(91)90022-L.

157. Locke E a, Latham GP. Building a Practically Useful Theory of Goal Setting and Task Motivation: A 35-Year Odyssey. *Am Psychol*. 2002;57(9):705-717. doi:10.1037/0003-066X.57.9.705.

158. Locke EA. Goal theory vs. control theory: Contrasting approaches to understanding work motivation. *Motiv Emot*. 1991;15(1):9-28. doi:10.1007/BF00991473.

159. Kiefe CI, Allison JJ, Williams OD, Person SD, Weaver MT, Weissman NW. Improving Quality Improvement Using Achievable Benchmarks For Physician Feedback: A Randomized Controlled Trial. *JAMA*. 2001;285(22):2871-2879. doi:10.1001/jama.285.22.2871.

160. Colquhoun H, Michie S, Sales A, et al. Reporting and design elements of audit and feedback interventions: A secondary review [published online Jan 25 2016]. *BMJ Qual Saf*. 2016. doi:10.1136/bmjqs-2015-005004.

161. Merkeley HL, Hemmett J, Cessford TA, et al. Multipronged strategy to reduce routine-priority blood testing in intensive care unit patients. *J Crit Care*. 2016;31(1):212-216. doi:10.1016/j.jcrc.2015.09.013.

162. Kotecha N, Shapiro JM, Cardasis J, Narayanswami G. Reducing Unnecessary Laboratory Testing in the Medical ICU. *Am J Med*. 2017;130(6):648-651. doi:10.1016/j.amjmed.2017.02.014.

163. Ottawa Health Science Network Research Ethics Board (OHSN-REB). Is your project Research or Quality Improvement ? Guideline & Checklist. http://www.ohri.ca/ohsn-reb/forms/Checklist Research or QI - revised November 2 2016.pdf. Published 2016. Accessed February 2, 2019.

164. Kanter M, Courneya PT. Perspective on Publishing Quality Improvement Efforts. *Perm Journal/Perm J Perm J*. 2017;2121(3):17-140. doi:10.7812/TPP/17-140.

165. Hempel S, Rubenstein L V., Shanman RM, et al. Identifying quality improvement intervention publications - A comparison of electronic search strategies. *Implement Sci*. 2011;6(85). doi:10.1186/1748-5908-6-85.

166. Lippi G, Bovo C, Ciaccio M. Inappropriateness in laboratory medicine: an elephant in the room? *Ann Transl Med*. 2017;5(4):82. doi:10.21037/atm.2017.02.04.

167. Naugler CT, Guo M. Mean Abnormal Result Rate: Proof of Concept of a New Metric for Benchmarking Selectivity in Laboratory Test Ordering. *Am J Clin Pathol*. 2016;145(4):568-573. doi:10.1093/AJCP/AQW041.

168. Islam R, Tinmouth AT, Francis JJ, et al. A cross-country comparison of intensive care physicians' beliefs about their transfusion behaviour: A qualitative study using the theoretical domains framework. *Implement Sci*. 2012;7:93. doi:10.1186/1748-5908-7-93.

169. Gude WT, Roos-Blom M-J, van der Veer SN, et al. Health professionals' perceptions about their clinical performance and the influence of audit and feedback on their intentions to improve

practice: a theory-based study in Dutch intensive care units. *Implement Sci*. 2018;13(1):33. doi:10.1186/s13012-018-0727-8.

170. Craig P, Dieppe P, Macintyre S, Mitchie S, Nazareth I, Petticrew M. Developing and evaluating complex interventions: The new Medical Research Council guidance. *BMJ*. 2008;337(a1655). doi:10.1136/bmj.a1655.

171. Grol R, Grimshaw J. From best evidence to best practice: Effective implementation of change in patients' care. *Lancet*. 2003;362(9391):1225-1230. doi:10.1016/S0140-6736(03)14546-1.

172. Atkins L, Francis J, Islam R, et al. A guide to using the Theoretical Domains Framework of behaviour change to investigate implementation problems. *Implement Sci*. 2017;12(1):77. doi:10.1186/s13012-017-0605-9.

173. Kluger AN, DeNisi A. Feedback Interventions: Toward the Understanding of a Double-Edged Sword. *Curr Dir Psychol Sci*. 1998;7(3):67-72. doi:10.1111/1467-8721.ep10772989.

174. Metcalfe J, Kornell N, Son LK. A cognitive-science based programme to enhance study efficacy in a high and low risk setting. *Eur J Cogn Psychol*. 2007;19(4-5):743-768. doi:10.1080/09541440701326063.

175. Yeh DD, Naraghi L, Larentzakis A, et al. Peer-to-peer physician feedback improves adherence to blood transfusion guidelines in the surgical intensive care unit. *J Trauma Acute Care Surg*. 2015;79(1):65-70. doi:10.1097/TA.0000000000000683.

176. Beaty CA, Haggerty KA, Moser MG, et al. Disclosure of Physician-Specific Behavior Improves Blood Utilization Protocol Adherence in Cardiac Surgery. *Ann Thorac Surg*. 2013;96(6):2168-2174. doi:10.1016/j.athoracsur.2013.06.080.

177. Sinuff T, Muscedere J, Rozmovits L, Dale CM, Scales DC. A Qualitative Study of the Variable Effects of Audit and Feedback in the ICU. *BMJ Qual Saf*. 2015;24(6):393-399. doi:10.1136/bmjqs-2015-003978.

178. Sheeran P, Milne S, Webb TL, Gollwitzer PM. Implementation Intentions and Health Behaviour. In: Conner M, Norman P, eds. *Predicting Health Behaviour: Research and Practice with Social Cognition Models*. Second. Maidenhead, UK: Open University Press; 2005:276-323.

179. Sniehotta FF. Towards a theory of intentional behaviour change: Plans, planning, and self-regulation. *Br J Health Psychol*. 2009;14(2):261-273. doi:10.1348/135910708X389042.

180. Kwasnicka D, Presseau J, White M, Sniehotta FF. Does planning how to cope with anticipated barriers facilitate health-related behaviour change ? A systematic review. *Health Psychol Rev*. 2013;7(2):129-145.

181. Potthoff S, Presseau J, Sniehotta FF, Johnston M, Elovainio M, Avery L. Planning to be routine : habit as a mediator of the planning-behaviour relationship in healthcare professionals. *Implement Sci*. 2017;12(1):24. doi:10.1186/s13012-017-0551-6.

182. French S, Green S, O'Connor D, et al. Developing theory-informed behaviour change interventions to implement evidence into practice: a systematic approach using the Theoretical Domains Framework. *Implement Sci*. 2012;7:38.

183. Scholte M, Neeleman-van der Steen CWM, van der Wees PJ, Nijhuis-van der Sanden MWG, Joze B. The Reasons behind the (Non) Use of Feedback Reports for Quality Improvement in Physical

Therapy: A Mixed-Method Study. *PLoS One*. 2016;11(8):e0161056. doi:10.1371/journal.pone.0161056.

184. Ivers NM, Tu K, Young J, et al. Feedback GAP: pragmatic, cluster- randomized trial of goal setting and action plans to increase the effectiveness of audit and feedback interventions in primary care. *Implement Sci*. 2013;8(142). doi:10.1186/1748-5908-8-142.

185. Boud D, Molloy E. Rethinking models of feedback for learning: the challenge of design. *Assess Eval High Educ*. 2013;38(6):698-712.

186. Hysong SJ, Best RG, Pugh JA. Audit and feedback and clinical practice guideline adherence: Making feedback actionable. *Implement Sci*. 2006;1:9. doi:10.1186/1748-5908-1-9.

187. Ivers NM, Desveaux L, Presseau J, et al. Testing feedback message framing and comparators to address prescribing of high-risk medications in nursing homes: protocol for a pragmatic, factorial, cluster-randomized trial. *Implement Sci*. 2017;12(1):86. doi:10.1186/s13012-017-0615-7.

188. Tracy JP, Albers MJ. Measuring Cognitive Load to Test the Usability of Web Sites. *Usability Inf Des*. 2006:256-260.

189. van der Weegen S, Verwey R, Tange HJ, Spreeuwenberg MD, de Witte LP. Usability testing of a monitoring and feedback tool to stimulate physical activity. *Patient Prefer Adherence*. 2014;8:311-322. doi:10.2147/PPA.S57961.

190. Brown B, Balatsoukas P, Williams R, Sperrin M, Buchan I. Interface design recommendations for computerised clinical audit and feedback: Hybrid usability evidence from a research-led system. *Int J Med Inform*. 2016;94:191-206. doi:10.1016/j.ijmedinf.2016.07.010.

191. Eccles M, Grimshaw J, Campbell M, Ramsay C. Research designs for studies evaluating the effectiveness of change and improvement strategies. *Qual Saf Heal Care*. 2003;12(1):47-52. http://qualitysafety.bmj.com/content/qhc/12/1/47.full.pdf%0Ahttp://www.ncbi.nlm.nih.gov/pubmed/12571345.

192. Tate DF, Lytle LA, Sherwood NE, et al. Deconstructing interventions: approaches to studying behavior change techniques across obesity interventions. *Transl Behav Med*. 2016;6(2):236-243. doi:10.1007/s13142-015-0369-1.

193. Paes BA, Modi A, Dunmore R. Changing physicians' behavior using combined strategies and an evidence-based protocol. *Arch Pediatr Adolesc Med*. 1994;148(12):1277-1280. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med3&NEWS=N&AN=7951806.

194. DeVon HA, Block ME, Moyle-Wright P, et al. A Psychometric Toolbox for Testing Validity and Reliability. *J Nurs Scholarsh*. 2007;39(2):155-164.

195. Eastern Ontario Regional Laboratory Association. EORLA- About Us. EORLA Website. http://www.eorla.ca/about-us/. Accessed May 3, 2016.

196. Goodacre S. Uncontrolled before-after studies: Discouraged by Cochrane and the EMJ. *Emerg Med J*. 2015;32(7):507-508. https://doi.org/10.1136/emermed-2015-204761.

197. Verstappen WHJM, Van Der Weijden T, Ter Riet G, Grimshaw J, Winkens R, Grol RPTM. Block design allowed for control of the Hawthorne effect in a randomized controlled trial of test ordering. *J Clin Epidemiol*. 2004;57(11):1119-1123. doi:10.1016/j.jclinepi.2004.03.009.

198. Sterne J, Egger M, Moher D (editors). Chapter 10: Addressing reporting biases. In: Higgins J, Green S, eds. *Cochrane Handbook for Systematic Reviews of Intervention. Version 5.1.0 (Updated March 2011)*. The Cochrane Collaboration; 2011. Available from www.handbook.cochrane.org.

199. Hoffmann TC, Glasziou PP, Boutron I, et al. Better reporting of interventions: Template for intervention description and replication (TIDieR) checklist and guide. *BMJ*. 2014;348:g1687. doi:10.1136/bmj.g1687.

200. Powell BJ, Waltz TJ, Chinman MJ, et al. A refined compilation of implementation strategies: Results from the Expert Recommendations for Implementing Change (ERIC) project. *Implement Sci*. 2015;10(21). doi:10.1186/s13012-015-0209-1.

201. Presseau J, Ivers NM, Newham JJ, Knittle K, Danko KJ, Grimshaw JM. Using a behaviour change techniques taxonomy to identify active ingredients within trials of implementation interventions for diabetes care. 2015:1-10. doi:10.1186/s13012-015-0248-7.

202. Michie S, Wood CE, Johnston M, Abraham C, Francis JJ, Hardeman W. Behaviour change techniques: The development and evaluation of a taxonomic method for reporting and describing behaviour change interventions (a suite of five studies involving consensus methods, randomised controlled trials and analysis of qualitative da. *Health Technol Assess (Rockv)*. 2015;19(99):1-187. doi:10.3310/hta19990.