Inspired by research. Inspiré par la recherche. Driven by compassion. Guidé par la compassion.

STUDY DESIGNS FOR EVALUATING EFFECTIVENESS OF AUDIT AND FEEDBACK

International Audit and Feedback Summit Amsterdam May 23, 2019

BEATRIZ GOULAO JEREMY GRIMSHAW CRAIG RAMSAY

Health Services Research Unit, University of Aberdeen Ottawa Hospital Research Institute



The Ottawa | L'Hôpital Hospital d'Ottawa



www.ottawahospital.on.ca | Affiliated with · Affilié à



OUTLINE

- 1. Introduction -5 min
- 2. Case study: RAPiD 5 min
- Considerations in choosing a study design 10 min

Audience participation 10 min

4. Non-randomized designs – 10 min

Audience participation 10 min

5. Randomized designs – 20 min

Audience participation 10 min

- 6. Case study: RAPiD 5 min
- 7. Future directions and wrap up 5 min

OVERVIEW OF THE WORKSHOP

- ► This workshop is intended to be interactive
- ▶ We will introduce the RAPiD study at the beginning of the workshop
- ► As we progress through the workshop, we will pause several times to allow you to discuss the material, in particular, to discuss how to design an evaluation of the RAPiD intervention
- ▶ We will ask 1-2 tables to report back on their discussions
- At the end of the workshop, we will reveal the actual study design that was used

1. INTRODUCTION

CONTEXT

► Setting:

• A&F being provided "in the real world"

▶ Interventions:

- Embedded into existing QI programmes
- Complex (multiple interacting components)
- Delivered at the level of the provider or site ("cluster")

Outcomes:

- Observed on multiple individuals (patients) per cluster
- Usually obtained from routinely collected sources

PURPOSE OF EVALUATION

Program evaluation

- Addressing local question, did our program appear to achieve our aims
- Research evaluation
 - Addressing generalizable question, does audit and feedback work (it does, stop asking this question), how, when and why does audit and feedback work, how can we optimize audit and feedback within specific settings.
 - Research evaluation will (almost always) also address the local question
- ▶ Implications for design choices
 - May need less confidence about causality when undertaking program evaluation

DIFFERENT TYPES OF QUESTIONS

0. Is there an association between providing feedback and prescription rates? 1. Is there a causal relationship between providing feedback and prescription rates? 2. Can we refine the type of feedback that produces the largest effect in prescription rates?

3. Can we generalise the results over varied health care professionals and settings?

2. CASE STUDY: THE RAPID STUDY



OPEN ACCESS

Citation: Elouafkaoui P, Young L, Newlands R, Duncon EM, Eldors A, Clorkoon JE, et al. (2016) An

PLOS MEDICINE

RESEARCH ARTICLE

An Audit and Feedback Intervention for Reducing Antibiotic Prescribing in General Dental Practice: The RAPiD

Paula Elouafkaoui^{1,2}, Linda Young¹*, Rumana Newlands³, Eilidh M. Duncan³, Andrew Elders⁴, Jan E. Clarkson^{1,2}, Craig R. Ramsay³, Translation Research in a Dental Setting (TRiaDS) Research Methodology Group¹

1 NHS Education for Scotland (NES), Dundee Dental Education Centre, Frankland Building, Dundee, United Kingdom, 2 Dental Health Services Research Unit (DHSRU), University of Dundee, Park Place, Dundee, United Kingdom, 3 Health Services Research Unit (HSRU), University of Aberdeen, Health Sciences Building, Foresterhill, Aberdeen, United Kingdom, 4 NMAHP Research Unit, Glasgow Caledonian University, Cowcaddens Road, Glasgow, United Kingdom

¶ Membership of the Translation Research in a Dental Setting (TRiaDS) Research Methodology Group is provided in the Acknowledgments. * linda.young@nes.scot.nhs.uk

SETTING



- Dental healthcare mostly provided via public insurance (National Health Service, NHS) –€750 million/year
- 1000 NHS primary care dental practices
- ► 3,200 dentists
- The RAPiD study aimed to assess the impact of individualised audit and feedback (A&F) interventions on dentists' antibiotic prescribing rates
- ▶ May 2013: Launched an A&F intervention
- Routine prescribing data are updated on a monthly basis

RAPID A&F INTERVENTION



Your prescribing rate is your monthly number of antibiotic items dispensed multiplied by 100 and divided by the average monthly number of claims made your monthly number of antibiotic items dispensed multiplied by 100 and divided by the average monthly number of claims made your monthly number of antibiotic items dispensed multiplied by 100 and divided by the average monthly number of claims made your monthly number of antibiotic items dispensed multiplied by 100 and divided by the average monthly number of claims made your monthly number of antibiotic items dispensed multiplied by 100 and divided by the average monthly number of claims made on your ordinary lists at this practice between April 2016 and March 2017 between April 2016 and March 2017 between April 2016 and March 2017 between April 2016 and March 2017. The health board rate is the overall ordinary list prescribing rate for current dentists in non-salaried practices in NHS Lothian. non-salaried practices in NHS Lothian. On-salaried practices in NHS Lothian. In Source: ISD Scotland. Data as at March 2017)

Prescribing courses of antibiotic treatment can encourage the development of antimicrobial resistance and therefore must be kept to a minimum. Graphical display of current prescribing practice, regional health board prescribing data and a written behavior change message

AUDIENCE PARTICIPATION – 10 MIN

- ► You are going to plan an evaluation of the effectiveness of the RAPiD intervention.
 - Discuss what you see as the key considerations in choosing a study design with respect to the RAPiD evaluation

KEY CONSIDERATIONS IN CHOOSING A STUDY DESIGN

- ► Can the delivery of the intervention be manipulated (i.e., can we use randomization)?
- ► How many independent providers/sites are available?
- Is there a requirement that the intervention be introduced at all sites (or can it be withheld from some sites)?
- Is it logistically feasible to introduce the intervention simultaneously across all sites?
- ► Are pre-intervention outcome data available to use in the evaluation?

THE COMPARATOR

▶ To evaluate effectiveness of an intervention, we need a comparator

► Three possible choices:

- A&F versus no A&F (not ideal)
- Usual A&F versus new A&F
- A&F + something else versus A&F alone

TWO MAIN TYPES OF STUDY DESIGNS



Minimizing bias (internal validity)

- Is the observed improvement actually caused by the A&F?
- Maximizing generalizability (external validity)
 - Will the A&F also work in other sites/providers and other patients?

5. NON-RANDOMIZED DESIGNS

► Major study designs:

- 1. Uncontrolled before and after
- 2. Controlled before and after
- 3. Interrupted time series (ITS)
- 4. Controlled interrupted time series
- 5. Multiple baseline interrupted time series

NON-RANDOMIZED DESIGNS

1. Uncontrolled before and after study

Months	12	24
Site		

2. Controlled before and after study

Months	12	24
Site 1		
Site 2		

3. Interrupted Time Series

Months	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
Site																								

4. Controlled Interrupted Time Series

Months	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
Site 1																								
Site 2																								



1. UNCONTROLLED BEFORE AND AFTER

Major threat to validity



1. UNCONTROLLED BEFORE AND AFTER

Major threat to validity



Apparent effect completely confounded with the secular trend

3. INTERRUPTED TIME SERIES



- Called "interrupted" time series because we look for an "interruption" in the line at the time of the intervention
- Look for either an immediate change or gradual change
- Can project what outcomes would have been had intervention not been introduced





trend, interrupted time-series analysis. Lancet, 385 (2015), pp. 1219–1228













3. INTERRUPTED TIME SERIES

- Sample size requirements:
 - Single site or multiple sites
 - Need relatively large numbers of observations per measurement (at least 50)
 - Need at least 8-12 measurement intervals pre and post
- ► Generally more difficult to conduct power calculations

3. INTERRUPTED TIME SERIES

Advantages:

- Can be used to evaluate intervention introduced at a single site or at the same time across the population
- Easy to use with routinely collected data over many time periods
- Can rule out pre-existing (secular) trends as an alternative explanation
- Clear graphical presentation of results, easy to explain
- Only need aggregate data

► Disadvantages:

- Cannot rule out possibility that another change occurred at the same time as the intervention
- Long study duration
- Difficult to interpret when there are few events per time period
- Difficult to interpret when data collection methods change over time
- Difficult to separate independent effects of different components of an intervention implemented close together in time

4. CONTROLLED INTERRUPTED TIME SERIES

Two major threats to validity of interrupted time series:

- Possibility that another change, occurring at the same time, is an alternative explanation for the observed changes
- Major shift in the characteristics of the population which coincided with the intervention
- ► Can be strengthened by adding one or more controls
 - External control: adding an interrupted time-series analysis for a comparison site which did not implement the intervention
 - Internal control: adding an interrupted time series analysis for an outcome not targeted by the intervention

Compare changes in the control with changes in the intervention series

COMMON METHODOLOGICAL PROBLEM: BASELINE IMBALANCE



5. MULTIPLE BASELINE INTERRUPTED TIME SERIES



► Multiple intervention sites with staggered implementation of intervention

- ► Look for an interruption at a particular time where intervention was introduced, accompanied by absence of an interruption at other sites
- ► Conduct an ITS analysis in each and pool the results (where possible)
- ▶ Looks like a stepped wedge design (but too few sites for stepped wedge)

MULTIPLE BASELINE INTERRUPTED TIME SERIES





5. MULTIPLE BASELINE INTERRUPTED TIME SERIES

Advantages:

- Can be used to evaluate intervention introduced at a small number of sites (too few for a randomized design)
- The greater the number of sites showing a change corresponding to the time at which the intervention was introduced, the more confident one can be that the intervention produced the observed changes (as opposed to some other influences)

Disadvantages:

- Can increase the overall study duration
- Can be difficult to interpret when sites are heterogeneous
- Works best when different sites operate independently of each other (no contamination)
- Can be difficult to interpret when interventions are implemented close together in time
- More difficult to produce a single estimate of intervention effect

AUDIENCE PARTICIPATION – 10 MIN

► How should we evaluate the effectiveness of the RAPiD implementation?

- Consider the 5 different non-randomized study designs with respect to the RAPiD evaluation
- Discuss possible designs to evaluate the RAPiD intervention

RANDOMISED CONTROLLED TRIALS



UNIT OF RANDOMIZATION

Two types of randomized controlled trials:

- Individual randomized trial
- Cluster randomized trial (CRT)

Individual randomization generally preferable (but not possible for site- or provider-level interventions such as A&F)









Practice 1 A+F Practice 2 A+F with message Practice 3 A+F

Practice 4 A+F with message



Randomisation at the practice level



Analysis at the health professional level



Practice 1

Practice 2

Practice 1



This Photo by Unknown Author is licensed under CC BY-ND



INTRACLUSTER CORRELATION

Independent members

Total dependence



Downsides?

We usually need to recruit more participants in a cluster randomised trial

▶ So.... Why are we using it?





RANDOMIZATION UNIT TRADE-OFFS



Decreasing intracluster correlation

4. RANDOMIZED DESIGNS

► Main cluster randomized trial (CRT) designs:

- 1. Two arm parallel design
- 2. Multi-arm parallel design
- 3. Parallel arm before and after design
- 4. Repeated measures parallel arm design
- 5. Stepped wedge design
- 6. Factorial trial design

PARALLEL DESIGNS

▶ Two arms

	Time	
Sites	1	
1		
		A&F intervent
		No interventio
Ν		

Does it work?

Multiple arms



Which version works the best?

PARALLEL MULTI-ARM DESIGNS

Advantages

 Allows comparison of multiple interventions or levels of intervention under similar circumstances

Disadvantages

- Need more sites to achieve the same power (due to use of multiple arms)
- Small differences between arms implies larger sample sizes required
- Analysis more complicated (need to account for multiple comparisons)

BEFORE AND AFTER PARALLEL ARM



 Add a pre-intervention measurement in both arms

BEFORE AND AFTER PARALLEL ARM

Advantages

- Can assess whether sites in different arms are comparable before intervention
- Utilizing the pre-intervention data in analysis can increase power
- Can assess whether sites who are dropped from the analysis (e.g., due to closures, mergers, attrition) are similar to those who remain

Disadvantages:

- More complex analysis
- Different methods of analysis are possible which may give different answers
- May extend the total study duration if no routine data available

STEPPED WEDGE

	Time											
Groups	1	2	3	4	5	6						
1												
2												
3												
4												
5												



- ► All sites start in control and end in intervention condition
- Sites cross to intervention sequentially and in random order
- ▶ Outcomes are assessed repeatedly in each site over time

STEPPED WEDGE: ADVANTAGES

- Uses randomization better than implementing the intervention at all sites without any randomization
- ▶ May increase power over parallel arm designs
- ▶ Delivery of the intervention can be spread out over time (e.g., by having only one site or a small number of sites cross over each time)

STEPPED WEDGE: DISADVANTAGES

- ► All sites must be ready to implement intervention at any time
- Can increase the total duration of the study (increase risk that external events may influence outcomes)
- ▶ Some sites have to wait a long time before receiving intervention
- ► Heavy data collection burden (unless using routinely collected data)
- ► More complex to analyze and interpret results (can be difficult to separate the effect of the intervention from the effect of secular trends)

FACTORIAL DESIGN

▶ 2x2 factorial design



INTERACTION











INTERACTION









÷

This Photo by Unknown Author is licensed under <u>CC BY-SA</u>

FACTORIAL DESIGN: NEXUS TRIAL

► 2x2 factorial design Randomize

Randomize	A&F	Not A&F				
Educational message	Message + A&F	Message only	Outcome : Number of radiograph			
No educational message	A&F only	Neither	patients per year			

INTERACTION



11%

S





This Photo by Unknown Author is licensed under CC BY-SA-NC

NO INTERACTION



This Photo by Unknown Author is licensed under CC BY-SA-NC

6. FACTORIAL DESIGNS

Advantages

- Multiple interventions tested in one trial (smaller sample size than if two separate trials)
- Allows examining possibility of interaction effects
- More participants exposed to potentially beneficial intervention

Disadvantages

- More complicated to analyze (must pre-specify whether pooled or fourarm comparison)
- Very difficult to guarantee no interaction took place (because usually there is insufficient power), so results can be difficult to interpret

AUDIENCE PARTICIPATION – 10 MIN

► How should we evaluate the effectiveness of the RAPiD implementation?

- Consider the different randomized designs with respect to the RAPiD evaluation
- Discuss possible randomized designs to evaluate the RAPiD intervention

RAPID TRIAL: RANDOMISED GROUP OPTIONS



INTERVENTION



The Ottawa | L'Hôpital Hospital d'Ottawa

wa

RESULTS

- ► At follow-up, the antibiotic prescribing rate of dentists who received individualised feedback was 5.7% lower than the antibiotic prescribing rate of dentists who did not receive individualised feedback.
- ► Feedback that included a written message synthesising and reiterating national guidance recommendations had the greatest effect

7. FUTURE DIRECTIONS

lvers et al. Implementation Science 2014, 9:14 http://www.implementationscience.com/content/9/1/14



DEBATE

Open Access

No more 'business as usual' with audit and feedback interventions: towards an agenda for a reinvigorated intervention

Noah M $\rm Ivers^{1*},$ Anne Sales², Heather Colquhoun³, Susan Michie⁴, Robbie Foy⁵, Jill J Francis⁶ and Jeremy M Grimshaw²

Abstract

Background: Audit and feedback interventions in healthcare have been found to be effective, but there has been little progress with respect to understanding their mechanisms of action or identifying their key 'active ingredients.' Discussion: Given the increasing use of audit and feedback to improve quality of care, it is imperative to focus further research on understanding how and when it works best. In this paper, we argue that continuing the 'business as usual' approach to evaluating two-arm trials of audit and feedback interventions against usual care for common problems and settings is unlikely to contribute new generalizable findings. Future audit and feedback trials should incorporate evidence- and theory-based best practices, and address known gaps in the literature.

Summary: We offer an agenda for high-priority research topics for implementation researchers that focuses on reviewing best practices for designing audit and feedback interventions to optimize effectiveness.

Keywords: Audit and feedback, Synthesis, Best practice, Implementation, Optimization

Background

Audit and feedback (A&F) involves providing a recipient with a summary of their performance over a specified period of time and is a common strategy to promote the implementation of evidence-based practices. A&F is used widely in healthcare by a range of stakeholders, including research funders and health system payers, delivery organizations, professional groups and researchers, to monitor and change health professionals' behaviour. both to increase accountability and to improve quality of care. A&F is an improvement over self-assessment [1] or self-monitoring [2] as it can provide objective data regarding discrepancies between current practice and target performance, as well as comparisons of performance to other health professionals. The recognition of suboptimal performance can act as a cue for action, encouraging those who are both motivated and capable to take action to reduce the discrepancy.

The effectiveness of A&F has been evaluated in the third update of a Cochrane review, which included 140 randomized trials of A&F conducted across many clinical conditions and settings around the world. The review found that A&F leads to a median 4.3% absolute improvement (interquartile range 0.5% to 16%) in provider compliance with desired practice [3]. One-quarter of A&F interventions had a relatively large, positive effect on quality of care, while another quarter had a negative or null effect. The challenge of identifying factors that differentiate more and less successful A&F interventions is exacerbated by poor reporting of both intervention components and contextual factors in the literature [4]. Furthermore, most A&F interventions tested in RCTs are designed without explicitly building on previous research or extant theory [5,6]. As a result, there has been little progress with respect to identifying the key ingredients for a successful A&F intervention or understanding the mechanisms of action of effective A&F interventions

IMPLEMENTATION LABORATORIES TO OPTIMISE AUDIT AND FEEDBACK

Reducing research waste with implementation laboratories

The Lancet REWARD (REduce research Waste And Reward and feedback intervention; the same is true for many examine how they work and make efforts to reduce waste substantial waste of scarce implementation research ensure and maximise efficiency. Research waste is undermining resources and promulgate evidence-practice gaps that efforts to improve the effectiveness of health systems, incurindividual and societal harms. A consistent finding in health services research is Health systems have a need for generalisable evidence inappropriate variations in care and evidence-practice about how to achieve the greatest possible impact with gaps. Implementation science-the study of methods their guality improvement initiatives.7 Implementation to promote the systematic uptake of clinical research intervention developers must make many decisions about findings and other evidence-based practices into routine content, format, and delivery of their intervention; even practice'-can inform health systems on how to reliably small modifications in these areas could influence the improve care and outcomes. However, the potential for effectiveness of the intervention.* Since the question of implementation science to improve the effectiveness of whether many common implementation strategies can health systems will not be realised until research waste in work has been answered, the time has come for a shift to the field is systematically addressed.

have known that audit and feedback is an effective way embedded within existing, large-scale initiatives. to improve care,² but researchers continue to undertake A promising solution is to develop implementation whether a particular version of audit and feedback can health systems delivering implementation strategies at work in a particular setting and for a particular purpose. scale and research teams. Implementation laboratories insufficient evidence on how best to design a new audit produce generalisable knowledge about factors-context,

For more on the Lance, REWARD Diligence) campaign has encouraged researchers to other implementation strategies.¹⁴ Such failures represent the lancet comi campaione

a comparative-effectiveness model for implementation A solid evidence base shows the effectiveness of research.⁹ Head-to-head trials that test different ways of common implementation strategies-eq, audit and designing and delivering implementation strategies are feedback," point of care reminders," educational meetings,4 needed to provide the evidence base for health system and educational outreach-but with substantial decision makers. Direct comparisons of implementation unexplained heterogeneity. Yet many current studies interventions will more efficiently move the field forward that evaluate implementation strategies against control than the current approach involving cumulating evidence create research waste because they do not build upon from fairly small trials for indirect analyses in systematic the current evidence base or address the key questions to reviews. However, the required sample sizes for such advance the field. For example, for more than a decade we research are difficult to achieve unless the research is

trials of audit and feedback versus usual care, testing laboratories that involve close collaboration between Such evaluations rarely incorporate relevant theory or best provide an opportunity to kick-start the field by ensuring practices⁶ in the design and delivery of the intervention that scholars meet both applied and scientific goals and do not address the question of how to optimise the of understanding what works better and why. Such effectiveness of audit and feedback. As a result, there is research can address health systems' priorities and

www.thelancet.com Vol 388 August 6, 2016



The Ottawa | L'Hôpitai d'Ottawa

62

IMPLEMENTATION LABORATORIES TO OPTIMISE AUDIT AND FEEDBACK





Affiliated with • Affilié à 🛍 uOttawa

IMPLEMENTATION LABORATORIES TO OPTIMISE AUDIT AND FEEDBACK

- Benefits for health system learning organisation; demonstrable improvements in its quality improvement activities; linkages to academic experts
- Benefits for implementation science ability to test important (but potentially subtle) variations in audit and feedback that may be important effect modifiers





CONCLUSIONS

- Many possible study designs that have strengths and weaknesses
- Choice of a particular design depends on research question and logistical considerations
- ▶ Generally, prefer a cluster randomized design
- Need special expertise to design and analyse appropriately

Inspired by research. Driven by compassion. Inspiré par la recherche. Guidé par la compassion.

THANK YOU FOR LISTENING!

BEATRIZ GOULAO JEREMY GRIMSHAW CRAIG RAMSAY

Health Services Research Unit, University of Aberdeen Ottawa Hospital Research Institute



The Ottawa | L'Hôpital Hospital d'Ottawa



www.ottawahospital.on.ca | Affiliated with · Affilié à

