Interpreting evidence and informing decisions

Learning Event #9 - March 2021

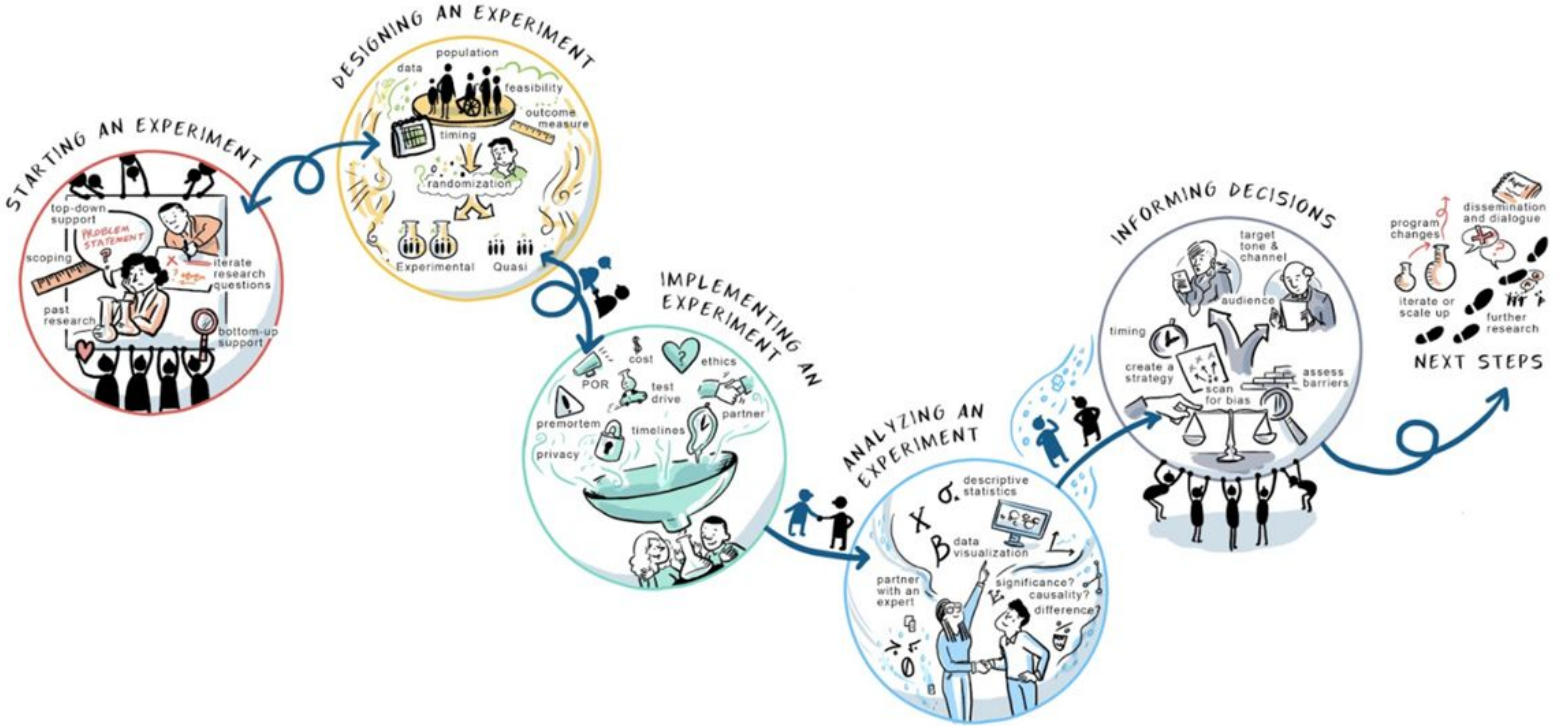
Pierre-Olivier Bédard, TBS

GC EXPÉRIMENTATION
WORKS | À L'OEUVRE

Outline

1. **Evidence Interpretation**
   Critical appraisal, risk of bias and error
2. **Evidence Dissemination**
   What should be done at this stage?
3. **Interactive Session & Discussion**
   Jamboard

The experimentation process



Evidence-informed decision-making

- We don’t necessarily know what works
- Even when we think we do, we don’t always have reliable evidence to support our claims
- Conducting an experiment in theory minimizes the risk of bias and risk of error

How reliable is your evidence?

Scared Straight programs became popular in the U.S. in the 1970s. They are based on the intuitive idea that perhaps people commit crimes because they are not aware of the consequences.

Teenagers visited inmates and observed prison life, with the goal of discouraging criminal behaviour later in life (‘scared straight’).

Significant resources were invested into the program, which was expanded in the 1980s and 1990s, continuing to receive positive coverage.

What does the evidence say?

"Several early studies, which looked at the criminal behaviours of participants before and after the programme, seemed to support these assumptions. Success rates were reported as being as high as 94%, and the programme was adopted in several countries, including the UK. None of these evaluations had a control group showing what would have happened to these participants if they had not participated in the programme."

"A meta-analysis of 7 US trials which randomly assigned half of the sample of at-risk children to the programme and found that "Scared Straight" in fact led to higher rates of offending behaviour: "doing nothing would have been better than exposing juveniles to the program"

Conclusion: Not only does this program not deter children from breaking the law, it actually **makes some participants more likely to break the law.** Ultimately, the design choice had an impact on the conclusions drawn.

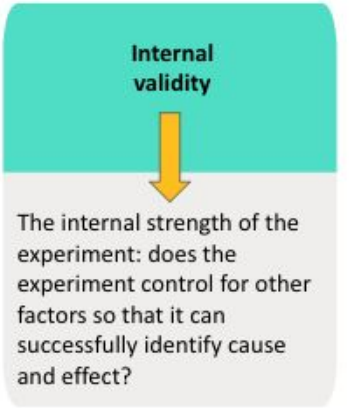
Critical appraisal

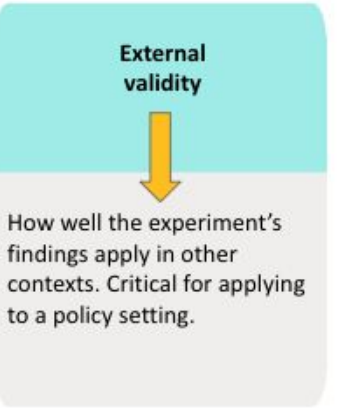
Before drawing conclusions and making recommendations, it is important to evaluate critically the evidence.

Critical appraisal tools helps you appraise the reliability, importance and applicability of evidence.

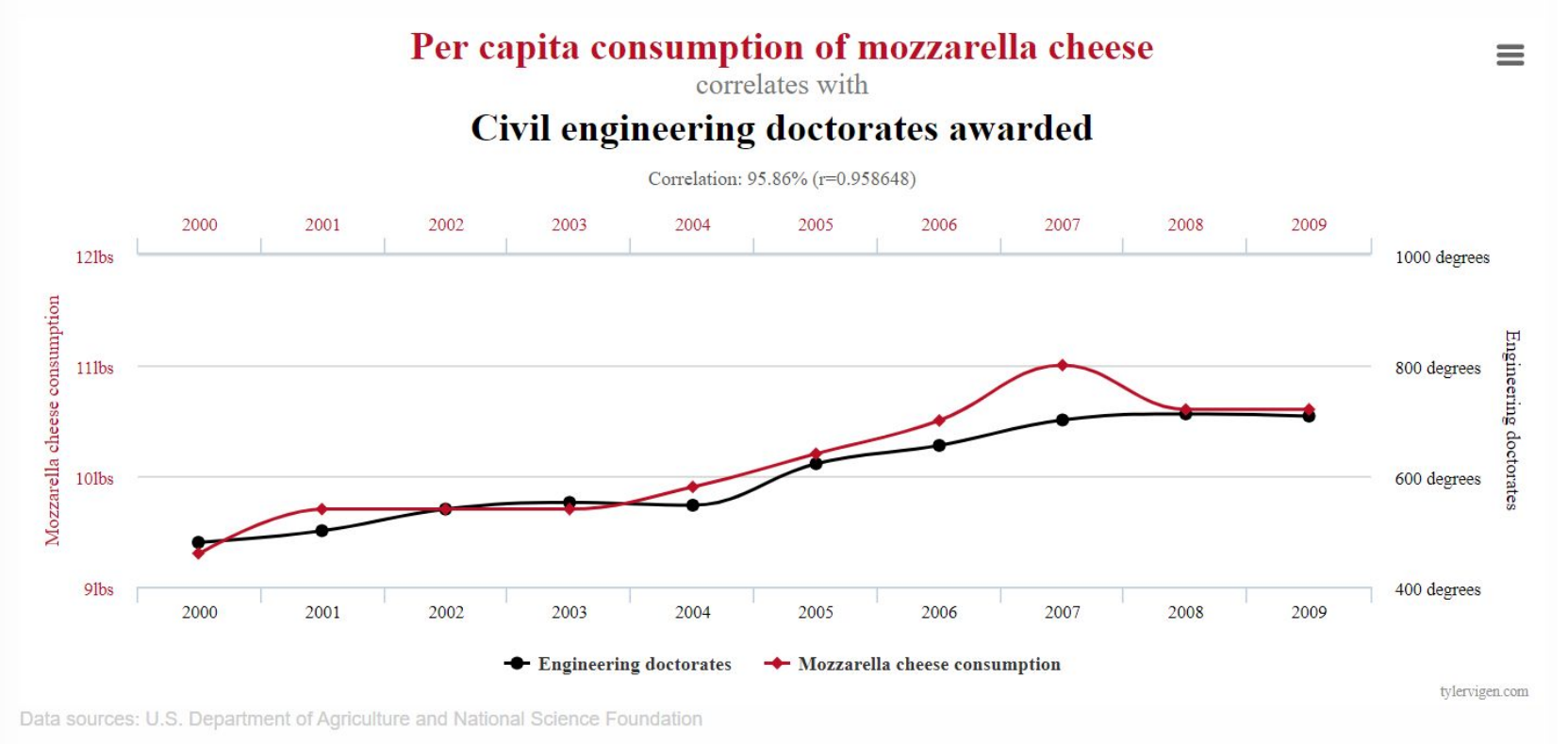
Key questions:

1. Does this study address a clearly focused question?
2. Did the study use valid methods to address this question?
3. Are the results of this study important, in terms of magnitude?
4. Are these valid, important results applicable to my population of interest?



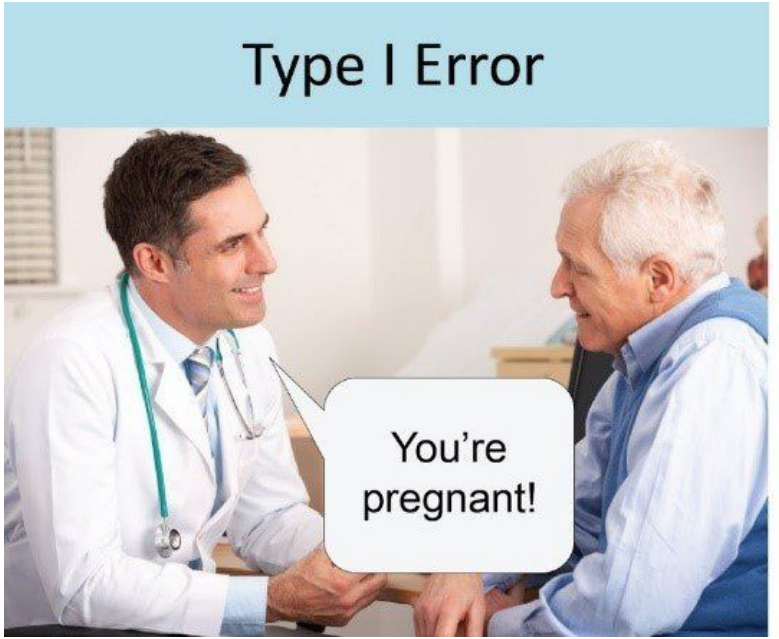


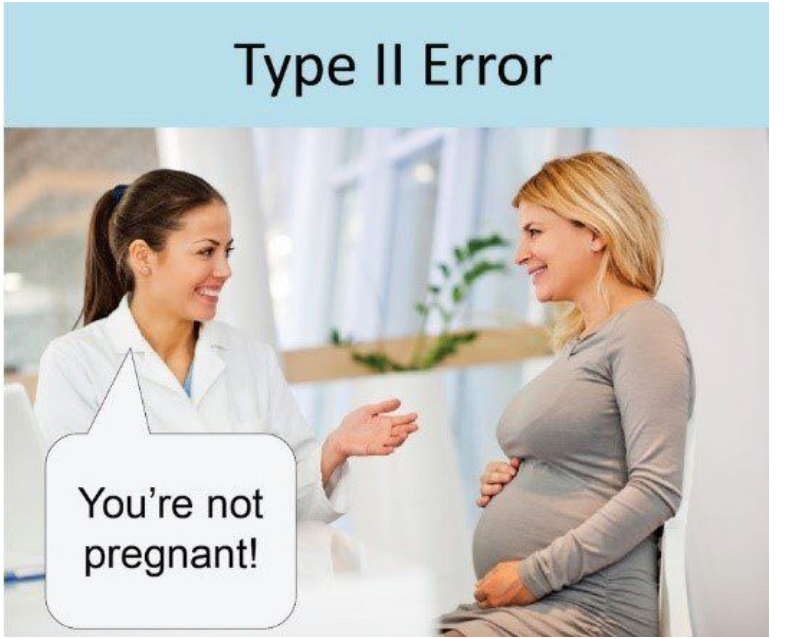
Spurious correlations



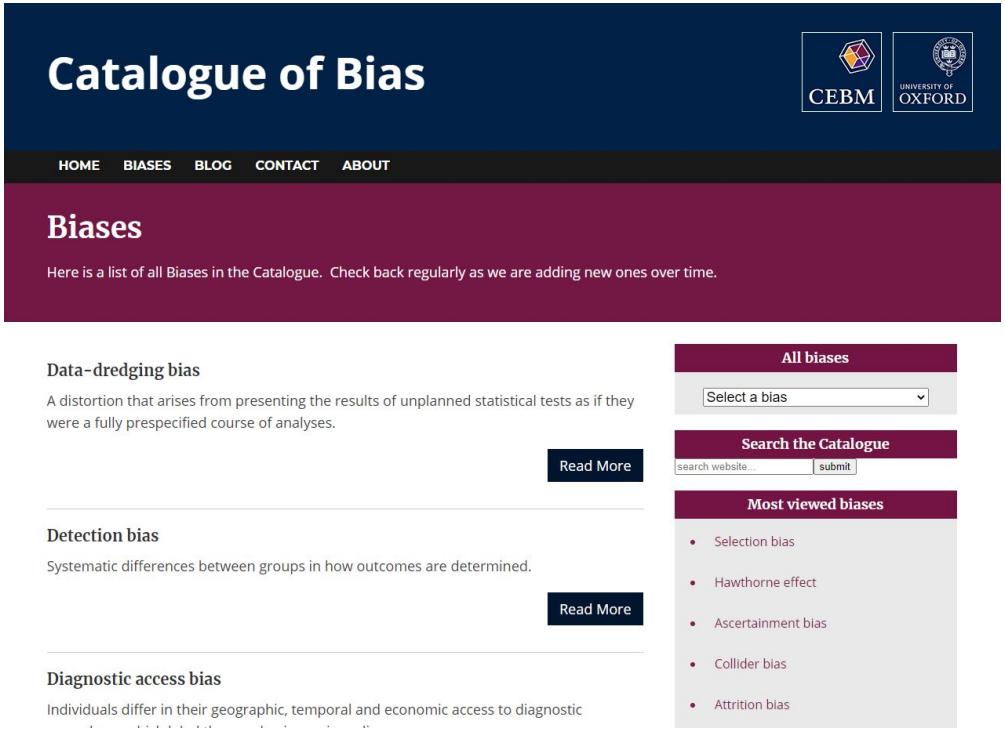
Data sources: U.S. Department of Agriculture and National Science Foundation

Are you drawing the right conclusion?





Risk of bias



https://catalogofbias.org/biases/

Example - Attrition bias

Attrition bias

Unequal loss of participants from study groups in a trial.

Background

Attrition occurs when participants leave during a study. It almost always happens to some extent.

Different rates of loss to follow-up in the exposure groups, or losses of different types of participants, whether at similar or different frequencies, may change the characteristics of the groups, irrespective of the exposure or intervention. Losses may be influenced by such factors as unsatisfactory treatment efficacy or intolerable adverse events.

When participants leave, it may not be known whether they continue or discontinue an intervention; there may be no data on outcomes for these participants after that time.

Systematic differences between people who leave the study and those who continue can introduce bias into a study's results – this is attrition bias. However, the results may not necessarily be biased, despite different drop-out rates in the groups. We discuss below how to assess the impact of different amounts of attrition.

In some cases, those who leave a study are likely to be different from those who continue. For instance, in an intervention study of diet in people with depression, those with more severe depression might find it harder to adhere to the diet regimen and therefore more likely to leave the study.

Preventive steps

Techniques for preventing losses follow-up include ensuring good communication between study staff and participants, accessibility to clinics, effective communication channels, incentives to continue, and ensuring that the study is of relevance to the participants.

However, for many studies, complete follow up is unlikely. In such cases, the reasons for attrition should be carefully considered. After the study has been completed, a number of analysis methods can be used to reduce the impact of attrition bias.

Intention to treat analysis: Because anything that happens after randomisation can affect the chance that a study participant has the outcome of interest, it is important that all patients (even those who fail to take their medicine or accidentally or intentionally receive the wrong treatment) are analysed in the groups to which they were allocated.

It is important that we not only look for the term 'intention-to-treat analysis' in the methods but also look at the results to ensure that the analysis was actually done.

Methods for dealing with missing data include last observation (or baseline value) carried forward, mixed models, imputation and sensitivity analysis using 'worst case' scenarios (assuming that those with no information all got worse) and 'best case' scenarios (assuming that all got better). Analysing data only from participants remaining in the study is called complete case analysis.

A rule of thumb states that <5% attrition leads to little bias, while >20% poses serious threats to validity. While this is useful, it is important to note that even small proportions of patients lost to follow-up can cause significant bias. One way to determine whether losses to follow-up can seriously affect results is to assume a worst-case scenario for the outcomes in those with missing data and look to see if the results would change. If this method doesn't change the study's conclusions, the loss to follow-up is likely not a threat to the study's validity.

https://catalogofbias.org/biases/

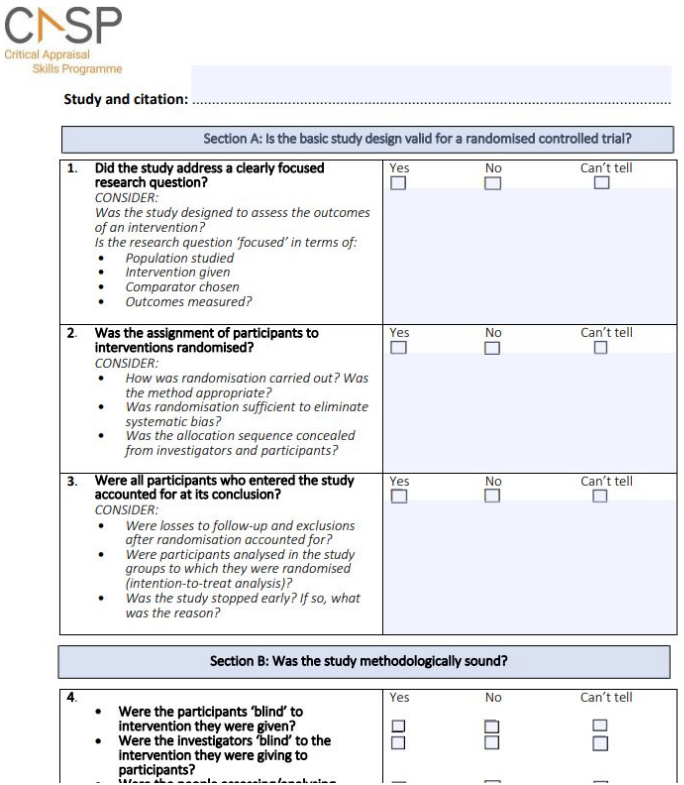
Critical appraisal tools

- Centre for Evidence-Based Medicine (CEBM)
  Tools for the critical appraisal of systematic reviews, diagnostic studies, prognosis studies, & RCTs

- Critical Appraisal Skills Programme (CASP) Checklists
  8 critical appraisal tools that are designed to be used when reading research. Tools for: Systematic Reviews, Randomised Controlled Trials, Cohort Studies, Case Control Studies, Economic Evaluations, Diagnostic Studies, Qualitative studies and Clinical Prediction Rule.

- Joanna Briggs Critical Appraisal Tools
  Contains 13 checklists on: Case Control Studies, Case Reports. Case Series, Cohort Studies, Diagnostic Test Accuracy Studies, Economic Evaluations, Prevalence Studies, Quasi-Experimental Studies (non-randomized experimental studies), Randomised Controlled Trials, Systematic Reviews, Text and Opinion, Analytical Cross Sectional Studies, Qualitative Research

Critical appraisal tools - Example (RCT)



CASP
Critical Appraisal Skills Programme

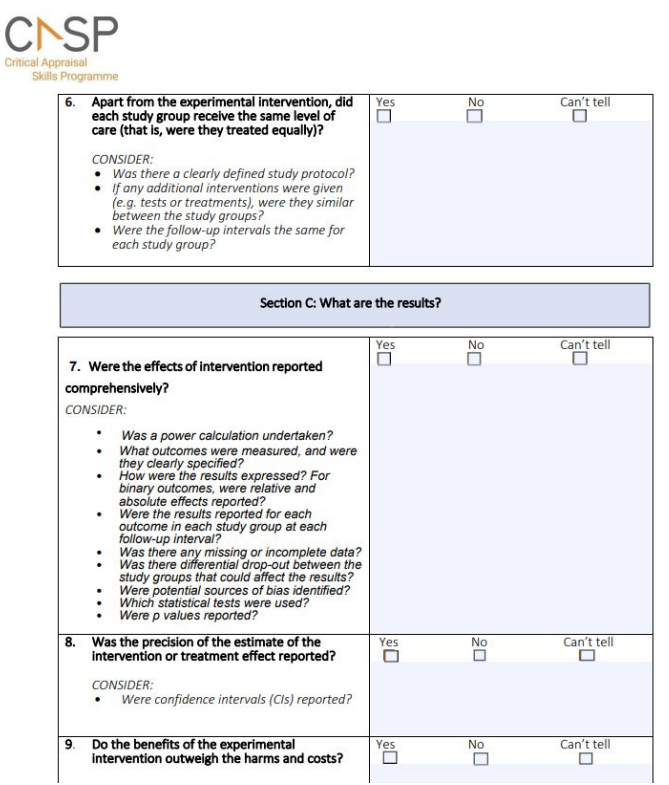
Study and citation: .....

Section A: Is the basic study design valid for a randomised controlled trial?

	Yes	No	Can't tell
1. Did the study address a clearly focused research question? *CONSIDER:* *Was the study designed to assess the outcomes of an intervention?* *Is the research question 'focused' in terms of:* • *Population studied* • *Intervention given* • *Comparator chosen* • *Outcomes measured?*	☐	☐	☐
2. Was the assignment of participants to interventions randomised? *CONSIDER:* • *How was randomisation carried out? Was the method appropriate?* • *Was randomisation sufficient to eliminate systematic bias?* • *Was the allocation sequence concealed from investigators and participants?*	☐	☐	☐
3. Were all participants who entered the study accounted for at its conclusion? *CONSIDER:* • *Were losses to follow-up and exclusions after randomisation accounted for?* • *Were participants analysed in the study groups to which they were randomised (intention-to-treat analysis)?* • *Was the study stopped early? If so, what was the reason?*	☐	☐	☐

Section B: Was the study methodologically sound?

	Yes	No	Can't tell
4. • **Were the participants 'blind' to intervention they were given?** • **Were the investigators 'blind' to the intervention they were giving to participants?** • **Were the people assessing/analysing**	☐ ☐	☐ ☐	☐ ☐



CASP
Critical Appraisal Skills Programme

	Yes	No	Can't tell
6. Apart from the experimental intervention, did each study group receive the same level of care (that is, were they treated equally)? *CONSIDER:* • *Was there a clearly defined study protocol?* • *If any additional interventions were given (e.g. tests or treatments), were they similar between the study groups?* • *Were the follow-up intervals the same for each study group?*	☐	☐	☐

Section C: What are the results?

	Yes	No	Can't tell
7. Were the effects of intervention reported comprehensively? *CONSIDER:* • *Was a power calculation undertaken?* • *What outcomes were measured, and were they clearly specified?* • *How were the results expressed? For binary outcomes, were relative and absolute effects reported?* • *Were the results reported for each outcome in each study group at each follow-up interval?* • *Was there any missing or incomplete data?* • *Was there differential drop-out between the study groups that could affect the results?* • *Were potential sources of bias identified?* • *Which statistical tests were used?* • *Were p values reported?*	☐	☐	☐
8. Was the precision of the estimate of the intervention or treatment effect reported? *CONSIDER:* • *Were confidence intervals (CIs) reported?*	☐	☐	☐
9. Do the benefits of the experimental intervention outweigh the harms and costs?	☐	☐	☐

What should be done at this stage?

Determine what you have learned and what comes next. Based on findings, should your program be modified or scaled up/down? Should your experiment be replicated? Should another experiment take place?

That will depend on the nature of your results.

What does your experiment say?

Null effect

- Suggests your intervention is not effective.
- May mean the intervention should be stopped, but it could also mean that the intervention is not being done in an effective way.
- As experiments typically focus on testing effectiveness, your results won't tell you *why* your intervention didn't work.
- In these cases, it can be worth following with more qualitative approaches to gather feedback and insights from participants.

Positive effect

- Suggests your intervention has the desired result. This is usually a good news story for an audience: their intervention works!
- Still, depending on your experiment you may also be able to give the managers insight into how to improve it further.
- For instance, you may have found an increase in web traffic and access to a form in your intervention group, but no overall increases in form submission. A next logical step would be to experiment with the form layout, display, language, etc.

Negative effect

- Suggests your intervention has the opposite of what you had expected. For instance, your outreach strategy didn't increase uptake in your program but decreased it.
- This should prompt a discussion on what is actually happening in your program to figure out why that is the case (as suggested above, using qualitative strategies).
- In some cases, this may create an ethical obligation to halt the intervention, if you are making people worse off: in others, there may be room for exploring how to improve it or address the shortcomings.

Not what you expected? Therefore, not useful?

Unexpected and Null Results Can Help Build Federal Evaluation Plans and Learning Agendas

Evaluation is a critical part of evidence-based policymaking. When conducting an evaluation of an intervention or program, our goal is to answer a specific question or test a specific hypothesis. After we design or introduce an intervention, we often expect it to result in changing outcomes in a certain direction (increase or decrease), or at a certain size or magnitude (such as a 2 percentage point increase). Yet even well-planned studies do not always meet expectations in terms of effect direction or size.

Recent research shows that null results are more common than we think, and occur for a variety of reasons. We know about this trend because agencies and researchers are increasingly publishing and discussing their null results more publicly.[1] Leading federal agencies publish null results, including the Department of Labor's Chief Evaluation Office (CEO), the Department of Education's Institute for Education Sciences (IES) and the Administration for Children and Families' Office of Planning, Research and Evaluation (OPRE). By sharing both expected and unexpected results, we can learn about what programs work, what effect sizes are realistic, and improve Federal evaluations.

What is an unexpected or null result?

- An unexpected result is one that runs counter to what prior evidence or informed hypotheses would suggest. This could mean an evaluation that shows no evidence of impact (null), a smaller or larger impact than prior evidence would suggest, or one that points to inconsistent program implementation. Such results can update our knowledge about the scope for change in large-scale government services.
- A null result, one type of unexpected result in an experiment, is when there is no statistically significant difference in outcomes between conditions: an intervention and a control (no-intervention) group, or between groups receiving two different versions of an intervention. This does NOT mean that one can conclude that the intervention has *no effect*. Instead, the observed impact was not large enough to stand out against background variation in the evaluation.

https://oes.gsa.gov/assets/files/unexpected-results-2-pager.pdf

What should be done at this stage?

Interpret the evidence with your stakeholders in mind. Understand what your key findings say and tailor those messages to resonate with different stakeholders.

Develop practical guidance informed by evidence. Make it easy for decision makers to interpret and use the evidence you provide.Identify next steps and key recommendations.

Disseminate your findings. A lot of people may benefit from learning about your experiment. Make sure that there is a corporate memory of the work so it is easy for others to access.

Share them publicly, if possible. Your work will contribute to the knowledge base and has the potential to shape program design for many years.

A single experiment is a pixel for an image, the more pixels you add to the broader body of evidence, the clearer the image will be for others.

Evidence dissemination tips

PLOS COMPUTATIONAL BIOLOGY

EDITORIAL

Ten simple rules for innovative dissemination of research

Tony Ross-Hellauer[1]*, Jonathan P. Tennant[2], Vilté Banelytė[3], Edit Gorogh[4], Daniela Luzi[5], Peter Kraker[6], Lucio Pisacane[5], Roberta Ruggieri[5], Electra Sifacaki[7], Michela Vignoli[8]

1 Open and Reproducible Research Group, Institute of Interactive Systems and Data Science, Graz University of Technology and Know-Center GmbH, Graz, Austria, **2** Center for Research and Interdisciplinarity, University of Paris, Paris, France, **3** Freelance Researcher, Vilnius, Lithuania, **4** University and National Library, University of Debrecen, Debrecen, Hungary, **5** Institute for Research on Population and Social Policies, National Research Council, Rome, Italy, **6** Open Knowledge Maps, Vienna, Austria, **7** National and Kapodistrian University of Athens, Athens, Greece, **8** Center for Digital Safety and Security, AIT Austrian Institute of Technology, Vienna, Austria

* tross@know-center.at

Author summary

How we communicate research is changing because of new (especially digital) possibilities. This article sets out 10 easy steps researchers can take to disseminate their work in novel and engaging ways, and hence increase the impact of their research on science and society.

OPEN ACCESS

Citation: Ross-Hellauer T, Tennant JP, Banelytė V, Gorogh E, Luzi D, Kraker P, et al. (2020) Ten simple rules for innovative dissemination of research. PLoS Comput Biol 16(4): e1007704. https://doi.org/10.1371/journal.pcbi.1007704

Editor: Russell Schwartz, Carnegie Mellon University, UNITED STATES

Published: April 16, 2020

Funding: This work was partly funded by the OpenUP project, which received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 710722. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: We have read the journal's policy and have the following conflicts: TR-H is Editor-in-Chief of the open access journal *Publications*. JT is the founder of the Open Science MOOC, and a former employee of ScienceOpen.

Introduction

As with virtually all areas of life, research dissemination has been disrupted by the internet and digitally networked technologies. The last two decades have seen the majority of scholarly journals move online, and scholarly books are increasingly found online as well as in print. However, these traditional communication vehicles have largely retained similar functions and formats during this transition. But digital dissemination can happen in a variety of ways beyond the traditional modes: social media have become more widely used among researchers [1,2,3], and the use of blogs and wikis as a specific form of 'open notebook science' has been popular for more than a decade [4].

Professional academic social networks such as ResearchGate and Academia.edu boast millions of users. New online formats for interaction with the wider public, such as TED talks broadcast via YouTube, often receive millions of views. Some researchers have even decided to make all of their research findings public in real time by keeping open notebooks [5,6]. In particular, digital technologies invoke new ways of reaching and involving audiences beyond their usual primary dissemination targets (i.e., other scholars) to actively involve peers or citizens who would otherwise remain out of reach for traditional methods of communication [7]. Adoption of these outlets and methods can also lead to new cross-disciplinary collaborations, helping to create new research, publication, and funding opportunities [8].

Beyond the increase in the use of web-based and computational technologies, other trends in research cultures have had a profound effect on dissemination. The push towards greater public understanding of science and research since the 1980s, and an emphasis on engagement and participation of non-research audiences have brought about new forms of dissemination [9]. These approaches include popular science magazines and science shows on television and

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7161944/

Group exercise - Context

COVID-19 has dramatically changed the way work is done in government. This new reality is placing increased pressure on the public service, but also presenting new opportunities for how teams can cope and thrive in a distributed work environment.

Multiple changes have taken place, but their impacts have yet to be measured/observed.

For instance, **the technologies we use now make it possible to be always ‘on’ and connected, make it harder to disconnect and have clearly delimited leisure/rest time.**

Group exercise - Context

- **Intervention**: Imagine that your organization is currently testing time-bound restrictions to email servers (i.e. no email being sent after 6:00pm or during weekends, unless urgent)
- **Design:** Departmental units are randomly assigned to have IT server limited, and not others
- **Outcomes of interest**: self-reported individual stress level
- **Main hypothesis:** Limiting email server activity will reduce employee stress levels

Imagine that two different departments implement this design, leading to two different results

Group exercise - Instructions

Go to this jamboard: https://jamboard.google.com/d/1ehXTgSzjkHWgNQfS4W2eWg7vYv1nx4ehls5ZXJVMt-U/viewer?f=0

Breakout room 1 - Scenario

Your team led the project in department A

Main hypothesis: Limiting email server activity will reduce employee stress levels

Results

- Decrease in stress levels
- Possibly due to removing expectations about receiving emails/tasks outside working hours

Context

- Management is not keen to maintain IT disruption in the long run

Breakout room 2 - Scenario

Your team led the project in department b

Main hypothesis: Limiting email server activity will reduce employee stress levels

Results

- Increase in stress levels
- Possibly due to having to manage batched emails coming in at the same time

Context

- Management is keen to move quickly on the results of the trial to implement best practices

How do you present the findings to your executives? What do you recommend doing next ?