

Note: Please excuse me for my "telegram-like" writing style. This is simply due to the intense additional workload I currently have due to the COVID-19 outbreak in Switzerland which forced me to keep communication more crisp than usual.

p1:

- More recent references regarding irreproducibility of animal research are warranted. References 1 & 2 are ~ 10 years old. Any updates since then?
- Passage 6: Is there a reference for the claim regarding reduced moral among research scientists due to irreproducibility?
- Passage 7: Randomized block design doesn't have reproducibility "built-in" per se. Irreproducibility might occur due biases that are not prevented by this randomisation protocol (e.g. performance bias or p-hacking).
- Passage 8: Randomised controlled experiments were not invented by Fisher. Fisher heavily popularised them, though (along with other influential statisticians such as Neyman, Pearson or Gosset). See for example: "Oakley (1998) Experimentation and social interventions: a forgotten but important history"

p2:

- Passage 2: Please indicate the page/chapter you're referring to in the referenced book.
- Passage 5: I'd rephrase the passage starting with "Fisher developed a number of ..." because the description of CR and RB designs will probably rather confusing for lay people. For example, the description of the RB design could be misunderstood by some in the way that each block must only consist of one subject and that said subject needs to receive each of the treatments tested.
- Passage 6: Reference for the claim that RB is the most common design across disciplines?
- Passage 6: The claim that the using either CR or RB does not result in irreproducibility is unfounded since there are other sources of bias that can render a study irreproducible even if CR or RB are used.
- Passage 7: Citation needed for this passage. Also: Increased power due to randomized block design is only given if the blocks that are chosen account for the variation that is supposed to be reduced. If the blocks are ill-chosen, the power would be lower.

p3:

- Passage 1: Citation for ANOVA by Fisher would be nice (alternatively, S. S. Wilks has a beautifully written paper "Certain Generalizations in the analysis of variance" from 1933)
- Passage 4: Calling the CR generally less powerful than the RB design can be misleading. Whether CR design is less powerful than RB design depends - among other things - on the correct choice of the blocks.

p4:

- Passage 2: Same remark as above: Correct choice of blocks is key when it comes to increasing power. RB with wrong blocks can be less powerful than CR design (see e.g. Krzywinski & Altman, Analysis of variance and blocking, 2014).
- Passage 3: Reference needed for claim that RB is most widely used.
- Passage 3: I'd change this to "repeatability built into it". As mentioned above, RB does not guarantee reproducibility per se.
- Passage 3: Claim "it is rarely used because scientists have not been told about it" is at odds with claim on page 2 that CB is "most common design across all disciplines".
- Passage 4: What are the references for the citations?

p6:

- passage 1: Which Wikipedia page is meant here? please add link.

- passage 4: biases that can be avoided should be specifically mentioned lest to give the impression that correct randomisation helps avoiding all possible experimental biases
- passage 6: Please add more details wrt to the selection of the publications. It's unclear what the reasoning was behind choosing these keywords was (why were there more search keywords linked with "mouse" than with "rat") and how the publications were chosen. For example, searching for "mouse" and "heart" on PubMedCentral yields ~240'000 publications between 2012 and 2019 - how were the 11 publications that were used for the study selected?
- passage 6: Why does the author conclude that this sample represents a "reasonably unbiased collection of published pre-clinical research papers"? How was this ensured?

p7:

- passages 4-7: references to cited sentences are missing