From the editor

We also ask that you add an explanation in your abstract that the article relates specifically to neuroimaging research.

Thank you for the suggestion. We have made the abstract more explicit.

Reviewer #1 (Niko Kriegeskorte)

This is an open review, available here:
http://nikokriegeskorte.org/2016/02/15/the-four-pillars-of-open-science/

NOTE: Because this review is so extensive, we have tried to distill the main points rather than pasting in the entire review.

(...)

Open notebook science: Open science is about enhancing the bandwidth and reducing the latency in our communication network. This means sharing more and at earlier stages, not only our data and code, but ultimately also our day-to-day incremental progress. This is called open notebook science and has been explored, by Cameron Neylon and Michael Nielson among others. Gorgolewski & Poldrack don't comment on this beautiful vision for an entirely different workflow and culture at all. Perhaps open notebook science is too far in the future? However, some are already practicing it. Surely, we should start exploring it in theory and considering what aspects of open notebook science we can integrate into our workflow. It would be great to have some pointers to practices and tools that help us move in this direction.

We thank the reviewer for pointing us to the idea of open notebook science. It is definitely interesting idea, but we are afraid that it is currently too unorthodox for most researchers. Our goal was to provide a practical guide that includes incremental changes to scientists' workflows that are motivated by with arguments for how those changes can benefit them. We stayed away from painting a vision of science in the distant future to avoid being discarded as detached from reality. By proposing small changes we hope to appeal to a wider range of readers who did not consider sharing their code, data, or publishing preprints before. In other words, we do not want the perfect to become the enemy of the good.

The scientific paper remains a critical component of scientific communication: Data and code sharing are essential, but will not replace communication through permanently citable scientific papers that link (increasingly accessible) data through analyses to novel insights and relate these insights to the literature.

Papers should simultaneously achieve clarity and transparency: The conceptual clarity of the the argument leading to an insight is often at a tension with the transparency of all the methodological details. Ideally, a paper will achieve both clarity and transparency, providing multiple levels of description: a main narrative that abstracts from the details, more detailed descriptions in the methods section, additional detail in the supplementary information, and full detail in the links to the open data and code, which together enable exact reproduction of the results in the figures. This is an ideal to aspire to. I wonder if any paper in our field has fully achieved it. If there is one, it should surely be cited.

We have certainly aspired to this ideal in some of our previous publications (Poldrack et al., 2012, PLOS Comp Biology; Poldrack et al., 2015, Nature Communications), as have others (Waskom et al, 2014) though we don't think that any of these have fully achieved the vision laid out by the authors.

Open access: Papers need to be openly accessible, so their insights can have the greatest positive impact on science and society. This is really a no brainer. The internet has greatly lowered the cost of publication, but the publishing industry has found ways to charge higher prices through a combination of paywalls and unreasonable open-access charges. I would add that every journal contains unique content, so the publishing industry runs hundreds of thousands of little monopolies - safe from competition. Many funding bodies require that studies they funded be published with open access. We need political initiatives that simply require all publicly funded research to be publicly accessible. In addition, we need publicly funded publication platforms that provide cost-effective alternatives to private publishing companies for editorial boards that run journals. Many journals are currently run by scientists whose salaries are funded by academic institutions and the public, but whose editorial work contributes to the profits of private publishers. In historical retrospect, future generations will marvel at the genius of an industry that managed for decades to employ a community without payment, take the fruits of their labour, and sell them back to that very community at exorbitant prices - or perhaps they will just note the idiocy of that community for playing along with this racket.

Although in general we do agree with the reviewer's diagnosis of the publishing landscape, this topic is somewhat afield from the goal of our paper and we do not feel that its inclusion would strengthen the paper. As mentioned above, our goal was to convince individual researchers that they can benefit from small changes in their workflow that coincidentally make science more open and benefit the whole community, not to argue for revolutionary changes.

(...)

Almost all journals support the posting of preprints: Although this is not widely known in the brain imaging and neuroscience communities, almost all major journals (including Nature, Science, Nature Neuroscience and most others) have preprint policies supportive of posting *preprints. Gorgolewski & Poldrack note that they "are not aware of any neuroscience journals that do not allow authors to deposit preprints before submission, although some journals such as Neuron and Current Biology consider each submission independently and thus one should contact the editor prior to submission." I would add that this reflects the fact that preprints are also advantageous to journals: They help catch errors and get the reception process and citation of the paper going earlier, boosting citations in the two-year window that matters for a journal's impact factor.*

This is an excellent point. We have added it to the paper.

4 Open reviews

The fourth pillar of open science is the open evaluation (OE, i.e. open peer review and rating) of scientific papers. This pillar is entirely overlooked in the present version of the Gorgolewski & Poldrack's commentary. However, peer review is an essential component of communication in science. Peer review is the process by which we prioritise the literature, guiding each field's attention, and steering scientific progress. Like other components of science, peer review is currently compromised by a lack of transparency, by inefficiency of information flow, and by unhealthy incentives. The movement for opening the peer review process is growing.

In traditional peer review, we judge anonymously, making inherently subjective decisions that decide about the publication of our competitors' work, under a cloak of secrecy and without ever having to answer for our judgments. It is easy to see that this does not provide ideal incentives for objectivity and constructive criticism. We've inherited secret peer review from the pre-internet age (when perhaps it made sense). Now we need to overcome this dysfunctional system. However, we've grown used to it and may be somewhat comfortable with it.

Transparent review means (1) that reviews are public communications and (2) that many of them are signed by their authors. Anonymous reviewing must remain an option, to enable scientists to avoid social consequences of negative judgments in certain scenarios. However, if our judgment is sound and constructively communicated, we should be able to stand by it. Just like in other domains, transparency is the antidote to corruption. Self-serving arguments won't fly in open reviewing, and even less so when the review is signed. Signing adds weight to a review. The reviewer's reputation is on the line, creating a strong incentive to be objective, to avoid any impression of self-serving judgment, and to attempt to be on the right side of history in one's judgment of another scientist's work. Signing also enables the reviewer to take credit for the hard work of reviewing.

The arguments for OE and a synopsis of 18 visions for how OE might be implemented are given in Kriegeskorte, Walther & Deca (2012). As for other components of open science, the primary obstacles to more open practices are not technological, but psychological, cultural, and political. Important journals like eLife and those of the PLoS family are experimenting with steps toward *opening the review process. New journals including, the Winnower, ScienceOpen, and F1000 Research already rely on postpublication peer review.*

We don't have to wait for journals to lead us. We have all the tools to reinvent the culture of peer review. The question is whether we can handle the challenges this poses. Here, in the spirit of Gorgolewki & Poldrack's practical guide, are some ways that we can make progress toward OE now by doing things a little differently.

Sign peer reviews you author: Signing our reviews is a major step out of the dark ages of peer review. It's easier said than done. How can we be as critical as we sometimes have to be and stand by our judgment? We can focus first on the strengths of a paper, then communicate all our critical arguments in a constructive manner. Some people feel that we must sign either all or none of our reviews. I think that position is unwise. It discourages beginning to sign and thus de facto cements the status quo. In addition, there are cases where the option to remain anonymous is needed, and as long as this option exists we cannot enforce signing anyway. What we can do is take anonymous comments with a grain of salt and give greater credence to signed reviews. It is better to sign sometimes than never. When I started to sign my reviews, I initially reserved the right to anonymity for myself. After all this was a unilateral act of openness; most of my peers do not sign their reviews.

However, after a while, I decided to sign all of my reviews, including negative ones.

Openly review papers that have preprints: When we read important papers as preprints, let's consider reviewing them openly. This can simultaneously serve our own and our collective thought process: an open notebook distilling the meaning of a paper, why its claims might or might not be reliable, how it relates to the literature, and what future steps it suggests. I use a blog. Alternatively or additionally, we can use PubMed Commons or PubPeer.

Make the reviews you write for journals open: When we are invited to do a review, we can check if the paper has been posted as a preprint. If not, we can contact the authors, asking them to consider posting. At the time of initial submission, the benefits tend to outweigh the risks of posting, so many authors will be open to this. Preprint posting is essential to open review. If a preprint is available, we can openly review it immediately and make the same review available to the journal to contribute to their decision process.

Reinvent peer review: What criteria define the quality of an open review? For example, what is this thing you're reading? A blog post? A peer review? Open notes on the essential points I would like to remember from the paper with my own ideas interwoven? All of the above. Ideally, an open review helps the reviewer, the authors, and the community think - by explaining the meaning of a paper in the context of the literature, judging the reliability of its claims, and suggesting future improvements. As we begin to review openly, we are reinventing peer review and the evaluation of scientific papers.

Invent peer rating: Eventually we will need quantitative measures evaluating papers. These should not be based on buzz and usage statistics, but reflect the careful judgement of peers *who are experts in the field, have considered the paper in detail, and ideally stand by their judgment. Quantitative judgment must be multidimensional and can be used to build a plurality of paper evaluation functions (Kriegeskorte 2012) that prioritise the literature from different perspectives. We need to invent suitable rating systems. For primary research papers, I use single digit ratings on multiple scales including reliability, importance, and novelty, using capital letters to indicate the scale in the following format: [R7I5].*

In principle we agree with many point that the reviewer made about the need for reforming the peer review system, but we also acknowledge many open questions about how things should be done. Our paper aims to be a practical guide that proposes immediately achievable changes to established practices rather than a vision of how science could work. We hope that in couple of years we will be able to give more practical advice on the topic of peer review. Nonetheless we acknowledge that researchers should be encouraged to express their opinions about preprints and peer reviewer manuscript and therefore we added a section describing different outlets for pre and post publication reviews.

> *An important part of the scientific method is peer review but with exception to some journals (eLife, GigaScience, ScienceOpen, and F1000Research) this procedure happens behind closed doors and thus leaves the reader without any information on how a published paper was evaluated (other than it was accepted). In addition reviewers do not get credit for their hard work. This situation can be remedied by publishing reviews performed for journals after the paper has been published. Several outlets exists that allow that. PubMed Commons allow registered and verified users of PubMed to provide comments under every paper indexed by PubMed. Those comments have to be signed so there is no option to remain anonymous (which is important for junior researchers afraid of a blowback after criticizing work from an established lab). Another option is PubPeer - a website that allow anyone to comment on any published paper or preprint. It supports both anonymous and signed comments so it's up to the reviewer to decide what is better for them. Finally there is Publons.com - a platform for tracking reviewers profiles and publishing reviews. Thanks to collaborations with many journals it is very easy to use and even allows you to get credit for publishing your reviews anonymously.*
>
> *All of those platforms can be used not only to share reviews solicited from reviewers by journals, but also to share comments and give feedback about already published work or preprints shared by other researchers. Peer review expanded to the whole community we can improve the quality, catch mistakes or help with the clarity of both preprints and already published work. Giving feedback on preprints can be especially useful when it comes to highlighting already published work that authors might have missed (which considering the number of papers published every year is not unlikely).*
>
> *Signing openly shared reviews can have some benefits when it comes to establishing one's reputation as an expert in the field. Well thought through and carefully worded reviews consisting of constructive criticism are hard to get by and extremely valuable. By sharing and signing reviews researchers can not only help their peers, but also boost their reputation*

> *which can potentially seen favourably by hiring committees and grant review boards. It is important to note that an option of anonymous reviews is very important since on many occasions it’s the only way for researchers to express concerns about validity of some work.*

Errors are normal

As we open our science and share more of it with the community, we run the risk of revealing more of our errors. From an idealistic perspective that's a good thing, enabling us learn more efficiently as individuals and as a community. However, in the current game of high-impact biomedical science there is an implicit pretense that major errors are unlikely. This is the reason why, in the rare case that a major error is revealed despite our lack of transparent practices, the current culture requires that everyone act surprised and the author be humiliated. Open science will teach us to drop these pretenses. We need to learn to own our mistakes (Marder 2015) and to be protective of others when errors are revealed. Opening science is an exciting creative challenge at many levels. It's about reinventing our culture to optimise our collective cognitive process. What could be more important or glamorous?

We could not agree more. We have added the following statement to the summarry to make this point explicit:

> *The scientific method is evolving towards a more transparent and collaborative endeavour. The age of digital communication allows us to go beyond printed summaries and dive deeper into underlying data and code. In this guide we hope to have shown that there are many improvements in scientific practice everyone can implement with relatively little added effort that will improve transparency, replicability and impact of their research. Even though the added transparency might in rare cases expose errors those are a natural part of the scientific process. As a community researchers should acknowledged their existence and try to learn from them instead of hiding them and antagonizing those who make them.*

A major relevant development regarding open science in the brain imaging community is the OHBM's Committee on Best Practices in Data Analysis and Sharing (COBIDAS), of which author Russ Poldrack and I are members. COBIDAS is attempting to define recommended practices for the neuroimaging community and has begun a broad dialogue with the community of researchers (see weblink above). It would be good to explain how COBIDAS fits in with the other developments.

We have added a mention of the report in context of reporting neuroimaging studies.

About a third of the cited papers are by the authors. This illustrates their substantial contribution and expertise in this field. I found all these papers worthy of citation in this context. However, I wonder if other groups that have made important contributions to this field should be more *broadly cited. I haven't followed this literature closely enough to give specific suggestions, but perhaps it's worth considering whether references should be added to important work by others.*

By incorporating feedback from all five reviewers we have diversified the list of citations.

As for the papers, the authors are directly involved in most of the cited web resources OpenfMRI, NeuroVault, NeuroStars.org. This is absolutely wonderful, and it might just be that there is not much else out there. Perhaps readers of this open review can leave pointers in the comments in case they are aware of other relevant resources. I would share these with the authors, so they can consider whether to include them in revision.

We only got some clarifying remarks on the data intake policies of FCP/INDI, but we have added a note about a repository for connectivity matrices.

> *The UCLA Multimodal Connectivity Database (Brown et al. 2012) provides similar service but for connectivity matrices (derived from fMRI or DWI data).*

Can the practical pointers be distilled into a table or figure that summarises the essentials? This would be a useful thing to print out and post next to our screens.

This is a great idea - we have added a box summarizing the main points.

"more than fair" -> "only fair"

Fixed.

Reviewer #2 (Brian Nosek)

This manuscript provides a useful, if too brief, overview of improving transparency and reproducibility of neuroimaging research. Comments that occurred to me as I read the manuscript are below. My main suggestions are to [1] be clearer about the value proposition of each step - how does it translate into something good for science and - ideally - good for the researcher too; [2] address more of the common concerns directly (e.g., how to manage being scooped, what is the value of doing the work to share data and code when others might find error); and [3] add a section on open workflow (a fourth pillar), particularly covering possibilities like preregistration to help deal with the core issues that opened the manuscript. And, while at it, why not give some advice for open materials - what had been delivered to the participants during the experimental protocol.

** The single paragraph in the introduction (third paragraph) may be a bit too telegraphic in giving the reader grounding in the nature of the problem. It would be helpful to revise that so that researchers can grasp the combination of [1] power of designs, publishing incentives, lure of positive results with [2] with enormous flexibility in analysis and reporting pipelines that can produce a literature that is much less reliable than desired. A full review not needed (can cite others such as Button et al., 2013 Nature Reviews Neuroscience for power; Ioannidis et al., 2014, TICS for the problems generally; etc.).*

We have extended the introduction and added the suggested citations:

> Having access to a plethora of denoising and modelling algorithms can be both good and bad. On one side there are many aspects of brain anatomy and function that we can extract and use as dependent variables which maximizes the chances of finding the right fit for a particular question. On the other side the passion driving brain researchers combined with methodological plurality can be a dangerous mix. Scientists rarely approach a problem without a theory, hypothesis, or a set of assumptions and those can implicitly drive them to implicitly choose results of some analyses over others leading to increased “researcher degrees of freedom” (Simmons et al. 2011). As Richard Feynman said “The first principle is that you must not fool yourself — and you are the easiest person to fool.”. Additionally neuroimaging like almost every other scientific field suffer from publications bias - “null” results are not being published - which leads to overestimated effect sizes (for review of this and other biases see (Ioannidis et al. 2014).
>
> Recent years have seen an increase in alarming signals about the lack of replicability in neuroscience, psychology, and other related fields [2]. Neuroimaging studies are usually statistically underpowered due to the high cost of data collection (Button et al. 2013), making the problem even harder. To avoid a widespread crisis in our field and consequently losing credibility in the public eye, we need to improve how we do science. This article aims to complement existing literature on the topic (Halchenko and Hanke 2015; Pernet and Poline 2015; Eglen et al. 2016) to be by compiling a practical guide for researchers at any stage of their careers that will help them make their research more reproducible and transparent while minimizing the additional effort that this might require.

** Publishing data: Open Science Framework connects researchers to the mentioned field agnostic repositories (figshare, dataverse) and others that researchers might want to use. Also, very good mention of data papers and pragmatics of when to consider writing them.*

The original version of the manuscript already included a paragraph introducing the concept of data papers, but we have extended the description of field agnostic repositories with the mention of OSF.

> *If for some reason field specific repositories are not an option we recommend using field agnostic repositories such as FigShare, Dryad, or DataVerse. When picking a repository one should think of long term data retention. No one can guarantee existence of a repository in the far future, but historical track record and the support of well established institutions can increase the chances that the data will be available in the decades to come. In addition a platform such as Open Science Framework (www.osf.io) can be used to link together datasets deposited in field agnostic repositories with code, pre-registrations and preprints (see below).*

** Code: "In addition to the lack of training, there are few incentives to spend the time necessary to generate high-quality and well-documented code. Changes in the incentive structure of science will take years, but in the meantime, perceived poor quality of code and lack of thorough documentation should not prevent scientists from publishing it [17] . Sharing undocumented code is a much better than not sharing code at all." The argument here could be made more compelling. As is, it probably will give the reader a sense that they really don't want to share their code (and the reference to incentives is not clear). Planning to share code is a way to increase likelihood of better documentation and, as a benefit to researchers, conduct internal review so that they are more likely to catch errors in advance - a benefit to them. Moreover, once code sharing is typical, then code reuse becomes typical, and everyone benefits from the development of standards and robust scripts.*

Producing high quality code can be a time consuming and expensive process. We regret that there are no clear incentives in the current academic system to produce high quality analysis code, but it is the reality and our practical guide needs to abide by it. We did already mention the advantage of planning ahead (and using version control system) in the previous version of the manuscript, but we manage to find one more compelling motivation for sharing code - citation rates:

> *Sharing undocumented code is a much better than not sharing code at all and can still provide benefits to the author. Perhaps the most compelling motivation for sharing code comes from citation rates. Papers accompanied by usable code are on average cited more often than their counterparts without the code (Vandewalle 2012).*

The code section also doesn't quite make the opening case for why sharing it is so important.

In the original version of the manuscript we have mentioned that sharing code is necessary to interpret original results, for reuse, and for scientific transparency and reproducibility (see below). We believe that those are strong arguments that should convince the reader about the importance of sharing code. Furthermore we believe that scientists understand the importance of sharing code, but are struggling with the incentives and practical solutions which we talk about extensively.

> *The code we write to analyze data is a vital part of the scientific process, and similar to data, is not only necessary to interpret and validate results, but can be also used to address new research questions. Therefore the sharing of code is as important as the sharing of data for scientific transparency and reproducibility.*

** Throughout the text, the benefits of openness and reproducibility should be made more explicitly. In many areas, the connection to benefits appears to be assumed.*

Throughout the paper we tried to explicitly list the benefits of open practices to the researcher (such as getting feedback on preprints, higher citation rates for shared data and code etc.). This aspect of the paper has been slightly expanded (see below). We also improved the explanation of the benefits for science as a whole in the introduction (see above). However, in many points we have referred the reader to other publications in order to keep the guide practical and focused on the perspective on the individual researcher.

** Might be helpful to note in data and code sharing sections which repositories allow embargoing (e.g., figshare, OSF) so that data and code can be deposited and privately shared with reviewers without exposing it before publication. This would allay scooping concerns (regardless of their reasonableness).*

We have added information about embargo functionality of figshare, Dryad, and OSF to both data and code sharing sections:

> *If one is concerned about losing competitive advantage by sharing data before the relevant manuscript will be accepted and published (so called “scooping”) one can consider setting an embargo period on the submitted dataset. OSF[1], figshare[2], and Dryad[3] support this functionality.*

> *We encourage scientists to use git rather than other VCS due to a passionate and rapidly growing community of scientists who use the GitHub.com platform, which is a freely available implementation of the git VCS system. In the simplest use case GitHub is a platform for sharing code (which is extremely simple for those who already use git as their VCS), but it also includes other features which make contributing to collaborative projects, reviewing, and testing code simple and efficient. The Open Science Framework mentioned above can be used to link together data and code related to a single project. It can also be*

[1] https://osf.io/faq/
[2] https://figshare.com/blog/The_future_of_figshare/166
[3] http://datadryad.org/pages/faq

used to set embargo period on the code so it could be submitted with the paper while minimising the risk of "scooping".

** The paper doesn't get into openness and reproducibility of the workflow itself - and that is where much of the action is in terms of the challenges for reproducibility (i.e., flexibility of analysis and selective reporting). It would be useful to have, for example, a section on preregistration, and discussion of open workflow more generally. A general application article that might provide some useful ideas that can be translated to specific considerations for neuroimaging is by the Open Science Collaboration: https://osf.io/nte3j/*

Thank you for this valuable comment and reference. Initially we planned not to mention preregistration, because we have limited personal experience with this mechanism and it does not have clear advantages for individual researchers. However, after some research we came up with a way of introducing this mechanism that would be appealing to broader audience:

Reduction of publication bias can be achieved even more effectively by adopting the pre-registration mechanism (Aarts et al. 2014). This way of doing research, originally adopted from clinical trials, involves writing and registering (in a third party repository) a study plan outlining details of data acquisition, subject exclusion criteria, and planned analyses even before that data have been acquired. This not only motivates researchers to formulate hypotheses before seeing data, but also allows for a clear distinction between results of hypothesis driven confirmatory analyses (include in the pre-registration) and exploratory analyses (added after seeing the data). It is worth mentioning that exploratory analyses are by no means inferior to confirmatory analyses. They are an important part of science, generating new hypotheses that could be confirmed by future studies. However exploratory analyses can suffer from bias (since their inception was influenced by the data itself) and thus require additional evidence. Unfortunately without preregistration confirmatory and exploratory analyses are often not labeled correctly. Preregistration also plays a vital role in highlighting hypothesis that turned out not to be confirmed by the data ("null effects").

It is clear that preregistration can help in research transparency and reproducibility by reducing biases. It is also important to acknowledge that putting together and registering a binding research plan requires a significant time investment from the researcher and thus is not common a common practice (with exception to replication studies (Boekel et al. 2015; Open Science Collaboration 2015)). There are however additional incentives for individual researchers to preregister their studies. Center for Open Sciences spearheaded the Registered Reports[4] initiative in 2012. According to this mechanism authors send their preregistartion reports (Introduction, Methods parts of a future paper and optionally analysis of pilot data) for peer review to a journal for peer review. Validity of the experimental plan is assessed and if deemed sufficient receives "In-principle acceptance" (IPA). The journal

[4] https://osf.io/8mpji

guarantees to publish the final version of the paper (after data collection and analysis) independently of the results (so even if the hypothesised effect was not found). This mechanism even though requires additional work brings stability to the publication process. Currently journals accepting in neuroimaging papers participating in the Registered Reports program include: AIMS Neuroscience (Chambers et al.), Attention, Perception, and Psychophysics, Cognition and Emotion (Fischer and van Reekum 2015), Cortex (Chambers 2013) and European Journal of Neuroscience. Additionally Center for Open Science started a Preregistration Challenge[5] providing $1000 reward for the first 1000 preregistered eligible studies. This initiative is independent of the Registered Reports, does not guarantee publication, but the list of eligible journals is much longer (includes such journals as PloS Biology, Hippocampus, or Stroke).

Reviewer #3

https://pubpeer.com/publications/CF1443E6D3B7CD0E1EA9EB7E1731C1

Interesting piece on the what open science can do to improve reproducibility - I guess many aspects have been discussed by others and the important point is the 'practical' in this respect I think the final version should have 'boxes' or similar with lists of tools/websites etc .. also maybe reference other tutorial papers on good practices etc .. see my own paper on the topic http://gigascience.biomedcentral.com/articles/10.1186/s13742-015-0055-8

We have added a box summarizing the main points in the paper.

We have added relevant citations to the introduction.

> This article aims to complement existing literature on the topic (Halchenko and Hanke 2015; Pernet and Poline 2015; Eglen et al. 2016) by compiling a practical guide for researchers at any stage of their careers that will help them make their research more reproducible and transparent while minimizing the additional effort that this might require.

specific comments:

- introduction: when talking about the plurality of analyses methods, why it is both good and bad?

We have extended the relevant paragraph to include a better introduction on the topic:

> *Having access to a plethora of denoising and modelling algorithms can be both good and bad. On one side there are many aspects of brain anatomy and function that we can extract and use as dependent variables which maximizes the chances of finding the right fit for a*

[5] https://cos.io/prereg/

> *particular question. On the other side the passion driving brain researchers combined with methodological plurality can be a dangerous mix. Scientists rarely approach a problem without a theory, hypothesis, or a set of assumptions and those can implicitly drive them to implicitly choose results of some analyses over others leading to increased "researcher degrees of freedom" (Simmons et al. 2011). As Richard Feynman said "The first principle is that you must not fool yourself — and you are the easiest person to fool.".*

- introduction: talking about power, do you have data on typical sample size? (would be good) + please ref articles about power and neuroimaging (not just fmri)

The only estimate of power in neuroimaging we are aware was included in Button et al. 2013. They found that median power across a few hundret papers wa 8%. We have included this information in the manuscript.

> *Neuroimaging studies generally have low statistical power (estimated at 8%), due to the high cost of data collection, which results in an inflation of the number of positive results that are false (Button et al. 2013).*

- consent forms: what is substantial risk? if by sharing the age, location and weird behaviour due to a brain injury, people leaving nearby the patient can recognize the person, is that a risk? no - still do we want that ? I don't think so . Here is an example of such issue: http://www.thelancet.com/journals/laninf/article/PIIS1473-3099%2810%2970243-7/abstract

When deciding how much to restrict access to data one has to take into account two factors: 1) how hard would it be to identify the participants (because it's never impossible) 2) if the participant was identified what consequences what that have (this is the "substantial risk" we were talking about). The case you brought up was unusual in both of those dimensions - due to carelessness of researchers identifying a participant was easy. The consequence of the participants identity being revealed could cause negative backlash in his community due to the disease the person was carrying. We have rephrased this paragraph to make it more clear:

> *We recommend using the restricted access version only for data and populations for which a) potential data re-identification is easy due to small sample and/or the level of detail of included variables (for example exact time and location of scanning) or b) re-identification would lead to negative consequences for the participants (for example in a study of HIV-positive subjects).*

- publishing data: maybe it is worth mentioning licensing - see Yaroslav's paper http://gigascience.biomedcentral.com/articles/10.1186/s13742-015-0072-7

Thank you for your suggestion. The lack of discussion of licensing was definitely an oversight on our side. We have added the following paragraphs to the paper to cover data...

> *Finally published data should be accompanied by an appropriate license. Data is treated differently by the legal system than creative works (i.e. papers, figures) and software and thus requires special licenses. Following the lead of major scientific institutions such as BioMed Central, CERN, or The British Library we recommend using unrestricted Public Domain license (such as CC0) for data. Using such license would maximize the impact of the shared data, by not imposing any restriction on how it can be used and combined with other data. The appropriate legal language that needs to accompany your data can be obtained from https://creativecommons.org/publicdomain/zero/1.0/. There are also other more restrictive license options (see http://www.dcc.ac.uk/resources/how-guides/license-research-data). However, additional restrictions can have unintended consequences. For example Non-Commercial clause in its most broad interpretation might prevent your data from being used for teaching or research at a private university. No-Derivatives clause, on the other hand, prevents your data from being combined in any form with other data (for example a brain template released under No-Derivatives license cannot be used as a coregistration target).*

as well as software licenses:

> *As with data, it is important to accompany your code with an appropriate license. Following (Halchenko and Hanke 2015) we recommend choosing a license that is compatible with the open source definition such as Apache 2.0 or General Public License (GPL)[6]. The most important concept to understand when choosing a license is "copyleft". A license with a "copyleft" property (such as GPL) allows derivatives of your software to be published, but only if done under the same license. This property limits the range of code your software can be combined with (due to license incompatibility) and thus can restrict the reusability of your code. Choosing an open source license and applying it to your code can be greatly simplified by using a service such as choosealicense.com.*

- how to deal with publications: if in a particular experiment you have N main contrasts and report only a subset, then you create a bias in literature (which means we should upload all maps on neurovault btw) = selective reporting ; sharing code and all maps is clearly a better way for meta-analyses (and power analyses)

We have extended the paragraph about reporting null results with your specific suggestions.

> *A clear narrative can be provided in the main body of the manuscript and the details of methods used together with null results and other analyses performed on the dataset can be included in the supplementary materials, as well as in the documentation of the shared*

---

[6] https://opensource.org/licenses

code. In this way, the main narrative of the paper is not obfuscated too many details and auxiliary analyses, but all of the results (even null ones) are available for the interested parties. Such results from extra analyses could include for example all of the additional contrasts that were not significant and thus not reported in the main body of the manuscript (of which unthresholded statistical maps should be shared for example using a platform such as NeuroVault). Often these extra analyses and null results may seem uninteresting from the author's point of view, but one cannot truly predict what other scientists can be interested in. In particular, the null results (which are difficult to publish independently) can contribute to growing body of evidence that can be used in the future to perform meta analyses.

- for previous paper published under the 'paywalled' model, consider self archiving the author version, making the article freely available, albeit without journal editing

This is a very interesting question. We have researched the topic and added the following paragraph to the manuscript:

To further improve accessibility and impact of research outputs one can also consider sharing papers that have already been published in subscription based journals. Unfortunately this can be difficult due to copyright transfer agreements many journals require from authors. Such agreement give the journal exclusive right to the content of the paper. However, each publisher uses a different set of rules and some of them allow limited sharing of your work you have surrender your rights to. For example Elsevier (publisher of NeuroImage) allows authors to publish their accepted manuscripts (without the journal formatting) on a non-commercial website, a blog or a preprint repository[7]. Wiley (publisher of Human Brain Mapping) has a similar policy for submitted manuscripts (before the paper gets accepted), but requires an embargo of 12 months before authors can share the accepted manuscript[8]. Policies for other journals might vary. SherPa/ROMEO (http://www.sherpa.ac.uk/romeo) is a databaset that allows authors to quickly check what the journal they published with allows to share and when.

There are multiple options when it comes to choosing a place to share manuscripts published with subscription based journals. Private websites, institutional repositories, and preprint servers seems to be well within the legal restrictions of most journals. Commercial websites such as researchgate.com and academia.edu remain a grey zone (with some reports of Elsevier taking legal actions toremove papers from one of them[9]). If the research has been at least partially funded by NIH one can deposit the manuscript in PubMed Central (respecting appropriate embargos)[10].

---

[7] https://www.elsevier.com/about/company-information/policies/sharing
[8] http://olabout.wiley.com/WileyCDA/Section/id-826716.html
[9] http://svpow.com/2013/12/06/elsevier-is-taking-down-papers-from-academia-edu/
[10] https://nihms.nih.gov

Reviewer #4 (Yaroslav Halchenko)

http://biorxiv.org/content/early/2016/02/12/039354#comment-2526132533

Well -- although it will sound like I am just trying to squeeze a citation in http://www.gigasciencejournal.com/content/4/1/31 *but I guess I wrote that piece for a reason.*

You state that pillars of "Open" science are data, code and papers... but oh well -- none of those alone make it "Open". Even deposition of all those online and publishing in "open access" journals doesn't make them sustainably open, since those might (and do) go away in X years, and if noone had permissions to duplicate, reuse, improve upon your works -- what "open" is that? It is as open as a door of a limo standing at your doorstep but which you cannot ride.

To guarantee that science product is open, it must be allowed for its widest dissemination and reuse. For that clear statement of copyright and license terms must be made, and no "exclusive licenses" be provided to take away those freedoms and place them into a single hand (which is often the case with publications in some "open access" journals). But this manuscript doesn't even mention a word "license". IMHO it is ignoring an elephant in the room ;)

Thank you for your suggestion. As noted above, we have now added discussion of licensing for data and software.

Reviewer #5 (Cameron Craddock)

http://biorxiv.org/content/early/2016/02/12/039354#comment-2579979522

Nice paper. I would like to clarify that although the 1000 Functional Connectomes and INDI projects do share a lot of resting state fMRI data, they are far from resting state only. INDI in particular includes a variety of task data, and very large repositories of structural MRI, diffusion, and arterial spin labelling data. We have yet to turn down any datasets based on the types of data they include. It might also be worthwhile to point out that INDI includes data from patient populations and with deep phenotyping.

We have changed the wording to "for any datasets that include resting state fMRI and T1 weighted scans". Thank you for your suggestion. (for full discussion see: http://biorxiv.org/content/early/2016/02/12/039354#comment-2579979522)