My personal journey into preregistration

Sjoerd Bruijn

Department of Human Movement Sciences.

@sjoerdmb

VU VRIJE UNIVERSITEIT AMSTERDAM | Department of Human Movement Sciences

How it started….   How its going…

Zrinka Potocanac @ZrinkaPot · 11 aug. 2017
Anyone with experience with registered reports in the field of posture and gait? @ISPGR @sjoerdmb @PieterMeyns @woutersinas @KimvanSchooten

6

Sjoerd Bruijn @sjoerdmb · 11 aug. 2017
I'm in doubt about this. Many cases: yes sure. But what's wrong with being honest that it was exploratory, when not preregistering? 1/2

3

Sjoerd Bruijn @sjoerdmb · 11 aug. 2017
Als antwoord op @sjoerdmb @ZrinkaPot en 4 anderen
Also, I feel that in most cases I do have these documents already, just not registered. Is that "bad"? I'm in doubt. 2/3

Sjoerd Bruijn @sjoerdmb · 11 aug. 2017
Als antwoord op @sjoerdmb @ZrinkaPot en 4 anderen
Just be honest about what's exploratory and planned seems best. Preregistering is a sad necessity because of possible dishonesty.

3   1

☆ Karvovskaya, E.   15 november 2020 om 21:56
pre-registration and VU Data Conversation
Aan: Bruijn, S.M.

Siri heeft nieuwe contactinfo gevonden in deze e-mail: E. Karvovskaya e.karvovskaya@vu.nl   zet in Contacten...

Dear Sjoerd,

My name is Lena Karvovskaya; I work as a community manager for research data management at the VU Amsterdam library.
Among other things, I organize Data Conversations, monthly meetings about open science, research transparency, and reproducibility.
One of the topics I would like to highlight at the Data Conversations is pre-registration practices in different disciplines.
Jaap van Dieën recommended that I contact you because I am looking for a speaker who could give a short talk about pre-registration.
I was wondering if you might be willing to give a short presentation about your experience with pre-registration. It would be fantastic if you could give a 10-15 talk in one of the upcoming Data Conversations meetings (January or later).
I usually organize the session in the following way: two short presentations, about 10-15 minutes each; a general Q&A session, and then a discussion in smaller groups in breakout rooms (for those who want to feel themselves part of the "broader university" and meet colleagues from other faculties and disciplines).
Please let me know if you would be interested in participating - I'd be happy to discuss the format in more detail. Thank you!
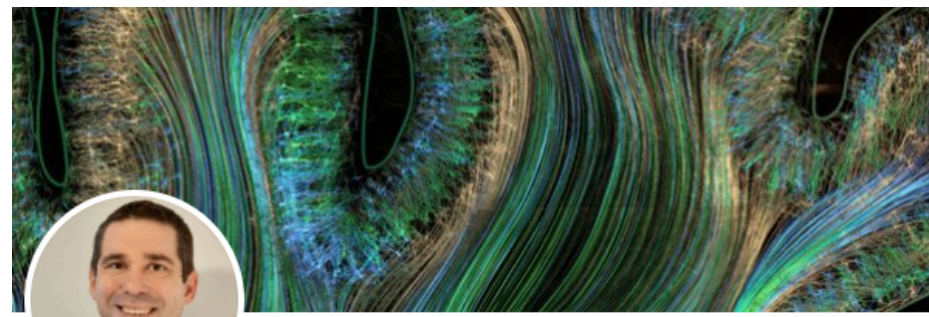Best wishes,
Lena

My goals with this presentation
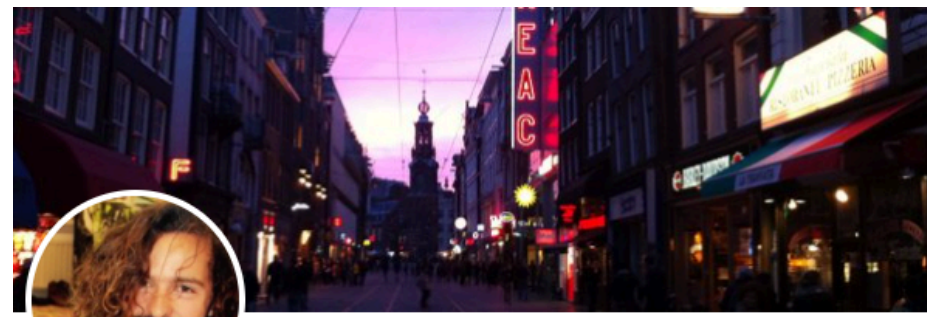
Tell you a bit about WHAT preregistration is

Tell you a bit about WHY you would preregister

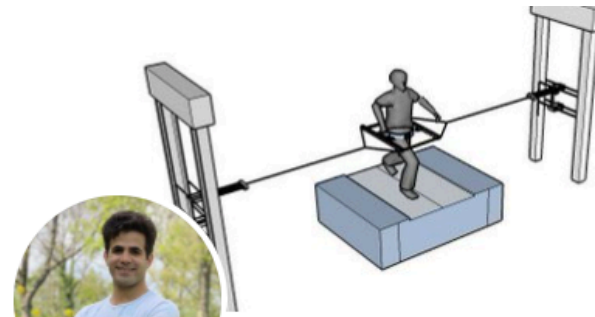Tell you my experiences in preregistration

I need to thank many people for this journey.

Matthieu Boisgontier
@MattBoisgontier Volgt jou
Neuroscience, Physical Activity, Open Science • Assistant Professor @uOttawa •
NBA fan • matthieuboisgontier.com

Nick Kluft
@Nick_Kluft Volgt jou
PhD candidate: Human Movement Sciences, at VU Amsterdam
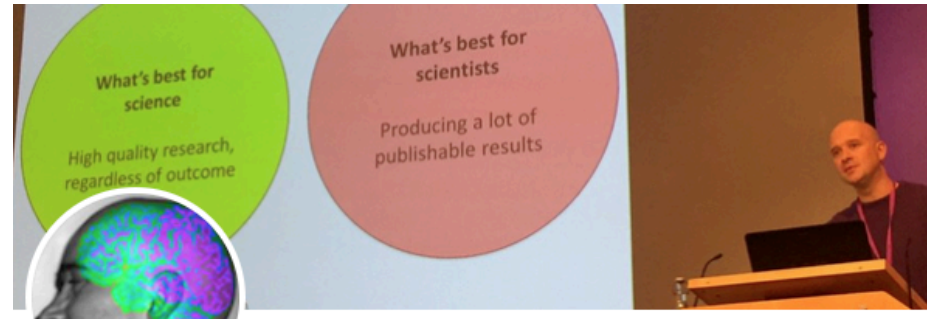📍 Amsterdam, Nederland 🔗 seniorenonderzoek.com

Mohammadreza Mahaki
@mohammad_mahaki Volgt jou
🧍 PhD student of Human Movement Sciences at @VU_FBW | 🥋 Karate Instructor
📍 Amsterdam, Nederland 🔗 Instagram.com/mohammadreza_m...

Moira
@LeeuwenMoira Volgt jou
PhD student Neuromechanics | Streetballer | Author of my imagination
📅 Lid geworden in augustus 2018

Jaap van Dieen
@DieenJaap Volgt jou
📍 Amsterdam, The Netherlands 🎂 Geboren op 15 september
📅 Lid geworden in oktober 2017

Chris Chambers ✔
@chrisdc77
Cognitive neuroscientist, Cardiff Uni. Australian by nature. Dad to three critters.
Former Guardian band member theguardian.com/science/head-q...

Brian Nosek
@BrianNosek
Executive Director @ Center for Open Science, Prof @ University of Virginia, co-
Founder of Project Implicit & the Society for Improving Psychological Science

Zrinka Potocanac
@ZrinkaPot Volgt jou
Human Movement Scientist with a background in Computer Science
Bio vertalen

And many others, who are not on twitter…

Disclaimer
Aka; I am not an expert

I'm just on twitter a whole lot

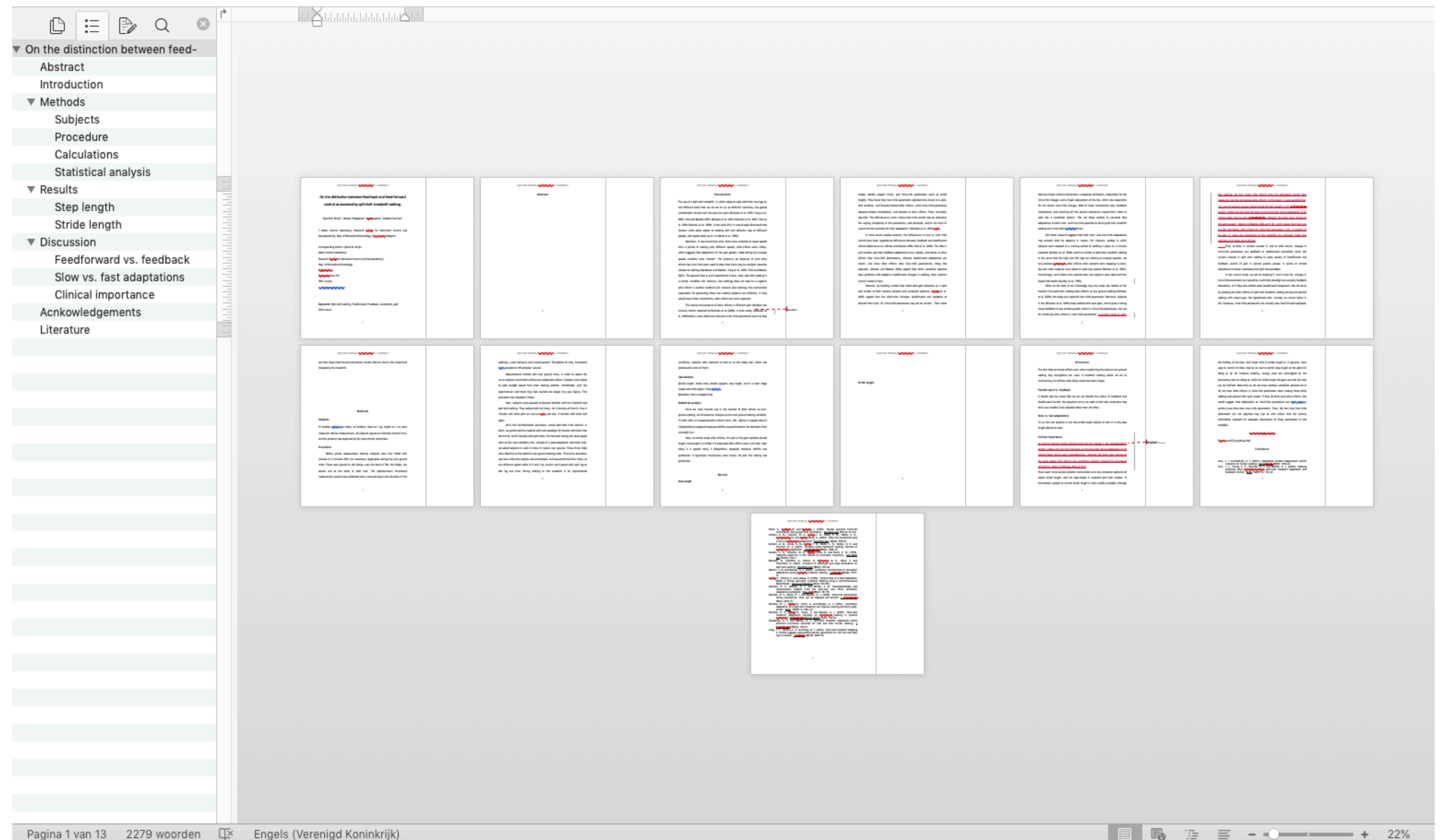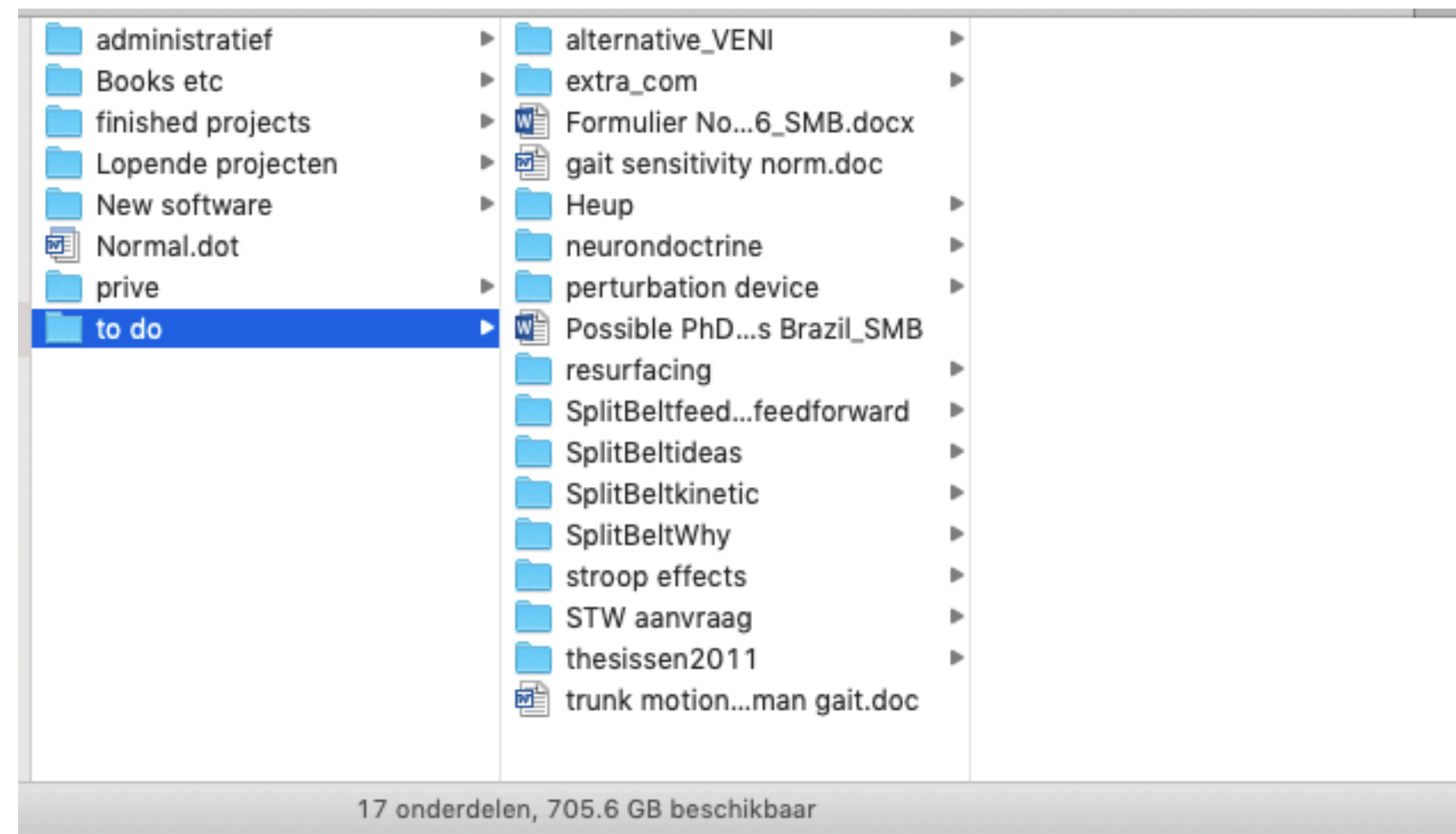And am trying to make my science better.

Along the way, I met the right PhD students, who wanted the same thing.

And worked with the right professors, who were open minded, and encouraging.

Normally, my slides contain no text, and only figures. I couldn't figure out(pun intended) how to do that for this presentation. Sorry!

What is preregistration?

My old way of doing

What you see here is a screenshot of a paper for which no data was collected at all yet (not even now… 9 years later… :-s)

preregistration

- Simply specifying up front what you will do, and how, and storing that document in a locked (time-stamped) format.

- You can ALWAYS do this, no matter which journal

Confirmatory Research
- Hypothesis testing
- Results are held to the highest standards
- Data-independent
- Minimizes false positives
- P-values retain diagnostic value
- Inferences may be drawn to wider population

Exploratory Research
- Hypothesis generating
- Results deserve to be replicated and confirmed
- Data-dependent
- Minimizes false negatives in order to find unexpected discoveries
- P-values lose diagnostic value
- Not useful for making inferences to any wider population

Preregistration allows the researcher to make a clear distinction between both modes of research.

When Can You Preregister?
- Right before your next round of data collection
- After you are asked to collect more data in peer review
- Before you begin analysis of an existing data set

Why Preregister?
- Makes your science better by increasing the credibility of your results
- Allows you to stake your claim to your ideas earlier
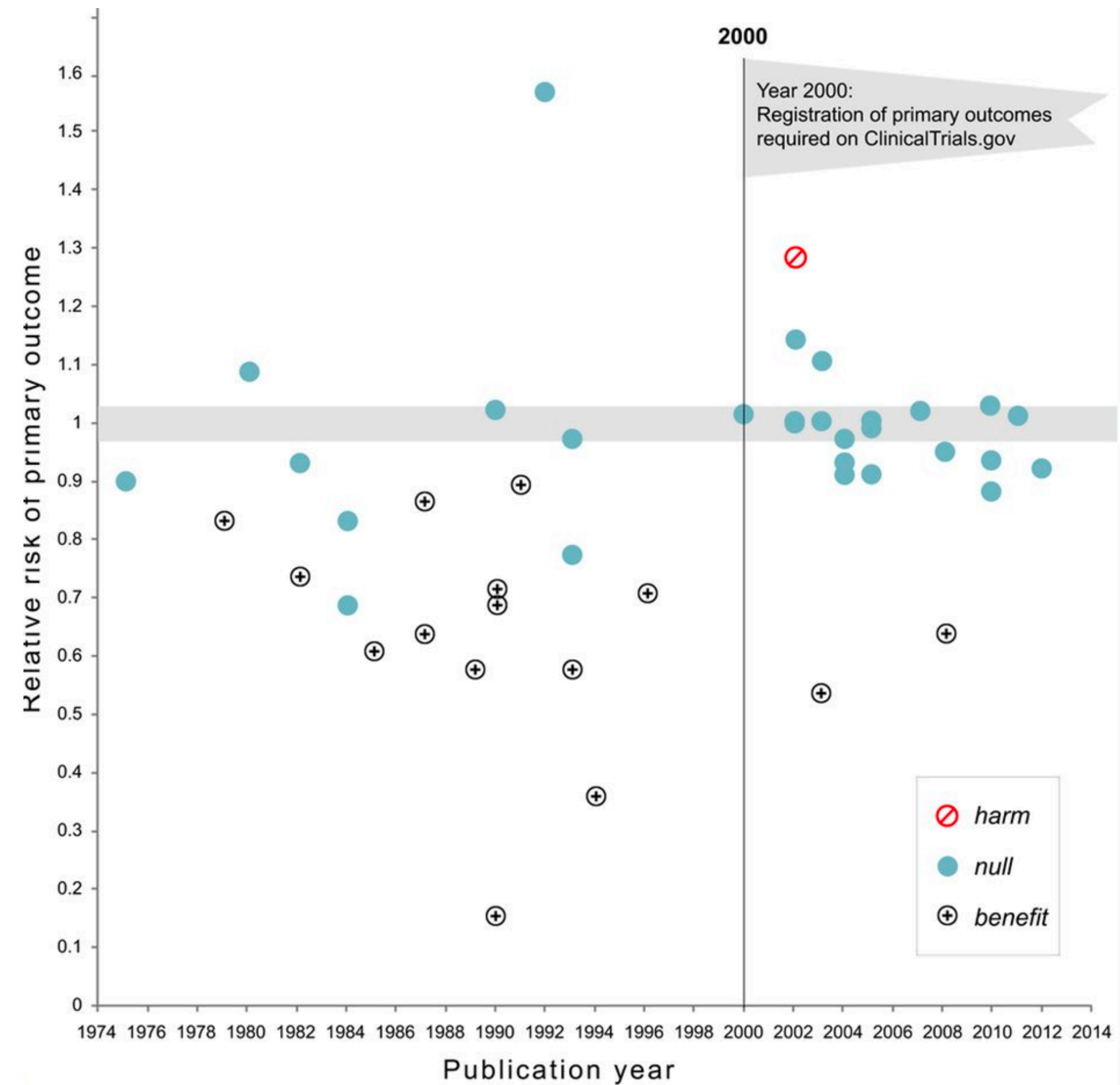- It's an easy way to plan for better research

https://osf.io/prereg/

Preregistration a small step?

https://www.aspredicted.org/create.php

Why preregistration?

Negative results

- This obviously doesn't help patients

- Nor does it help us (how many times have you based your results on findings that were most likely bogus?)

- Nor does it help tax payer money to get what it's worth

- But still, journals seem to prefer 'positive' findings.

- How to solve this?

    - A willingness to see 'negative' results as equally important as positive results

    - Preregistration

    - Registered reports

    - (Other stats (ways to prove H0 is more likely, e.g. equivalence testing and or bayesian stats))

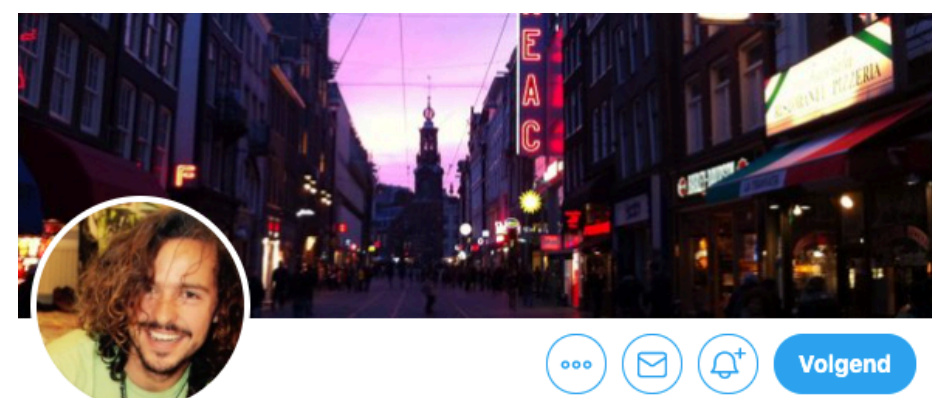# My experiences

My experiences with:

Authoring

Reviewers

Editors

The author
Those "pesky" sample sizes.

How many observations will be collected or what will determine sample size? No need to justify decision, but be precise about exactly how the number will be determined.
The sample size is set using a similar approach as described in Wagenmakers (2007); which comprehends the monitoring of the Bayes factor during data collection, until a threshold of meaningful evidence has been reached. We set this threshold to a Bayes factor (BF10, or BF01) of 10, as generally this is considered as being strong evidence in favour of either hypothesis. First, a number of twenty healthy older adults (age 65-85 years) will be included, and this sample will be extended until the BF exceeds the selected threshold, or a maximum of fifty participants has been reached.

Nick Kluft Volgt jou
@Nick_Kluft Volgt jou
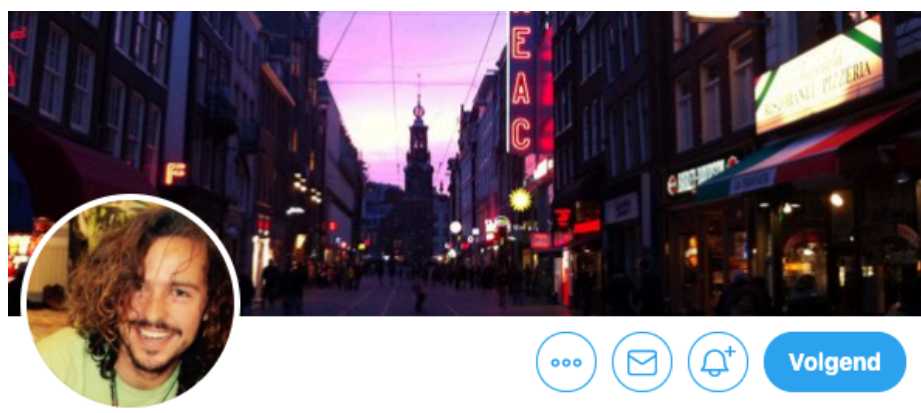PhD candidate: Human Movement Sciences, at VU Amsterdam
Amsterdam, Nederland · seniorenonderzoek.com

The "forgetful" author
Aka "deviations from preregistration"

Deviations from the preregistration document

Three aspects that deviate from the registered document should be noted. First, we planned to determine the slope of the psychometric curve at hcrit to reflect the consistency of the strategy selection. In contrast to earlier studies[5,40], we found an overall lower critical height in the present study. This led to a shift of the psychometric curve towards zero, which makes the slope of the curve less reliable as there were fewer data available to fit the lower end of the curve (due to the inability to evaluate stepping down at negative step heights). Hence, we omitted the consistency of strategy selection from further analysis. Second, for the handling of missing data, visual evaluation of the interpolated data demonstrated that the resulting trajectories were adequate, and there was no need to continue fitting linked-segment models, as suggested in the preregistered document. Third, for the between participant comparison, we planned to categorise fear group on the basis of the physiological arousal data. However, the grouping cutoffs appeared arbitrary. Instead we performed a linear regression, since no cutoffs are needed in a linear regression model and this analysis is analogous to the planned analysis.

The "nasty" reviewer 2

1.There is an over-emphasis on "hypothesis" testing throughout the manuscript that is unwarranted and diminishes the credibility of the manuscript. This work is primarily exploratory in nature. This is perfectly fine, but the manuscript presents all of this work as being "hypothsesis" driven, which it mostly was not. For example, Lines 91-100: For "hypothesis" (2), the proposed "higher aforementioned correlations in walking" cannot be predicted a priori. Likewise, "hypotheses" (4) and (5) are purely conjecture. This comes across as "HARK-ing" (the unscientific process of contriving hypotheses after the results are known). This unwarranted over-emphasis on "hypotheses" continues throughout the Results and Discussion sections in particular. As the dependent measures addressed here have not yet been assessed in running, an exploratory study to determine how running is similar/different from walking is perfectly legitimate, but the manuscript must be written to present the work as such.

Response: We appreciate the reviewer's concern about the credibility of our submitted manuscript. One suggested method for deterring HARKing, hypothesizing after the results are known, is pre-registration of the research proposal. We did not officially preregister our proposal, but did put the project proposal on an OSF page (*https://osf.io/mvkex/.*) which has a timestamp. From this, it can be appreciated that our hypotheses were formulated before the data (which all have a date AFTER the said timestamp) were collected. Thus, we do not agree with the reviewer here that we are HARKing. To make this clear to the reader also, we have now put a short text in the introduction, which reads; *"our initial research proposal for this project can be found at* https://osf.io/mvkex/.*"*

Mohammadreza Mahaki
@mohammad_mahaki Volgt jou
👤 PhD student of Human Movement Sciences at @VU_FBW | 🥋 Karate Instructor
📍 Amsterdam, Nederland 🔗 Instagram.com/mohammadreza_m...

The "cool" reviewer
or; sharing data ad code as well.

3. The authors made all data and analysis scripts available, which is great. Looking over what's in this cloud drive, however, I noticed some weird issues. One issue is that some of the variables have very large jumps, as seen in software/Plots, for example the Right Arm Swing in Subject 1, Trial 1 or Subject 14, Trial 5. This might be a problem with calculating angles from the rigid body orientation given by three markers on the cluster, since the jumps seem to be roughly around 90deg. Similar jumps occur in ML Pelvis Displacement, though, e.g. Subject 1, Trial 6; Subject 5, Trial 9; Subject 9, Trial 4. Another issue is gaps in the data, where some of the trajectories will just disappear for some of the gait cycle, e.g. in Right Arm Swing in Subject 10, Trial 9 or Transverse Pelvis Rotation Subject 1, Trial 4, where around 60% of the gait cycle *all* data is missing, similarly for Subject 6, Trial 5 around 0-20%. I did not go through the analysis code in detail, so it is possible that these are just intermediate results, before such issues have been weeded out by the authors, although the readme.docx seems to suggest that this is not the case. If this is the case and these artifacts are still part of the data as analyzed in the manuscript, then I suggest that the authors go back to the data processing stage and take a close and careful look at where they come from and how to avoid them. In some cases, removing a small number of problematic gait cycles might be sufficient, but in other cases, all data seems to be missing for part of the gait cycle, and I don't know of a good way to deal with this.

We thank the reviewer for spotting these errors in our data analysis. This is one of the reasons why we also share the data (and code); to make sure that (due to some unforeseen circumstances) we don't end up publishing rubbish. So, we are really happy that you spotted this mistake. Indeed, some of the data was quite noisy, part of which was caused by malfunctioning of the equipment (there were renovations on the floor where the lab is located, and only after these, we discovered that dust on our Optotrak lenses may have caused us quite some problems

Mohammadreza Mahaki
@mohammad_mahaki Volgt jou
🧍 PhD student of Human Movement Sciences at @VU_FBW | 🥋 Karate Instructor
📍 Amsterdam, Nederland 🔗 Instagram.com/mohammadreza_m...

The appreciative editor

Dear Dr. Bruijn,

I am pleased to inform you that your manuscript has been judged scientifically suitable for publication and will be formally accepted for publication once it meets all outstanding technical requirements (formatting, etc.). These requirements (if any) are determined by the production office and are separate from the content review provided by the academic editorial process.

Personally, I am happy to congratulate you on a very methodical and thorough study, and commend your group for pre-registration. Very nice work.

Moira
@LeeuwenMoira Volgt jou
PhD student Neuromechanics | Streetballer | Author of my imagination
Lid geworden in augustus 2018

Takeaways;

Should you preregister?

Obviously; yes

But, I would go further
And also share data and code whenever possible.

This costs very little time (they should be organized well anyway), and can greatly benefit you and others.

If you want to know more, a good place to start is :

https://www.cos.io/initiatives/prereg

@sjoerdmb   <u>**s.m.bruijn@vu.nl**</u>