Reproducible Manuscripts

Alexander Mark Weber

2024-06-26

Notes

Bridge-In

Mentimeter

Scientific Method

Scientific Method

What is the scientific method (broadly)?

1. Define a question
2. Gather information and resources (observe)
3. Form an explanatory hypothesis
4. Test the hypothesis by performing an experiment and collecting data in a reproducible manner
5. Analyze the data
6. Interpret the data and draw conclusions that serve as a starting point for a new hypothesis
7. Publish results
8. Retest (frequently done by other scientists)

Scientific Method

7. Publish Results
8. Retest (frequently done by other scientists)

Problem


Essay

Why Most Published Research Finding
Are False

John P. A. Ioannidis

- In 2011, John Ioannidis[1] published
- Why?
  - Studies are underpowered
  - Current incentives lead scientists to publish quantity over quality
  - No incentives for scientists to replicate other studies
  - More...

John suggested that the majority of all published papers at the time were likely not true. Or put another way, wouldn't be reproduced

Problem

Was he right?

- In 2015, the Open Science Collaboration sampled studies from prominent journals to estimate the replicability of psychological research.[2]

Problem

Problem

```
# Load ggplot2
library(ggplot2)
library(tidyverse)
```

[1] Ioannidis JPA (2005) Why most published research findings are false. PLoS Med 2(8): e124.
[2] Open Science Collaboration. Estimating the reproducibility of psychological science. Science 349, aac4716 (2015).

```
-- Attaching packages --------------------------------------- tidyverse 1.3.1 --

v tibble  3.2.1     v dplyr   1.1.4
v tidyr   1.3.1     v stringr 1.5.1
v readr   2.1.5     v forcats 0.5.1
v purrr   1.0.2

-- Conflicts ------------------------------------------ tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()    masks stats::lag()
```

```r
# Create Data
data <- data.frame(
  group=c("Successful", "Unsuccessful"),
  value=c(39,61)
)

data <- data |>
  arrange(desc(group)) |>
  mutate(prop = value / sum(data$value) *100) |>
  mutate(ypos = cumsum(prop)- 0.5*prop )

# Basic piechart
ggplot(data, aes(x="", y=value, fill=group)) +
  geom_bar(stat="identity", width=1, color="white") +
  coord_polar("y", start=0) + theme_void() +
  theme(legend.text = element_text(size=18), legend.title = element_blank())+

  geom_text(aes(y = ypos, label = value), color = "white", size=6) +
  scale_fill_brewer(palette="Set1")
```

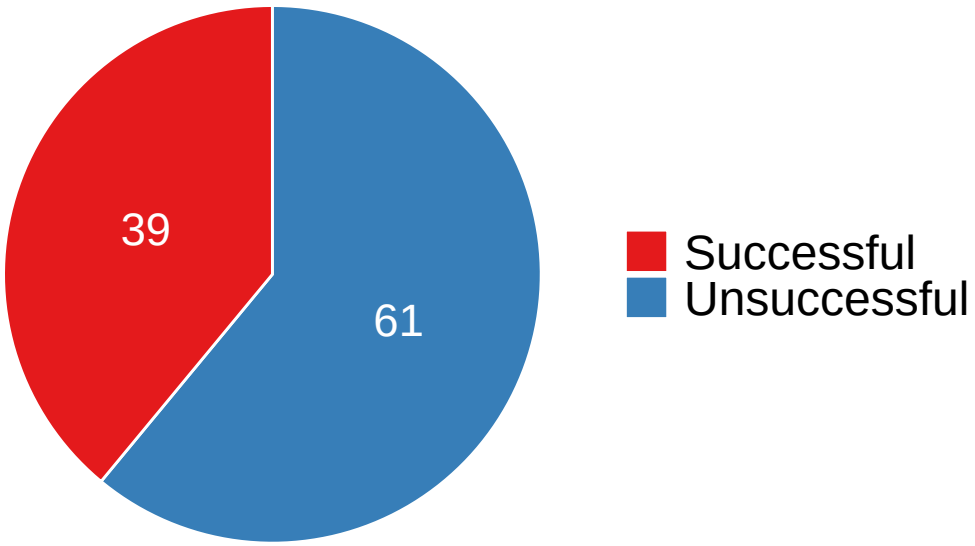


Figure 1

Out of 100 independently performed replications, only 39% were subjectively labelled as successful replications, and on average, the effects were roughly half the original size [3]

[https://www.nature.com/articles/s44271-023-00003-2]

Out of 100 independently performed replications, only 39% were subjectively labelled as successful replications, and on average, the effects were roughly half the original size

Problem

Not just in Psychology:

- animal behaviour[4];
- cancer biology[5];
- economics[6]
- pharmaceutical industry[7]

[3]https://www.nature.com/articles/s44271-023-00003-2

[4]Farrar, B. G., Boeckle, M. & Clayton, N. S. Replications in comparative cognition: what should we expect and how can we improve? Anim. Behav. Cognit. 7, 1 (2020).

[5]Errington, T. M. et al. Investigating the replicability of preclinical cancer biology. Elife 10, e71601 (2021).

[6]Camerer, C. F. et al. Evaluating replicability of laboratory experiments in economics. Science 351, 1433–1436 (2016).

[7]Begley CG, Ellis LM (2012) Drug development: Raise standards for preclinical cancer research. Nature 483: 531–533. doi: 10.1038/483531a PMID: 22460880

- neuroscience[8]
- neuroimaging[9]
- clinical trials[10]

Problem

- For clinical trials: 44% contained at least some flawed data:[11]
  - impossible statistics,
  - incorrect calculations,
  - or duplicated numbers or figures
  - 26% of trials were impossible to judge: either due to incompetence or faked data

Problem

- Publishing irreproducible results is worse than not publishing: more difficult to eliminate an idea than it is to introduce it[12]
- Spurious results can mislead other researchers who conduct follow-up investigations or try to integrate findings into broader theories.

For the clinical trials; for more than 150 trials, the author of the paper got access to anonymized individual participant data (IPD). By studying the IPD spreadsheets, he judged that 44% of these trials contained at least some flawed data: impossible statistics, incorrect calculations or duplicated numbers or figures, for instance. And 26% of the papers had problems that were so widespread that the trial was impossible to trust, he judged — either because the authors were incompetent, or because they had faked the data.

What Can We Do?

- Many solutions are needed; far outside the scope of this talk

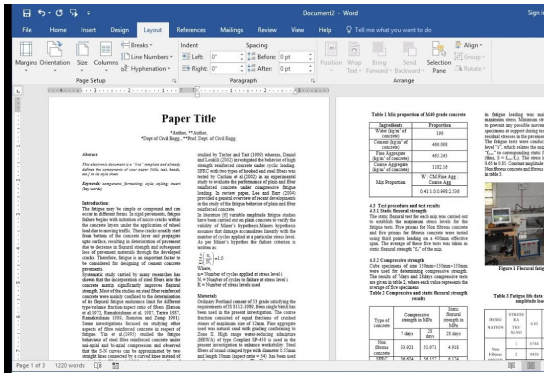
[8] K.S. Button, J.P.A. Ioannidis, C. Mokrysz, B.A. Nosek, J. Flint, E.S.J. Robinson, M.R. Munafò. Power failure: Why small sample size undermines the reliability of neuroscience. Nat Rev Neurosci, 14 (2013), pp. 365-376

[9] Marek, S., Tervo-Clemmens, B., Calabro, F.J. et al. Reproducible brain-wide association studies require thousands of individuals. Nature 603, 654–660 (2022). https://doi.org/10.1038/s41586-022-04492-9

[10] Carlisle, J. B. Anaesthesia 76, 472–479 (2021).

[11] Carlisle, J. B. Anaesthesia 76, 472–479 (2021).

[12] C. Piller. Disgraced COVID-19 studies are still routinely cited. Science, 371 (2021), pp. 331-332; E.M. Bucci. On zombie papers. Cell Death Dis, 10 (2019), p. 189; S.B. Nissen, T. Magidson, K. Gross, C.T. Bergstrom. Publication bias and the canonization of false facts. eLife, 5 (2016), Article e21451



- **One thing we can do is change the way we write papers.**
- Currently, papers are written and published in a way that results in **errors** and the inability to **computationally reproduce** results.

What Can We Do?

- **Errors**: a 2016 paper by Nuijten et al.[13] found that
  - nearly **half** of all papers had errors in them;
  - over **10%** of p-values in published papers were inconsistent with the reported details of the statistical test
  - 1.6% were what they called "grossly" inconsistent, e.g. difference between the p-value and the test statistic meant that one implied statistical significance and the other did not

What Can We Do?

- **Computational reproducibility**: a 2021 paper by Hardwicke et al.[14] attempted to reproduce results from 25 published papers that publicly shared their data and code:
  - found substantial numerical discrepancies between reported statistical values and values obtained from reproduction attempts in **64%** of these papers

[13] Nuijten, Michèle B, Chris HJ Hartgerink, Marcel ALM van Assen, Sacha Epskamp, and Jelte M Wicherts. 2016. "The Prevalence of Statistical Reporting Errors in Psychology (1985–2013)." Behavior Research Methods 48 (4). Springer: 1205–26.

[14] T.E. Hardwicke, M. Bohn, K. MacDonald, E. Hembacher, M.B. Nuijten, B.N. Peloquin, et al. Analytic reproducibility in articles receiving open data badges at the journal Psychological Science: An observational study. R Soc Open Sci, 8 (2021), Article 201494

What Can We Do?

This is where **Reproducible Papers** come in...

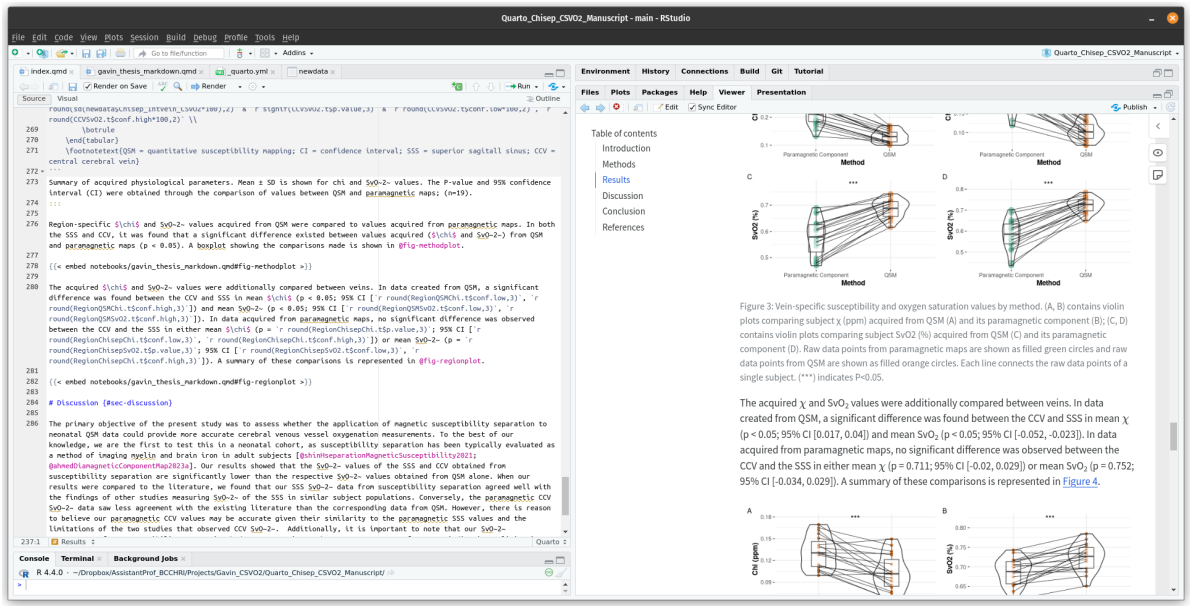


Figure 3: Vein-specific susceptibility and oxygen saturation values by method. (A, B) contains violin plots comparing subject χ (ppm) acquired from QSM (A) and its paramagnetic component (B); (C, D) contains violin plots comparing subject SvO2 (%) acquired from QSM (C) and its paramagnetic component (D). Raw data points from paramagnetic maps are shown as filled green circles and raw data points from QSM are shown as filled orange circles. Each line connects the raw data points of a single subject. (***) indicates P<0.05.

The acquired χ and SvO_2 values were additionally compared between veins. In data created from QSM, a significant difference was found between the CCV and SSS in mean χ (p < 0.05; 95% CI [0.017, 0.04]) and mean SvO_2 (p < 0.05; 95% CI [-0.052, -0.023]). In data acquired from paramagnetic maps, no significant difference was observed between the CCV and the SSS in either mean χ (p = 0.711; 95% CI [-0.02, 0.029]) or mean SvO_2 (p = 0.752; 95% CI [-0.034, 0.029]). A summary of these comparisons is represented in Figure 4.

Learning Goals

Learning Goals

By the end of the talk, the audience should:

- Know what a reproducible manuscript is,
- Understand some reasons why scientists should be writing their manuscripts this way,
- Know what Markdown, Knitr, Pandoc, LaTeX, Jupyter Notebook, R/RMarkdown, and Quarto are,
- Know the basics of the syntax for Markdown, R and Quarto,
- See how to integrate author information, code, equations, tables, images, and citations
- Be able to start writing your next manuscript using Quarto Manuscripts

Introduction

What is a reproducible manuscript?

- Reports the scientific findings

- Provides all (or almost all) the necessary data, code, and methodologies required to create those findings (i.e. data, stats, figures, tables, etc.)
- Transparent and organized
- Enables others to **replicate** and **verify** the results of your study independently

What does it look like?

```
---
title: "The Application of Magnetic Susceptibility Separation for Measuring Cerebral Oxygenat
titlerunning: "CSVO2"
author:
  - name: Thomas Gavin Carmichael
    orcid: 0009-0008-6849-5333
    corresponding: false
    email: tgcarmichael@outlook.com
    roles:
      - writing - original draft
      - formal analysis
      - methodology
      - validation
      - visualization
    affiliations:
      - ref: 1
      - ref: 2
    degrees:
      - HBSc
  - name: Alexander Rauscher
    orcid: 0000-0002-1961-8252
    email: rauscher@physics.ubc.ca
    corresponding: false
    roles:
      - writing - review & editing
    affiliations:
      - ref: 3
    degrees:
      - PhD
      - MSc
  - name: Ruth E Grunau
    orcid: 0000-0002-5428-9212
    corresponding: false
    email: rgrunau@mail.ubc.ca
    roles:
      - writing - review & editing
      - funding acquisition
    affiliations:
      - ref: 2
      - ref: 3
```

```
  - name: Alexander Mark Weber
    orcid: 0000-0001-7295-0775
    corresponding: true
    email: aweber@bcchr.ca
    roles:
      - project administration
      - supervision
      - validation
      - visualization
      - resources
      - methodology
      - formal analysis
      - funding acquisition
      - writing - review & editing
      - conceptualization
      - data curation
      - investigation
    affiliations:
      - ref: 2
      - ref: 3
    degrees:
      - PhD
      - MSc
affiliations:
  - id: 1
    name: The University of British Columbia
    department: Integrated Sciences
    address: 2329 West Mall
    city: Vancouver
    region: BC
    country: Canada
    postal-code: V6T 1Z4
  - id: 3
    name: The University of British Columbia
    department: Pediatrics
    address: 2329 West Mall
    city: Vancouver
    region: BC
    country: Canada
    postal-code: V6T 1Z4
  - id: 2
    department: BC Children's Hospital Research Institute
```

```
    name: The University of British Columbia
    address: 938 West 28th Avenue
    city: Vancouver
    state: BC
    country: Canada
    postal-code: V5Z 4H4

keywords:
  - Quantitative Susceptbility Mapping
  - Preterm
  - Newborn
  - Cerebral Venous Oxygen Saturation
abstract: |
  **Background**: Quantitative susceptibility mapping (QSM) is a magnetic resonance imaging

  **Methods**: 19 neonates born preterm were scanned on a 3T research MRI at term equivalent

  **Results**: The mean SvO~2~ values of the SSS and CCV calculated from QSM images were fou

  **Conclusion**: SSS SvO~2~ values derived from paramagnetic components agreed well with th

plain-language-summary: |

key-points:
  -
date: last-modified
bibliography: [ Gavin_Thesis_Ref.bib ]
citation:
  container-title: Unpublished
number-sections: false
notebook-links: true
---

```{r setup, include=FALSE}
# options
knitr::opts_chunk$set(
  # fig.width=8, fig.height=5,
  # out.width="50%",
  # fig.align="center",
  echo=FALSE,
  message=FALSE,
  warning=FALSE
```
```

```
  # cache=TRUE
  )
set.seed(1234) # reproducible
options(knitr.kable.NA = '') # how kable handles NA
options(reticulate.repl.quiet = TRUE)
```

```{r libraries}
#libraries
library(tidyverse) # ggplot2, dplyr, tidyr, readr, purrr, tibble, stringr, forcats
theme_set( theme_minimal() ) # ggplot theme
library(broom) # for nice summaries
library(knitr) #
library(kableExtra) # more tables options. Can cause problems
library(Rmpfr)
library(gt)
# library(reticulate) # incorporate Python
# use_virtualenv('./pyvenv_csvo', required = TRUE) # load pythong venv from path

# library(neurobase)
```

```{r}
load("notebooks/results.RData")
```

```{r}
# function to make rounding means and sd easier
rndmean <- function(clm) {
  return(round(mean(clm),2))
}
rndsd <- function(clm) {
  return(round(sd(clm),2))
}
```

Introduction {#sec-intro}

<!-- should be around 5-6 paragraphs. Aim for 460 words -->

Preterm birth

````
Abnormal brain development is a significant concern for parents with children born preterm, a

In the present study, we set out to determine whether a QSM image alone, or the paramagnetic


# Methods {#sec-data-methods}

The study was approved by the Clinical Research Ethics Board at the University of British Col

## Study population

Participant data comes from a previous study [@zhu-etal-cmro2]. Participants consisted of pre

## Image acquisition

MR imaging was performed on a 3.0 Tesla General Electric Discovery MR750 scanner (scanner so

```{r}
#| label: tbl-mri
#| tbl-cap: "Technical parameters for MR imaging pulse sequences"
df <- data.frame(Scan = character(), T1w = character(), T2w = character(), pcASL = character
df[1,] <- c("Sequence", "3D FSPGR", "3D CUBE", "Multi-shot 3D fast spin-echo", "3D spoiled GR
df[2,] <- c("Phase-encoding direction", "Coronal", "Sagittal", "Axial", "Axial")
df[3,] <- c("TR (ms)", "7.74", "2,300", "4,680", "30.9")
df[4,] <- c("TE (ms)", "2.97", "66.29", "10.55", "5.24")
df[5,] <- c("Flip angle", "12\U00B0", "90\U00B0", "111\U00B0", "20\U00B0")
df[6,] <- c("FOV (cm)", "20", "20", "24", "25")
df[7,] <- c("Acquisition matrix", "512 x 512", "256 x 256", "128 x 128", "256 x 256")
df[8,] <- c("In-plane resolution (mm)", "0.39 x 0.39", "0.78 x 0.78", "1.875 x 1.875", "0.97
df[9,] <- c("Slice thickness (mm)", "1", "1", "4", "2, reconstructed to 1 with zero filling
df[10,] <- c("Number of slices", "126", "106", "50", "92")
df[11,] <- c("Additional parameters", "n/a", "n/a", "1,450 ms label period;\n 2,025 ms pulse
df[12,] <- c("Scan duration", "4 min 39 s", "5 min 1 s", "5 min 26 s", "5 min 29 s")

footnotetext="T1w = T1-weighted; T2w = T2-weighted; pcASL = pseudo-continuous arterial spin l

if (knitr::is_latex_output()) {
colnames(df)[1] <- ""
df[11,] <- linebreak(df[11,])
df |>
  kbl(format = "latex",
    booktabs = TRUE,
````
````

```
    longtable = TRUE,
    linesep = "",
    align = "l",
    escape = FALSE) |>
    kable_styling(font_size = 8, position = "center", latex_options = c("hold_position", "sca
    footnote(general_title = "",
           footnote_as_chunk = TRUE,
           threeparttable = TRUE,
           general = footnotetext) |>
    column_spec(1, width="8em") |>
    column_spec(4, width="9em") |>
    column_spec(5, width="9em")
} else {
df |>
    mutate(across(everything(), ~ str_replace_all(., "\n", "<br>"))) |>
    gt(rowname_col = "Scan") |> tab_footnote(footnotetext) |>
    fmt_markdown(columns = TRUE) |>
    tab_options(quarto.disable_processing = TRUE)
}
```

The MRI scan protocol comprised of the following sequences: a T1-weighted scan, a T2-weighted

Image analysis

The raw DICOM files acquired from the scanning procedure were converted to NIfTI (Neuroimagi

{{< embed notebooks/Figures.ipynb#fig-graph >}}

First, the fifth echo SWI magnitude file was processed using FSL's (v. 6.0.7.3) [@woolrichBay

STI Suite (v. 3.0) [@liIntegratedLaplacianBased2014], was used to process the final QSM image

To isolate the paramagnetic component of subjects' QSM data, the χ-separation toolbox [@

{{< embed notebooks/Figures.ipynb#fig-sample >}}

Once the mean susceptibility values of the SSS and CCV were obtained from the subjects' QSM i

$$
SvO_{2} = 1 - \frac{\Delta \chi _{blood} - (\Delta \chi _{oxy} * Hct)}{\Delta \chi _{do} * He
$$ {#eq-svo}

```
where $\Delta \chi _{blood}$ is the vessel's measured susceptibility, $\Delta \chi _{oxy}$ is

## Statistical analysis

Statistical analysis of the acquired data was performed using R and RStudio (v. 2023.09.1 Bu

<!-- the vessel-specific SvO2 values determined through QSM and those determined from the pa

# Results {#sec-results}

A total sample size of `r length(newdata$Subject)` infants were scanned, with a mean ($\pm$ s

```{r}
#| label: tbl-dem
#| tbl-cap: Demographic and clinical characteristic of the study sample.

df <- data.frame(Variable = character(), "Subject" = character(), stringsAsFactors = FALSE)
df[1,] <- c("Gestational age, weeks (mean \U00B1 SD)", paste0(rndmean(newdata$GA), " \U00B1 "
df[2,] <- c("Corrected gestational age on scan day, weeks (mean \U00B1 SD) ", paste0(rndmean
df[3,] <- c("Number of male neonates (\\%)", paste0(sum(newdata$Sex == "M"), " (", (round(sum
df[4,] <- c("Birth weight, g (mean \U00B1 SD)", paste0(rndmean(newdata$BW), " \U00B1 ", rndsd
df[5,] <- c("Weight on scan day, g (mean \U00B1 SD)", paste0(rndmean(newdata$Weight.on.Scan.D
df[6,] <- c("Days spent in NICU (median, IQR)", paste0(median(newdata$Total_Days_NICU), ", "
df[7,] <- c("Days on ventilation (median, IQR)", paste0(median(newdata$Total_Days_Ventilation
df <- df |> rename("Subject data (n = 19)" = Subject)
footnotetext="SD = standard deviation; IQR = inter quartile range"

if (knitr::is_latex_output()) {
df |>
  kbl(format = "latex",
    booktabs = TRUE,
    longtable = TRUE,
    linesep = "",
    align = "lc",
    escape = FALSE) |>
    kable_styling(font_size = 9, position = "center", latex_options = c("hold_position", "sca
    footnote(general_title = "",
           footnote_as_chunk = TRUE,
           threeparttable = TRUE,
           general = footnotetext)
} else {
df |>
```
```

```
    mutate(across(everything(), ~ str_replace_all(., "\n", "<br>"))) |>
    gt() |>
    cols_align(align = "center", columns = c("Subject data (n = 19)")) |>
    tab_footnote(footnotetext) |>
    fmt_markdown(columns = TRUE) |>
    tab_options(quarto.disable_processing = TRUE)
}
```

The mean SvO~2~ values for the SSS and the CCV were found to be `r rndmean(newdata$Gavin_SSS

```{r}
#| label: tbl-chistats
#| tbl-cap: Summary of acquired physiological parameters. Mean $\pm$ SD is shown for chi and

df <- data.frame(Region = character(), Measure = character(), QSM = character(), pmap = chara
df[1,] <- c("SSS", "Chi (ppm)", paste0(rndmean(newdata$Gavin_SSS_Chi), " \U00B1 ", rndsd(newd
df[2,] <- c("SSS", "SvO\U2082 (\\%)", paste0(rndmean(newdata$Gavin_SSSVein_CSvO2*100), " \U00
df[3,] <- c("CCV", "Chi (ppm)", paste0(rndmean(newdata$Gavin_IntVein_Chi), " \U00B1 ", rndsd
df[4,] <- c("CCV", "SvO\U2082 (\\%)", paste0(rndmean(newdata$Gavin_IntVein_CSvO2*100), " \U00
footnotetext="QSM = quantitative susceptibility mapping; CI = confidence interval; SSS = supe

if (knitr::is_latex_output()) {
df <- df |> rename("Paramagnetic map" = pmap, "p-value" = pvalue, "95\\\U0025 CI" = CI)
df |>
  kbl(format = "latex",
    booktabs = TRUE,
    longtable = TRUE,
    linesep = "",
    align = "llcccc",
    escape = FALSE) |>
    kable_styling(font_size = 9, position = "center", latex_options = c("hold_position", "sca
    footnote(general_title = "",
           footnote_as_chunk = TRUE,
           threeparttable = TRUE,
           general = footnotetext)
} else {
df <- df |> rename("Paramagnetic map" = pmap, "p-value" = pvalue, "95\U0025 CI" = CI)
df |>
    mutate(across(everything(), ~ str_replace_all(., "\n", "<br>"))) |>
    gt() |>
    cols_align(align = "center", columns = c("QSM", "Paramagnetic map", "p-value", "95\U0025
```

```
    tab_footnote(footnotetext) |>
    fmt_markdown(columns = TRUE) |>
    tab_options(quarto.disable_processing = TRUE)
}
```

Region-specific χ and SvO~2~ values acquired from QSM were compared to values acquired

{{< embed notebooks/gavin_thesis_markdown.qmd#fig-methodplot >}}

The acquired χ and SvO~2~ values were additionally compared between veins. In data creat

{{< embed notebooks/gavin_thesis_markdown.qmd#fig-regionplot >}}

Discussion {#sec-discussion}

The primary objective of the present study was to assess whether the application of magnetic

Conclusion {#sec-conclusion}

References {.unnumbered}

::: {#refs}
:::

What are some other benefits?

- Already mentioned:
  - reducing **errors** from copy-pasting results to paper
  - anyone can see how I obtained my results or figures by **reviewing my code** (bonus: learn how others made their figures!)
- Easy to restructure, rewrite, revise:
  - no need to tweak reported values, tables, or figures by hand
  - remove barrier to re-running analyses (thanks to Reviewer #2); speed up resubmission

What are some other benefits?

- easy cross-referencing and citations