|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Plan** | **Checklist** | **Chapter title** | **Learning objectives** | **Abstract (100-200 words)** |
|  | Should be able to make links from the Exercises to ch 23  Seems that formatting of shaded bit is OK if line space between html and text is removed  Ch 6 Arrowsmith program is example of this working with the original formatting.  Ch 8 has maze bright eg using ::: {#custom} and does succeed in embedding a figure in the shaded box  Ch 9 vignette worked with <p id="vignette">  <!-- reminder to self: blogposts referenced in the book should be uploaded to figshare or another place with DOI--> |  |  |  |
|  | All acknowledgements complete | Preface: Why did we write this book? | n/a | n/a |
| 9/6 | Fix vignette text boxes | Chapter 1: Introduction | **Learning objectives**:   * Understand and explain the importance of evaluating interventions * Recognise and distinguish between examples of random error and systemic bias | Three vignettes are used to introduce the problem confronting a practitioner: clients may improve with intervention, but it can be hard to know if they would have improved anyhow. The importance of evaluating interventions is discussed – to ensure clients receive evidence-based interventions that will not do harm, to avoid wasting resources on ineffective interventions, and to provide a stimulus to develop new methods. The chapter then introduces the key concept of measurement error, and the distinction between random error and systematic bias. |
| 9/6 | Fix vignette text boxes  Possibly add something about Ioannidis study of nutrition?  Add links to Exercises in ch 23 | Chapter 2: Why observational studies can be misleading | **Learning objectives**:   * Describe potential confounds in observational studies * Reflect on how these can be controlled for in experimental research designs | A further three vignettes illustrate observational studies, and the ways in which reliance on such evidence can be misleading. The notion of a confound – a factor that induces a spurious correlation between intervention and outcome, because it has an impact on both – is introduced.  Readers are encouraged to find examples of newspapers reports of factors that are either thought to lead to health risks or benefits, and to reflect on possible confounds. |
| 10/6 | See email re target diagram – attribution needed. May also want more on validity – idea that measures don’t always measure what the title suggests  Still lots of problems formatting boxes, but may be best to fix these in LaTex  Do a version of MLU plot with error bars? Might be useful for ch 23? We do have the data | Chapter 3: How to select an outcome measure | **Learning objectives**:   * Understand the key concepts of reliability, validity, sensitivity and practicality * Evaluate test content in relation to research goals | This chapter notes the key questions to ask of an outcome measure:  The reader is introduced to types of measurement, and the properties nominal, ordinal, interval and ratio scales, before moving on to consider questions to ask of a potential outcome measure:   1. Is the measure reliable? 2. Is it valid? i.e., does it measure what I want to measure? 3. Are there norms? 4. Is it sensitive? 5. Is it practical? |
| 11/6 |  | Chapter 4: Improvement due to nonspecific effects of intervention | **Learning objectives:** -  Be aware of non-specific intervention effects and how these can be separated from the specific impact of an intervention.  Understand the concept of mechanism measures, as distinct from broader outcome measures. | The chapter starts with a brief introduction to placebo effects, which are well-known in medicine, where one can show that an inert pill may be as effective as an active drug. It is argued that similar nonspecific effects can occur with behavioural interventions, where there may be positive benefits from involvement of a teacher or therapist. The Hawthorne effect is briefly discussed: this was first described in a study of factory working conditions, which showed almost any change led to greater productivity: however, subsequent evaluation of the study has cast some doubt on the findings.  Placebo effects may be detected by use of appropriate controls. In addition, the reader is introduced to the idea of incorporating measures of mechanism into study design, to check whether an intervention does change the underlying processes that are postulated to lead to benefits. |
| 12/6 | Maybe look more closely at EasyPeasy intervention report by EEF, which is used to illustrate idea of mechanism measures.  For this and previous chapters, may just remove options re bw vs colour as this creates problems. Better just treat bw as default. As explained in comment in index, will save colour versions and then if we want to make a version with these we just change folder names | Chapter 5: Limitation of the pre-post design: biases related to systematic change | **Learning objectives:**  - Describe three limitations of common pre-post intervention designs and how these can be addressed | The focus of this chapter is three reasons why a person’s score on an outcome might improve after intervention, even if the intervention is ineffective. The first is spontaneous improvement, which may be particularly evident in children, where maturation can lead to greater skills. It is also an issue when studying people who are recovering from brain injury.  The second factor is practice on the measures used to evaluate outcome. Simple familiarity with language tests, for instance, might lead someone to perform much better on re-testing, as they know what to expect, and understand the task better.  Finally, we consider regression to the mean. This is a purely statistical phenomenon, whereby if our measures are not highly reliable, there will be a trend for those who start out with low scores to improve on retesting. |
| 13/6 | Text box Arrowsmith needs formatting – see if removing line between section command has worked – YES! | Chapter 6: Controlling unwanted effects with a control group | **Learning objectives**: By the end of this chapter, you will be able to:   * Explain the importance of control groups * Understand the strengths and weaknesses of study designs using untreated or active controls | The chapter starts by discussing ethical objections that are sometimes raised against the inclusion of a control group in intervention studies. These arguments are problematic because they ignore possible harms from administering unevidenced treatments. It is unethical to invest the time of professionals and clients, and to waste research funds, on studies that cannot determine if the intervention is effective.  The chapter then moves on to discuss the differences between having untreated controls vs active controls who are given an alternative intervention. |
| 13/6 | Chapter is a bit boring. Any way to liven it up – maybe with an example? | Chapter 7: Controlling for selection bias: randomized assignment to intervention | **Learning objectives**: By the end of this chapter, you will be able to:   * Explain why non-random allocation to intervention is problematic * Be aware of the differences between simple randomization, stratification and minimization | A key element of a randomized controlled trial is the process of randomization, which is the topic of this chapter. It is important to assign people to control and experimental groups in a way that minimizes bias.  The advantages and disadvantages of a range of randomization methods are discussed, including simple randomization, permuted blocks, unequal randomization, stratification, adaptive randomization and minimization. The notion of randomizing by groups (clusters) rather than individuals is introduced, to prepare the way for discussion of Cluster Randomized Controlled Trials in Chapter 16. |
| 13/6 |  | Chapter 8: The researcher as a source of bias | **Learning objectives**: By the end of this chapter, you will be able to:   * Explain the concept of experimenter bias, and describe ways in which it can be counteracted * Appreciate how conflict of interest may lead researchers to report unduly rosy findings | Chapter 8 notes that in medical trials there is widespread use of “allocation concealment” or blinding, so that the researcher is kept unaware of who is in an intervention group, but this may not be feasible in the case of behavioural interventions. Researchers need to be aware that, even with the best of intentions, they may bias results in favour of an intervention group, at the point of recruitment to the study, administering assessments or data analysis. The notion of conflict of interest (COI) is discussed, noting that this extends beyond financial considerations. |
| 14/6 | Wonder if EEF has any useful guidance on sampling  Decided to delete exercise 3 – found a v good website by Eekhout that covers missing data very comprehensively and have now pointed readers to that. | Chapter 9: Further potential for bias: volunteers, dropouts, and missing data | **Learning objectives**: By the end of this chapter, you will be able to:  - Distinguish potential sources of bias in who gets included in a study  - Understand different approaches to handling missing data in an intervention study | The topic of Chapter 9 is how bias can creep into an intervention evaluation at the recruitment stage or during the course of the study. The chapter then moves on to discuss how to analyse datasets with missing data and the terms “per protocol”, “intention to treat” and “instrumental variable analysis” are introduced. Imputation of missing data is discussed, noting the importance of determining the reason for missingness when deciding how to proceed. The CONSORT flow diagram is introduced. |
| 16/6 | (moved forward – was after Power/Phacking)  Would be nice to have a cartoon! | Chapter 10: The randomized controlled trial as a method for controlling biases | **Learning objective:** By the end of this chapter, you will be able to explain which biases from previous chapters are adequately controlled in a randomized controlled trial | This short chapter illustrates how the biases and sources of error discussed in previous chapters are controlled for in a classic Randomized Controlled Trial. It also emphasises the importance of full reporting of study details, as emphasised in CONSORT. |
| 19/6 | (previously included as part of RCT chapter; now stand-alone and incorporates some of the stats concepts from chapter on power). I’ve done a lot of work on this but there always seems more to do. But I’ll return after a break.  Has several sections in boxes to be formatted | Chapter 11: Analysing a two-group RCT | **Learning objectives:** By the end of this chapter, you will be able to:  - Appreciate why we need to measure variation as well as average effects of an intervention  - Understand the terms "standard deviation", "confidence interval" and "p-value"  - Interpret output from a t-test  - Explain why Analysis of Covariance is often recommended to analyse outcome data | Statistical analysis of trial results is introduced by first emphasising how analysis looks at variation in scores related to the intervention relative to within-group variation. This leads to an explanation of standard deviation and confidence intervals. Null Hypothesis Significance Testing is briefly introduced, and illustrated with a real dataset from a repository. This is used to show importance of plotting data, and to illustrate the difference between analysing difference scores vs Analysis of Covariance treating the baseline as a covariate. Brief mention is made of nonparametric tests and linear mixed models. |
| 15/6 | Figure 12.6 should be embedded in the grey box.  Need better explaining of exercises    Like ch 11, still a bit flaky in places. | Chapter 12 How big a sample do I need? Sampling, statistical power and type II errors | **Learning objectives**: By the end of this chapter, you will be able to:  - Understand how effect size is measured and how it relates to statistical power  - Explain what a type II error is, and how it may arise when sample size is too small | An extreme example – comparing an intervention study with 2 cases per group with one with 1000 cases per group – is used to explain why we can’t just use mean differences to evaluate intervention. We need also to consider sample size, which influences how well we estimate variation. The concept of effect size is introduced, using simulated data showing how two groups with the same mean difference may differ in terms of how much overlap there is between groups.  This chapter also explains how Type II errors are high when sample size is small, before introducing the formal definition of statistical power. The chapter concludes with advice on how to increase statistical power. |
| 20/6 | Cartoon for ‘did not survive Bonferroni correction’  Would also be good to have an exercise with real data – ideally showing Cis for several tests. I have downloaded some data from Cesaerian study from PONE – they omitted 5 variables so could be used re Bonferroni. But can add this later maybe | Chapter 13: False positives, p-hacking and multiple comparisons | **Learning objectives**: By the end of this chapter, you will be able to:  - Explain what a type I error is and how it relates to p-hacking  - Understand why we need to correct for multiple comparisons, and how this can be done | Understanding of Type I and Type II errors is reinforced using the story of the Boy who Cried Wolf: the villagers first thought there was a wolf when there wasn’t (Type I) but subsequently thought there was no wolf when there was (Type II).  Reasons for false positives are discussed, and the importance of interpretating p-values in context is emphasised. This leads to a discussion of familywise error rate and the need to correct for multiple testing. The Bonferroni correction, MEff and principal component methods are described. The chapter concludes by noting the potential to increase statistical power by appropriate analysis of multiple outcomes. |
| 20/6 |  | Chapter 14: Drawbacks of the two-arm RCT | **Learning objectives**: By the end of this chapter, you will be able to:  - Explain the difference between treatment efficacy and treatment effectiveness  - Consider whether the RCT approach is feasible for evaluating a specific intervention | The conflict is noted between the requirements of an idealised RCT and feasibility of what clinicians can do. Large and homogeneous samples are not always possible to achieve – in part this is due to limited resources, but it may also be the case that a condition of interest is characterised by wide individual variation. These considerations lead to a discussion of the contrast between efficacy and effectiveness, and the development of pragmatic trials that are more compatible with real-world practice. |
| 22/6 |  | Chapter 15: Moderators and mediators of intervention effects | **Learning objectives**: By the end of this chapter, you will be able to:  - Explain the difference between mediators and moderators  - Identify potential mediators and moderators for interventions of interest | This chapter provides a simple overview of Moderators and Mediators, emphasising the contrast between them, and the dangers of applying moderator-mediator analysis uncritically. |
| 23/6 |  | Chapter 16: Adaptive Designs | **Learning objectives**: By the end of this chapter, you will be able to:  - Understand the characteristics of an adaptive design and recognise when it may be applicable | This short chapter discusses the pros and cons of adaptive trial designs, in which there is a planned interim analysis, which can be used to drop ineffective intervention arms, or modify the proportions of individuals allocated to specific interventions. |
| 24/6 | Maybe other exercises? | Chapter 17: Cluster Randomized Controlled Trials | **Learning objectives**: By the end of this chapter, you will be able to:  - Understand when a cluster RCT is preferable to a regular RCT  - Explain the limitations of a cluster RCT | The reader is introduced to the notion of sampling by cluster (e.g. schools, hospitals, cities) rather than by individuals. It is noted that this method is unavoidable when intervention is delivered to whole groups, e.g. classrooms. Advantages of the method are outlined, including avoidance of contamination between interventions, and ease of administration. Disadvantages, notably inefficiency and need for large samples, are then described. |
| 25/6 | Need to finalise answer to exercises | Chapter 18: Cross-over designs | **Learning objectives**: By the end of this chapter, you will be able to:  - Understand how a cross-over RCT works, and specify its advantages and disadvantages compared to a regular RCT | The cross-over design is introduced in the context of a drug trial, where treatment effects are expected to be short-lasting. In that case, we can use a ‘washout’ period after the first phase, and then reverse the assignment of intervention and control groups. It is noted that this method is seldom suitable in the context of fields such as speech and language therapy, where intervention is intended to produce lasting effects. The wait-list control method is an alternative that provides some of the same advantages as the cross-over design. |
| 26/6 | Figs created in base R less good – can we save at higher dpi?  NEED RESPONSES TO EXERCISES | Chapter 19: Single case designs | **Learning objectives**: By the end of this chapter, you will be able to:  - Explain how different types of single case design can control for the kinds of bias discussed in previous chapters  - Understand the limitations of single case designs for interventions that aim to have persistent effects, and how to counteract these | It is noted that classic clinical trial methods work best with large homogenous groups, whereas many clinical and health practitioners work with heterogeneous populations, who may need personally-tailored interventions. The single case design method is introduced with description of a simple A-B-A design that is common in drug studies, where the impact of an intervention is transient. It is noted that this is seldom appropriate for behavioural interventions, where the goal is to produce longer-term change. An alternative approach is to compare outcomes on trained vs untrained items. Such approaches may be combined with methods that stagger the timing of introduction of different intervention elements.  The N-of-1 approach is often extended into a case series, where the focus is not on group outcomes, but on assessing an intervention with a heterogeneous sample. Benefits and drawbacks of the single case approach are summarised. |
| 27/6 | Why this at top of chapter?  ---  output:  pdf\_document: default  html\_document: default  ?cartoon file drawer  Cites bishopblog | Chapter 20: Can you trust the published literature? | **Learning objectives**: By the end of this chapter, you will be able to:  - Understand how publication bias distorts the published literature, leading to intervention effects being overestimated  - Appreciate the importance of citing negative as well as positive results when reviewing the literature on an intervention | The notion of publication bias – the tendency for only positive findings to make it into the research literature – is discussed, and it is noted that it persists, despite having been identified as a problem decades ago.  Citation bias is a related problem: even when articles get published, they are less likely to get cited by others if they have negative findings. |
| 28/6 |  | Chapter 21: Preregistration and Registered Reports | **Learning objectives:** By the end of this chapter, you will be able to:  - Understand how pre-registration of study protocols can help counteract publication bias  - Explain how a Registered Report differs from a standard pre-registered study | This chapter introduces the notion of study registration, which has been used for some years in clinical trials research, but is starting to be adopted in other areas. A good pre-registration specifies a protocol that details rationale, methods and analysis plan, which is registered and time-stamped before data is collected. This makes it possible to detect any decisions about analysis that were made after the data were inspected. Registered reports are a more stringent form of pre-registration, in which the protocol is peer-reviewed before data is collected, allowing the researchers to benefit from expert feedback. If the reviewers and editor are satisfied, the article than can be accepted “in principle” provided the authors follow the planned protocol – or provide convincing reasons for not doing so. The benefits of the registered report format are discussed. |
| 29/6 | Check Hemingway citation | Chapter 22: Reviewing the literature before you start | **Learning objectives**: By the end of this chapter, you will be able to:  - Describe the key features of a Systematic Review  - Understand the main criteria used for quality appraisal in a Systematic Review | The chapter starts by describing the features of a systematic review, a method of synthesizing research literature developed by the Cochrane Collaboration in the field of clinical trials.  A full systematic review is only feasible in fields where there is a reasonable amount of research on interventions, and where there are adequate resources to do the screening and analysis of published papers. We argue that, nevertheless, some general principles of good practice can be taken from systematic review guidelines and applied to smaller-scale literature reviews. It is noted that the goal is to guard against random error and systematic bias in the synthesis of the literature. |
| 30/6 |  | Chapter 23 |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

Ch 16 notes

```{=html}

<!--

## More niche trial designs

### Just-in-Time adaptive interventions (JITAI)

### Micro-Randomized Trials (MRT)

### Sequential, Multiple Assignment, Randomized Trial (SMART)

- A review of this approach in the field of autism is provided by @

Kasari, C., Sturm, A., & Shih, W. (2018). SMARTer Approach to Personalizing Intervention for Children With Autism Spectrum Disorder. Journal of Speech, Language, and Hearing Research, 61(11), 2629–2640. https://doi.org/10.1044/2018\_JSLHR-L-RSAUT-18-0029

- more general review for education:

Nahum-Shani, I., Almirall, D., & Buckley, J. (2019). An Introduction to Adaptive Interventions and SMART Designs in Education (p. 45). US Department of Education: Institute of Education Sciences.

The Nahum-Shani is a v useful introduction, but I do worry that this approach does just divide people into 'responders' and 'nonresponders' in exactly the way we argued against in chapter 1! - ie assuming measures are perfectly reliable and so can be validly used to subdivide cases.

possibly relevant:

https://pubs.asha.org/doi/full/10.1044/2018\_JSLHR-L-RSAUT-18-0029?url\_ver=Z39.88-2003&rfr\_id=ori%3Arid%3Acrossref.org&rfr\_dat=cr\_pub++0pubmed

https://ies.ed.gov/ncser/pubs/2020001/pdf/2020001.pdf

https://academy.pubs.asha.org/2013/11/need-for-adaptive-research-designs-in-speech-language-pathology/

https://www.youtube.com/watch?v=ythfdGd63hU

ch 18

```{=html}

<!---We definitely need to cover this, as it is not uncommon esp in educational contexts.

But thinking about this, I'm not sure how it differs from regular crossover apart from making it easier to recruit. And I guess it is possible to make it adaptive so the waitlist people only get intervention if analysis at end of phase 1 shows it is promising.

Also wonder about the analysis: I googled and someone said this,

"Have three timepoints for everyone, t0 at baseline, t1 when the control group switches to treatment, and t2 after about double that time. At t1 you have a between subject treatment comparison of treatment versus control adjusted for the t0 baseline covariate. You also have a within subject comparison within the control-treatment group. The right linear mixed model can combine those estimates under certain assumptions. And stronger assumptions can be made of course if blinding occurs and neither group knows who is on control or treatment to time t1." -->

```