

Bonnes pratiques et science ouverte : Un document d'accompagnement au PRD



(Image traduite de <https://vusci.blog/2020/04/17/reproducibilitea-blends-and-flavors/>)

Document préparé par : Rémi Thériault
Département de psychologie, Université du Québec à Montréal
Version : 6 juin 2023

Pour des commentaires, questions, ou rétroactions sur ce document, veuillez contacter Rémi Thériault (theriault.remi@courrier.uqam.ca).



Table des matières

Bonnes pratiques et science ouverte : Un document d'accompagnement au PRD.....	3
Comment utiliser ce document lors d'une rencontre de PRD?	4
Contexte	4
La recherche exploratoire vs. la recherche confirmatoire.....	5
Que sont les pratiques de recherche questionnables?	7
Les bonnes pratiques en recherche et la science ouverte	10
Le préenregistrement	13
Le rapport enregistré	15
Le Projet de recherche doctoral	16
Les données ouvertes	17
Les matériels ouverts	20
L'utilisation du logiciel R	21
La prépublication	23
La publication en libre accès.....	26
Les bénéfices (et limites) d'embrasser la science ouverte	28
OK, et maintenant?	29
« Checklist » à partager à votre direction, comité, ou collègues	30
Remerciements.....	32
Références	32

Bonnes pratiques et science ouverte : Un document d'accompagnement au PRD

Ce document est un guide visant à vous informer et vous motiver à adopter les bonnes pratiques reliées à la science ouverte, et ainsi à améliorer la qualité des thèses doctorales. À ce titre, l'objectif est de vous convaincre de l'importance et des avantages d'intégrer des pratiques de recherche ouvertes et transparentes et à éviter les pratiques de recherche questionnables, dont plusieurs sont désormais considérées comme contraires à l'éthique.

Ce guide est inspiré, et complémente, le document « [Statistical Methods in Theses: Guidelines and Explanations](#) » (Al-Aidroos et al., 2018) publié sur le site du département de psychologie de l'Université de Guelph et signé par des professeures et professeurs qui enseignent les cours de statistiques ou de méthodologie : Naseem Al-Aidroos, PhD, Christopher Fiacconi, PhD, Deborah Powell, PhD, Harvey Marmurek, PhD, Ian Newby-Clark, PhD, Jeffrey Spence, PhD, David Stanley, PhD, et Lana Trick, PhD. Il peut être très constructif de consulter le document de Al-Aidroos et collègues après avoir terminé la lecture du présent guide pour les gens désirant approfondir leurs apprentissages.

Ce guide s'adresse en particulier aux étudiantes et étudiants du doctorat en psychologie de l'Université du Québec à Montréal. L'intention est que ce document soit suggéré comme lecture dans le cadre du Projet de recherche doctoral (PRD). Ce document final sera éventuellement disponible sous format PDF sur le site du Département de psychologie de l'Université du Québec à Montréal et celui de l'Association générale des étudiants et étudiantes aux cycles supérieurs en psychologie (AGEPSY-CS).

Il est suggéré de lire ce document avant le cours PSY8331 – *Séminaire lié à l'essai ou à la thèse*, et préféablement, dès le début du doctorat, puisque cela peut aider à mieux encadrer le projet de recherche proposé. Il est également recommandé d'apporter ce document avec soi lors des rencontres avec sa direction de recherche, ainsi qu'avec les membres du comité de PRD. Ceci permettra d'aider tous les partis concernés à mieux juger de la qualité et du caractère de répliquabilité et de transparence du projet de recherche proposé.

Comment utiliser ce document lors d'une rencontre de PRD?

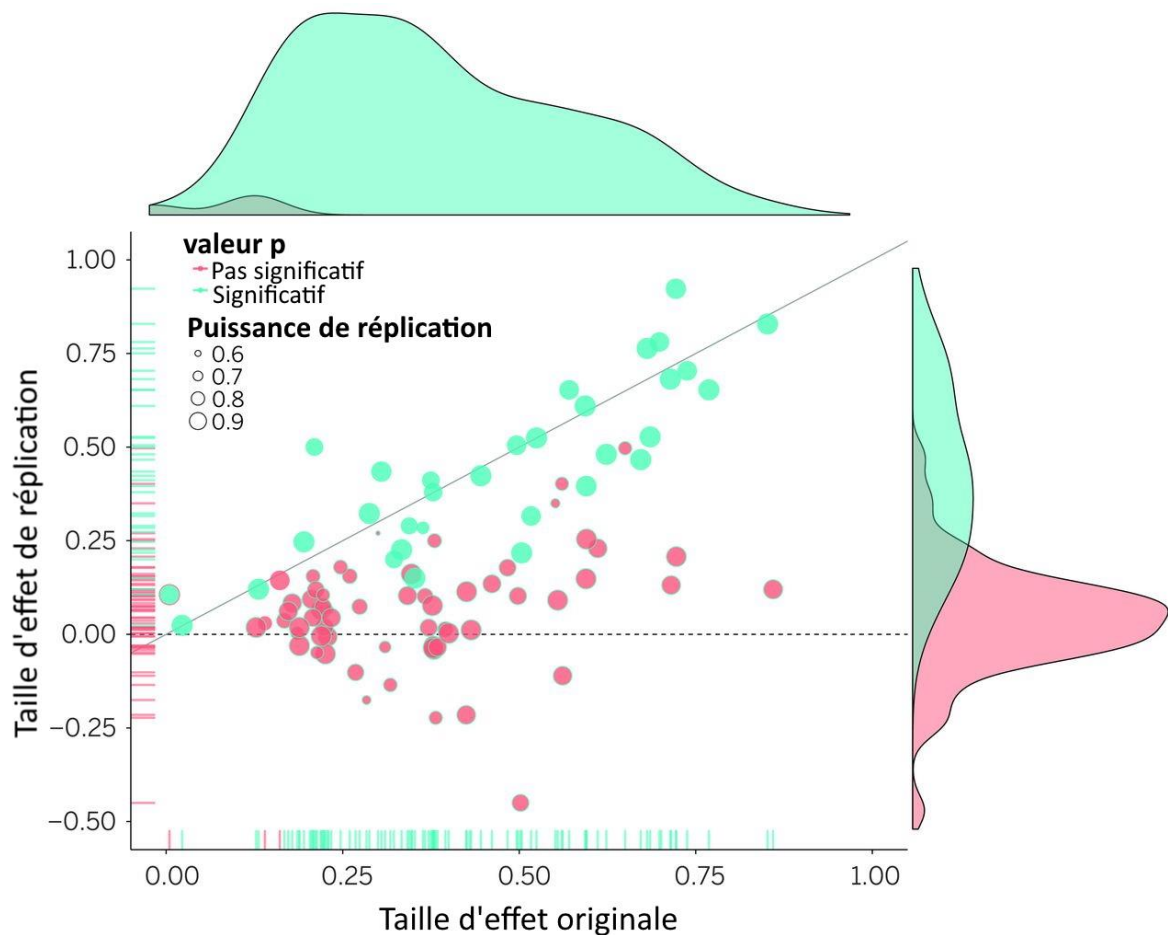
1. Par exemple, vous pourriez imprimer une copie de ce document et l'apporter à vos rencontres de PRD.
2. Au cours de ces rencontres, vous pourriez utiliser le document comme point de repère pour demander l'aide des membres de votre comité présents afin de résoudre de potentiels problèmes à l'avance.
3. Vous pourriez également revenir à ce document et aux problèmes qu'il soulève (en particulier vis-à-vis des étapes d'analyse et de diffusion) lorsque vous vous préparez pour l'examen doctoral ou bien encore pour la soutenance doctorale.

Contexte

Vous avez peut-être déjà entendu parler de la « crise de réplication » en psychologie et dans d'autres domaines scientifiques (Camerer et al., 2018; Ioannidis, 2005; Open Science Collaboration, 2015). Essentiellement, plusieurs membres de la communauté scientifique se sont penchés sur des résultats classiques de la littérature en psychologie avec l'intention de les répliquer, mais de nombreuses études n'ont pas été reproduites. La Figure 1 en page suivante, par exemple, illustre que les tailles d'effet de nombreuses études de réplication sont souvent beaucoup plus petites que les tailles d'effet des études originales, et même que de nombreuses tentatives de réplifications sont incapables de répliquer les effets originaux. Ceci a poussé les équipes de recherche à demander une plus grande rigueur dans notre discipline, avec la croyance que l'échec de réplication serait dû notamment à des pratiques de recherche questionnables (c.-à-d., qui réduisent notre confiance dans les résultats) et aux biais de publication (soit la tendance des revues scientifiques de ne publier que les résultats significatifs; Ferguson & Heene, 2012). Cette tendance de ne publier que les résultats significatifs aurait tendance à mettre de la pression sur les chercheuses et chercheurs pour trouver de tels résultats à tout prix, quitte à s'engager dans des pratiques questionnables. En contrepartie, un nouveau mouvement est né de ce choc : la science ouverte. Celui-ci met au cœur de sa démarche la transparence, la reproductibilité et les bonnes pratiques de recherche. Le présent guide couvre brièvement les pratiques questionnables avant d'explorer plus en profondeur les bonnes pratiques associées à la science ouverte. Il fait ensuite le pont avec la réalité uqamienne, et en particulier, avec le PRD, que les personnes doctorantes en psychologie doivent compléter au plus tard dans leur troisième année.

Figure 1

Taille d'effet de l'étude originale par rapport à la taille d'effet de la réplication (coefficients de corrélation)



Note. Figure adaptée de Open Science Collaboration (2015). Chaque cercle représente une corrélation entre la taille d'effet d'une étude originale et la taille d'effet d'une étude de réplication. La ligne diagonale représente la taille d'effet de réplication égale à la taille d'effet d'origine. La ligne pointillée représente une taille d'effet de réplication de 0. Les points sous la ligne pointillée étaient des effets dans la direction opposée à l'original. Les diagrammes de densité sont séparés par des effets significatifs (bleu) et non significatifs (rouge). Ce que l'on tire de cette figure, c'est que les tailles d'effet de nombreuses réplifications sont souvent plus petites que les tailles d'effet des études originales, et pareillement, de nombreuses tentatives de réplifications ne trouvent pas d'effet significatif.

La recherche exploratoire vs. confirmatoire

Pour mieux comprendre le débat sur les pratiques questionnables et la science ouverte, il faut comprendre la nuance entre la recherche exploratoire et confirmatoire (Nosek et al., 2018; Wagenmakers et al., 2012). La recherche *exploratoire* vise à identifier de possibles patrons dans

les données pour générer de nouvelles hypothèses qui pourront éventuellement être testées (*hypothesis-generating framework*). C'est la première étape. La recherche *confirmatoire* quant à elle vient normalement en deuxième et vise à tester les hypothèses formulées basé sur la théorie ou sur des recherches exploratoires précédentes (*hypothesis-testing framework*). Les deux approches ont leur places, mais celles-ci devraient généralement être équilibrées, de sorte que les résultats exploratoires sont éventuellement confirmés avec un certain degré de certitude, qui permet des applications plus concrètes de ces résultats. À ce titre, tant que les résultats de recherches exploratoires ne sont pas confirmés, ils ne devraient pas informer les politiques publiques ou les interventions cliniques (ou si c'est le cas, avec énormément de prudence; Chambers, 2018). En effet, nous voulons être bien sûr qu'un nouveau type de thérapie par exemple pourra vraiment aider les clients plutôt que perdre leur temps, ou pire, leur nuire!

Pour l'exercice, imaginons un instant que la majorité des recherches en psychologie étaient exploratoires. Il serait à ce moment-là difficile de développer des certitudes vis-à-vis de nos trouvailles scientifiques. On aurait de très nombreuses « pistes intéressantes et prometteuses », mais on ne saurait pas lesquelles seraient véritables, reproductibles, ou bénéfiques. Vous comprenez peut-être maintenant un peu mieux l'enjeu actuel : nous nous trouvons exactement dans cette situation hypothétique. Alors que la plupart des articles scientifiques sont écrits avec une allure d'autorité et de certitude, comme si l'équipe de recherche avait tout parfaitement prédit depuis le début, de nombreux articles ne révèlent pas l'ampleur de la teneur exploratoire de leurs démarches, ce qui amène à tort le lectorat à attribuer un haut degré de confiance à leurs recherches. Ce n'est donc pas surprenant que nous nous retrouvions maintenant avec une crise de réplication collective.

De prime abord, il faut clarifier que ce guide s'applique plus particulièrement aux types de recherche qui s'inscrivent dans cette approche confirmatoire par hypothèse, notamment la recherche empirique quantitative usant de données primaires. À ce titre, les présentes recommandations s'appliquent moins bien à d'autres types de recherche, comme la recherche qualitative, explicitement exploratoire, ou usant de données secondaires. Ne vous méprenez pas : la recherche exploratoire a de nombreux avantages. Sans elle, de nombreuses découvertes scientifiques importantes ou développements théoriques auraient été manqués. Mais tout est question d'équilibre et surtout, de *transparence*. Si vous décidez de conduire de la recherche exploratoire, faites preuve de transparence et communiquez-le clairement. En revanche, si

l'intention est belle et bien de conduire une recherche confirmatoire—ou du moins avec cette prétention—faites de votre mieux pour éviter les pratiques de recherche questionnables.

Que sont les pratiques de recherche questionnables?

Les pratiques de recherche questionnables sont des pratiques qui peuvent miner la validité et l'interprétation des résultats, mais qui ne constituent pas une fraude ou de la méconduite en recherche. À ce titre, elles peuvent dans certains cas constituer une zone grise. Cependant, ces pratiques sont à éviter lorsque possible puisqu'elles peuvent miner la confiance dans les résultats de recherche puisqu'elles permettent aux chercheur.es d'en quelque sorte « manipuler » (gamer) le système pour obtenir des résultats significatifs. Les chercheuses et chercheurs peuvent s'engager dans ces pratiques sans mauvaise intention et même sans s'en rendre compte, d'où l'importance de la conscientisation sur ce sujet. Ces pratiques peuvent également créer des dilemmes éthiques puisque de nombreux chercheurs et chercheuses et étudiants et étudiantes ressentent de la pression à publier des résultats significatifs. Voici quelques exemples dans le tableau ci-dessous (p. ex., Schwab et al., 2022; Wicherts et al., 2016) :

Pratique questionnable	Explication
	Orienter les décisions de l'équipe de recherche tout au long du processus d'analyse dans l'intention de favoriser une valeur p significative (c.-à-d., de capitaliser sur la chance et les « degrés de liberté », la flexibilité de l'équipe de recherche).
<i>Le « p hacking »</i>	L'équipe de recherche doit prendre de nombreuses décisions : comment gérer les inclusions et exclusions, les données manquantes, les données aberrantes, les assumptions statistiques, le fait de standardiser les données (ou pas), le fait de tester différentes stratégies pour quantifier

	<p>une variable, le choix précis des tests statistiques et de leurs paramètres, le fait de corriger pour de multiple tests (ou pas), le seuil de signification, etc. Si l'équipe de recherche fait constamment des choix basés sur l'option qui dans chaque cas favorise une valeur p plus significative, il s'agit de p hacking.</p> <p>Cela, plutôt que de se tenir à un plan expérimental et d'analyse déterminé à l'avance en fonction de raisons théoriques et méthodologiques (p. ex., une méthode identifiée comme étant la pratique généralement recommandée). Cependant, il est difficile de tout prévoir. Nous recommandons donc de faire de son mieux, tout en restant raisonnable.</p>
<i>L'arrêt arbitraire de la collecte de données</i>	<p>Examiner les données (ou pire, refaire les analyses à plusieurs reprises) pendant la collecte de données, et arrêter celle-ci aussitôt que les résultats désirés ressortent ou au contraire poursuivre la collecte afin d'obtenir un résultat significatif, plutôt que de s'en tenir à une taille d'échantillon prédéterminée.</p>
<i>Changer dans le modèle le rôle des variables de manière arbitraire</i>	<p>Modifier quelles variables sont indépendantes, dépendantes, médiatrices, modératrices, ou bien des covariables en se basant sur le résultat le plus intéressant ou le plus significatif plutôt que sur le modèle théorique défini a priori.</p>

<i>Le rapport sélectif des variables</i>	Ne pas rapporter toutes les variables qui ont été incluses dans l'étude (s'applique aux collectes de données primaires).
<i>Le rapport sélectif des conditions expérimentales ou sous-groupes</i>	Ne pas rapporter tous les groupes ou sous-groupes expérimentaux.
<i>Le rapport sélectif des analyses</i>	Ne pas rapporter toutes les analyses ou tests statistiques qui ont été effectuées
<i>Le rapport sélectif des études</i>	Ne pas rapporter toutes les études qui ont été effectuées
<i>Le « HARKing » (<u>H</u>ypothesizing <u>A</u>fter the <u>R</u>esults are <u>K</u>nown)</i>	Changer ou générer nos hypothèses en fonction des résultats de tests statistiques exploratoires.
<i>Recruter des échantillons trop petits pour la méthode d'analyse choisie</i>	<p>Car des échantillons trop petits mènent à une trop faible puissance statistique. Si la taille d'échantillon doit être limitée dû à des contraintes financières ou logistiques, alors il faudrait prévoir utiliser une méthode alternative plus appropriée pour la taille d'échantillon réaliste. Avec trop peu de puissance statistique, il ne sera pas possible de détecter un effet, même s'il existe réellement, gaspillant ainsi de précieuses ressources financières et de temps.</p> <p>Heureusement, une analyse de puissance est généralement requise lors du PRD.</p>
<i>Ne pas communiquer explicitement si les analyses sont confirmatoires ou exploratoires</i>	Ne pas spécifier quelles sont les analyses exploratoires peut amener les gens qui liront votre article à penser qu'il s'agit d'une étude confirmatoire (avec une plus grande crédibilité) même si elle est en réalité exploratoire, ce qui est trompeur.

Gardez en tête que la différence entre la recherche exploratoire et confirmatoire en est une de degrés, et nécessairement, certaines décisions improvisées seront nécessaires même dans les cas les plus rigoureux. En ce sens, les exemples de pratiques de recherche questionnables ci-haut sont à titre indicatif et non pas absolus : vous pourriez avoir de très bonnes raisons d'en appliquer quelques-unes. Celles-ci deviennent surtout problématiques lorsqu'elles s'accumulent ou qu'elles viennent affecter l'interprétation des résultats de manière importante. Il faut donc rester raisonnable tout en priorisant la transparence dans nos prises de décisions.

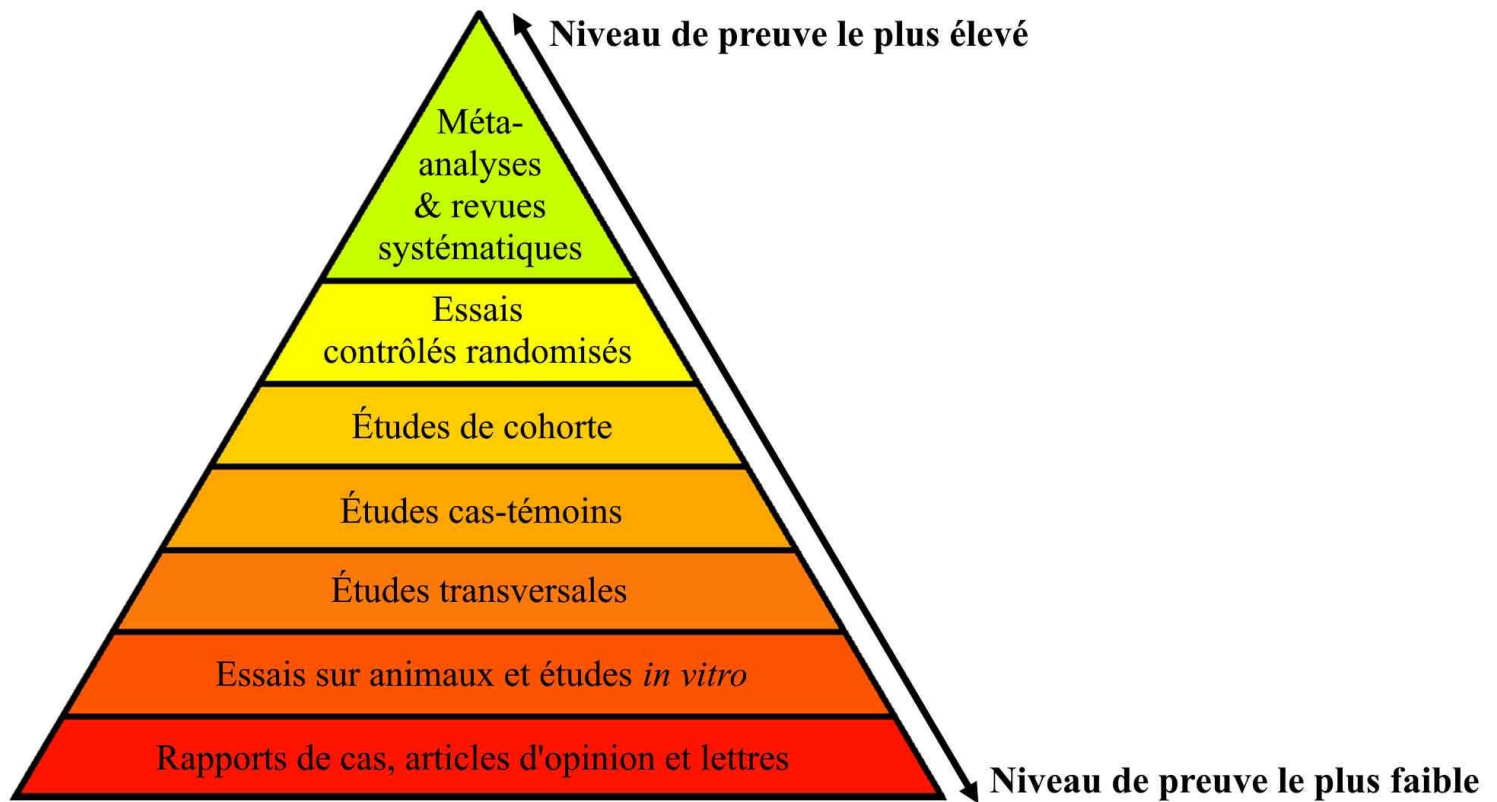
Les bonnes pratiques en recherche et la science ouverte

Vu l'omniprésence des pratiques questionnables en recherche, de nombreuses équipes de recherche ressentent qu'elles ne peuvent se fier aux résultats des recherches passées et présentes (Haefel, 2022). En revanche, la science ouverte, un mouvement visant à rendre l'ensemble du processus scientifique transparent et accessible, permet d'adresser directement plusieurs de ces pratiques questionnables (Schwab et al., 2022). À ce titre, elle met au cœur de sa démarche la transparence, la reproductibilité, et les bonnes pratiques de recherche. C'est pourquoi de nombreuses personnes croient que c'est la science ouverte qui pourra rétablir la confiance en la science (particulièrement en psychologie). Une facette importante de la science ouverte, c'est qu'elle permet de distinguer clairement la recherche exploratoire de la recherche confirmatoire. Comme souligné précédemment, certain.es croient ainsi qu'une proportion grandissante d'études devraient être confirmatoires plutôt qu'exploratoires (Wagenmakers et al., 2012). Chambers (2018) a, par exemple, proposé une nouvelle pyramide hiérarchique de la preuve scientifique visant à informer nos réflexions, notre perception, et notre discours autour des données probantes (Figures 2 et 3).

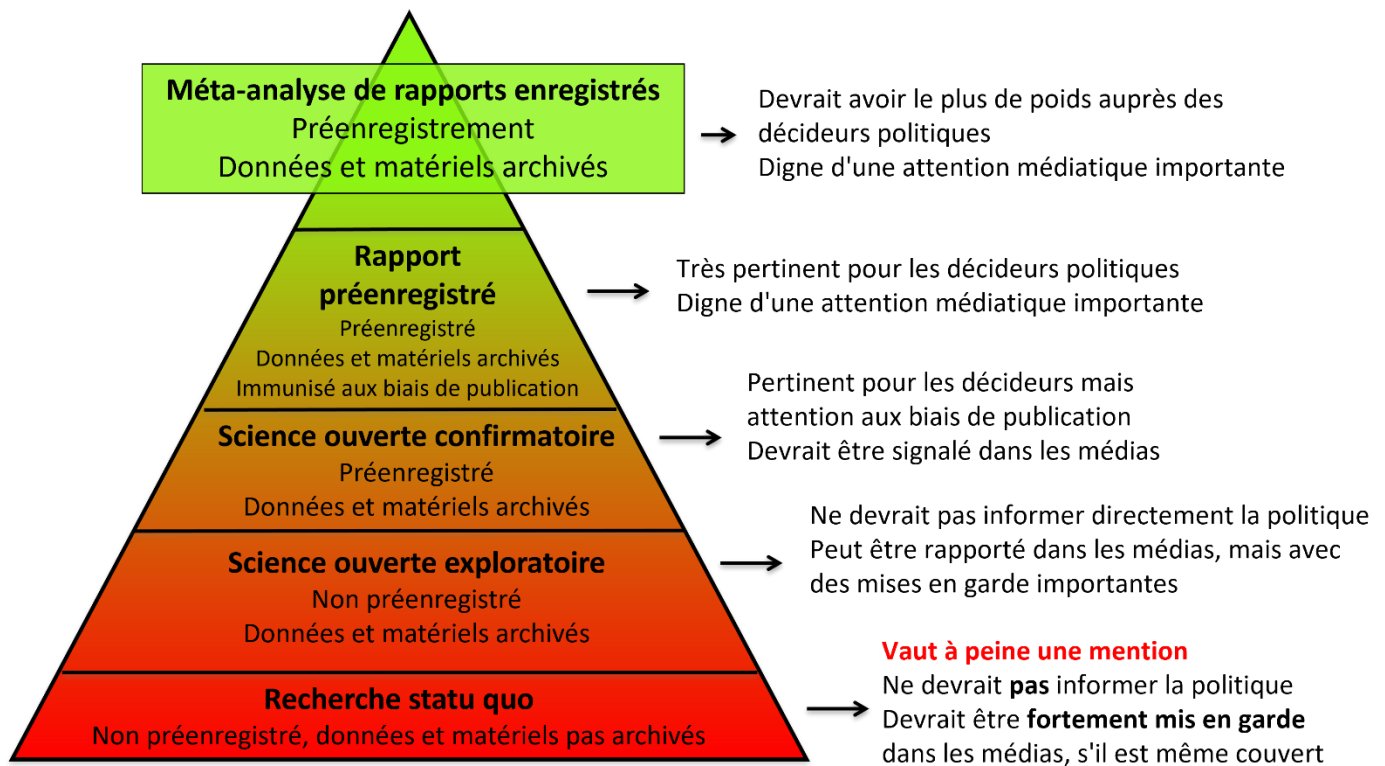
La science ouverte recommande de nombreuses pratiques, dont certaines continuent d'évoluer. Les plus populaires incluent (a) le préenregistrement de l'étude; (b) le rapport enregistré; (c) la mise en ligne publique des données désidentifiées; (d) la mise en ligne des matériels et de la syntaxe; (e) l'utilisation de logiciels libres; (f) la prépublication; et (g) la publication en libre accès. Dans les sections suivantes, nous décrivons chacune de ces pratiques un peu plus en détails (voir également la « checklist » à partager à votre direction, comité, ou collègues à la fin de ce document).

Figure 2

La hiérarchie conventionnelle de la preuve scientifique



Note. Figure adaptée de thelogicofscience.com.

Figure 3*La nouvelle hiérarchie de la preuve scientifique**Note.* Figure adaptée de Chambers (2018).

Le préenregistrement

Le préenregistrement de sa recherche est une étape (un peu plus intimidante pour certaines personnes) qui arrive tôt dans le processus de la recherche, soit avant de mettre ses données ou son script d'analyse en ligne (puisque théoriquement la collecte de données n'a pas encore débuté!). Il s'agit de s'engager publiquement à suivre un plan d'analyse spécifique. On spécifie donc : le devis de recherche (p. ex., expérimental, transversal, longitudinal), nos variables indépendantes (nos prédicteurs ou nos conditions expérimentales), nos variables intermédiaires (modérateurs, médiateurs), si applicable, nos variables dépendantes (sur lesquelles on pense observer un effet d'intérêt), nos hypothèses (donc les liens attendus entre toutes ces variables), la taille d'échantillon et comment celle-ci a été déterminée, ainsi que nos stratégies de recrutement, nos critères d'inclusion et d'exclusion, toute transformations prévues aux données, et le choix des analyses statistiques en tant que tel. À lui seul, le préenregistrement est le meilleur outil contre les pratiques questionnables, car il enlève beaucoup de « degrés de liberté » aux équipes de recherche en les encourageant à penser à ces éléments à l'avance, ce qui permet de grandement augmenter la confiance dans les résultats des analyses statistiques. Comme le souligne Goldacre (2009) :



What's truly extraordinary is that almost all of these problems—the suppression of negative results, data dredging, hiding unhelpful data, and more—could largely be solved with one very simple intervention that would cost almost nothing: a clinical trial register, public, open, and properly enforced (...) Before you even start your study, you publish the 'protocol' for it, the methods section of the paper, somewhere public. This means that everyone can see what you're going to do in your trial, what you're going to measure, how, in how many people, and so on, before you start. The problems of publication bias, duplicate publication and hidden data on side-effects—which all cause unnecessary death and suffering—would be eradicated overnight, in one fell swoop. If you registered a trial, and conducted it, but it didn't appear in the literature, it would stick out like a sore thumb. (Goldacre, 2009, pp. 220–221)

Bien que certaines personnes voient le préenregistrement comme du travail additionnel, en réalité, on ne fait que déplacer le travail dans le temps (Quintana, 2020). Plutôt que de se presser de collecter les données, pour ensuite réfléchir aux analyses et aux hypothèses, on réfléchit d'abord aux analyses et hypothèses, et après on débute la collecte des données. En

d'autres termes, cette pratique a un effet neutre ou positif en termes de gain de temps, mais permet un énorme gain pour la science.

Votre préenregistrement ne sera jamais parfait, et c'est OK!

Malgré nos meilleures intentions lors du préenregistrement de notre étude, il y aura (presque) toujours des imprévus de toutes sortes qui nous contraindront à dévier de notre protocole original. Ce pourrait être par exemple des problèmes de puissance inattendus ou bien de nouvelles conclusions de d'autres études récentes qui entraînent la nécessité d'ajuster le modèle à tester. Ceci est tout à fait normal et acceptable! Il suffit simplement d'être transparent face à ces déviations et le noter et de le justifier dans votre manuscrit. Mettez-vous à la place des gens qui liront votre article : si vous justifiez bien vos décisions, les gens seront raisonnables et il ne devrait pas y avoir de problème.

En plus de la transparence accrue et du risque réduit de décisions ad hoc, cette pratique a aussi l'avantage d'aider à détecter des faiblesses potentielles de l'étude avant que la collecte de données ne commence. Elle donne également une plus grande place à la réflexivité et la théorie lors de la phase de conceptualisation de la recherche pour s'assurer que celle-ci réponde vraiment aux objectifs visés. Cet aspect est intéressant puisque d'autres scientifiques croient justement que les équipes de recherche doivent [donner à la réflexivité et la théorie leur juste place dans le processus de recherche](#). C'est que cette étape est malheureusement trop souvent complétée à la hâte, dans l'espoir de publier rapidement (Rawat & Meena, 2014; Sarewitz, 2016).

Enfin, le préenregistrement des recherches est de plus en plus prisé par la communauté scientifique. Alors qu'il peut être normal de se demander si le préenregistrement pourrait nuire aux chances de publier, c'est plutôt le contraire : toutes choses étant égales par ailleurs, des résultats non-significatifs ont plus de chance d'être publiés en cas de préenregistrement. De plus, même en cas de résultats préenregistrés non-significatifs, rappelez-vous que vous pouvez également inclure des résultats exploratoires. Il n'y a donc rien à perdre à s'engager dans ce processus. Ainsi, pour encourager le préenregistrement, de plus en plus de journaux ([comme les journaux de l'American Psychological Association, l'APA](#)) offrent des « badges » (insignes) pour encourager les pratiques de science ouverte. Le badge « Préenregistré » est parmi les plus prestigieux.

Qu'en est-il de l'analyse de données secondaires?

La recommandation du préenregistrement s'applique très bien aux études impliquant de nouvelles collectes de données (données primaires). Toutefois, pour les études impliquant l'analyse de données secondaires à partir de données existantes, de nombreuses analyses préliminaires sont nécessaires pour examiner les variables disponibles et leurs propriétés psychométriques, ainsi que la taille de l'échantillon disponible, afin d'examiner la faisabilité de l'étude envisagée. Ceci fait de l'analyse de données secondaires une situation spéciale où le concept de préenregistrement est plus difficile à appliquer.

Néanmoins, [cela est tout à fait possible](#), et un modèle prêt à être utilisé a été développé à cet effet sur l'Open Science Framework. Il existe même un nombre croissant de directives pour encadrer et réduire les biais lors de l'analyse de données secondaires (voir p. ex. Baldwin et al., 2022 pour une analyse des défis et solutions possibles). De nouveaux paradigmes voient également le jour, comme le « Explore and Confirm Analysis Workflow (ECAW) », où dans une première étape une institution de gestion des données donne accès à un échantillon d'une plus grande banque de données pour permettre l'exploration des données et formuler des hypothèses. Dans une deuxième étape, une fois les hypothèses clairement formulées, l'institution peut donner le reste des données pour formellement tester les hypothèses (Thibault et al., 2023).

Le rapport enregistré

Le rapport enregistré (Nosek & Lakens, 2014) est très similaire au préenregistrement, mais c'est une version encore plus rigoureuse, c'est donc la bonne pratique la plus valorisée et primée : le « gold standard » de la science ouverte. Cependant, ce ne sont pas tous les revues scientifiques qui offrent cette possibilité à l'heure actuelle. La différence avec le préenregistrement, c'est que le rapport enregistré vous demande d'écrire d'abord votre introduction et votre méthodologie, et puis de soumettre ce manuscrit incomplet à un journal. Le journal enverra donc ce document en révision par les pairs, puis votre manuscrit sera accepté ou rejeté seulement sur la pertinence théorique et la force méthodologique de l'étude, peu importe les résultats, qu'ils soient significatifs ou non. Souvent, votre protocole peut être accepté à condition d'y apporter des modifications et des améliorations, ce qui permet au final d'avoir une étude révisée par les pairs en avance et en particulier, encore plus robuste.

Ceci est intéressant car en garantissant la publication a priori, le biais des revues scientifiques de ne publier que les résultats significatifs (effet tiroir) est éliminé à la source (Ferguson & Heene, 2012). Cela étant, bien que le rapport enregistré devienne également de plus



en plus populaire et désirable, il demande une certaine organisation puisqu'il faut réfléchir à beaucoup d'éléments, et notamment rédiger la revue de littérature, avant même de collecter la première donnée! Certes, ceci est bien la manière dont nous devrions théoriquement faire de la science, mais en pratique, ce n'est malheureusement pas aussi fréquent que cela devrait l'être. Bien que de multiples révisions à l'étape de la révision par les pairs pourrait augmenter la qualité de l'étude, cela pourrait également retarder la collecte de données dans certains cas. Il faut donc bien réfléchir aux pour et aux contres de cette stratégie. Ne ressentez donc pas de pression à emprunter cette voie. Vous pourrez néanmoins garder le rapport enregistré en tête comme la pratique à adopter dans un monde idéal.

Le Projet de recherche doctoral

Le Projet de recherche doctoral (PRD) se définit comme suit (Département de psychologie de l'Université du Québec à Montréal, 2020, p. 4) :

Le PRD est un document différent de la thèse qui a pour objectif de préciser le cadre théorique, les objectifs et la méthode de l'étude (ou études) qui sera réalisée comme essai ou thèse de doctorat en psychologie. L'élaboration du document sera faite de manière soutenue, mais non exhaustive, car il s'agit en fait d'annoncer les grandes lignes du projet de recherche à venir. Il faut garder en mémoire que le PRD demeure la proposition d'un « Projet de recherche doctorale ». Ainsi, un PED devrait faire entre 20 et 25 pages, et un PTD devrait faire entre 30 et 35 pages (annexes et références exclues).

Le PRD est complété dans le cadre du cours PSY8331 – Séminaire lié à l'essai ou à la thèse. Ce cours inclut désormais également l'examen doctoral, complété au semestre suivant. La forme du PRD varie en fonction du programme doctoral : il s'agit du Projet de Thèse Doctoral (PTD), pour les étudiants dans les profils recherche (3091) et scientifique-professionnel (3291), ou bien le Projet d'Essai Doctoral (PED), pour les étudiants dans le profil professionnel (3191). Il existe un document fort informatif sur [la politique d'encadrement du PRD](#), disponible sur le site du département de psychologie de l'UQAM.

Le PRD est intéressant comme exercice car il agit un peu comme un rapport enregistré. Comme vous écrivez déjà votre introduction et votre méthodologie dans le contexte de votre PRD, il n'y a donc de là qu'un pas pour soumettre un rapport enregistré, ou du moins, préenregistrer votre recherche, par exemple sur [Open Science Framework](#). Vous êtes donc fortement encouragé à discuter avec votre directeur de l'option de soumettre votre PRD comme rapport enregistré une fois approuvé par votre comité de PRD.

Les données ouvertes

La transparence, la reproductibilité et l'accessibilité deviennent des considérations de plus en plus importantes en ce qui concerne la diffusion de la recherche. À ce titre, les trois Conseils de recherche du Canada (CRSH, CRSNG, IRSC) ont publié une exigence selon laquelle toutes les recherches financées par des fonds publics doivent être disponibles au public, et ce, immédiatement suivant la publication suivant la révision par les pairs. De nombreuses revues exigent également désormais que l'ensemble des données brutes (à l'exception des données sensibles ou identificatoires) soient incluses dans les soumissions de publications, ou bien offrent le badge « Données ouvertes » pour encourager cette pratique. Dans sa [*Déclaration de principes des trois organismes sur la gestion des données numériques*](#), les trois Conseils indiquent aussi spécifiquement de partager les données de recherche sur un dépôt public, et ce, tôt dans le processus (Gouvernement du Canada, Section 3) :



Il faut normalement préserver les données résultant du financement d'un organisme sur une plateforme ou dans un dépôt publiquement accessible, sécurisé et structuré, afin que d'autres chercheurs puissent les trouver et s'en servir.

Pour déterminer si des données doivent être préservées et partagées, les chercheurs devraient examiner quelles données sont nécessaires pour valider les conclusions et les résultats de la recherche et pour répliquer et réutiliser les résultats. Ils devraient se pencher sur les avantages potentiels des données pour leur propre domaine de recherche ou d'autres domaines de recherche et pour la société en général. Ils devraient également déterminer si des obligations éthiques, juridiques ou commerciales interdisent le partage et la préservation des données, et si les données doivent être dépersonnalisées ou mises à disposition avec accès restreint.

Les données doivent être partagées le plus tôt possible au cours du processus de recherche, tandis qu'elles sont considérées comme informatives et de qualité appropriée.

Notez que le simple partage des données n'est pas suffisant si elles restent inaccessibles ou bien incompréhensibles. C'est pourquoi il est important de respecter les [*principes FAIR*](#) (« Findability, Accessibility, Interoperability, and Reusability »), soit qui répondent aux principes de trouvabilité, d'accessibilité, d'interopérabilité et de réutilisation (Wilkinson et al., 2016) et de [*gestion des données de recherche*](#). Pour ce faire, il est possible de mettre les données brutes désidentifiées et métadonnées sur une banque de données publiques telle que le [*Open*](#)

[Science Framework \(OSF\)](#). Il suffit de créer un projet OSF pour son étude, puis de choisir un serveur canadien (Montréal), avant de téléverser les données, et d'ouvrir le projet au grand public. Il est même possible d'obtenir un identifiant doi, pour ce jeu de données, si désiré.

Pour les données ouvertes, même s'il est possible de demander au comité éthique une approbation a posteriori (après la collecte de données) pour mettre les données désidentifiées en ligne, cela n'est pas idéal, puisque les personnes participantes à l'étude n'ont techniquement pas donné leur accord explicite pour une telle utilisation des données. Plutôt, nous vous invitons à être proactifs et à inclure une section sur le partage public des données dans vos formulaires de consentement futurs. Voici un exemple que vous pouvez réutiliser dans vos propres demandes (qui a initialement été suggéré en anglais par le comité éthique de l'Université McGill). Version longue :

UTILISATION SECONDAIRE ET DIFFUSION PUBLIQUE DES DONNÉES

Les organismes subventionnaires et les éditeurs demandent souvent aux chercheurs de rendre leurs données de recherche accessibles à d'autres chercheurs à la fin de leur étude. Rendre les données de recherche accessibles à d'autres permet aux chercheurs qualifiés de reproduire les découvertes scientifiques et stimule l'exploration des jeux de données existants. Ils pourraient notamment utiliser les données pour répondre à des questions différentes de celles examinées dans cette étude. Conformément à ces exigences, nous rendrons les données disponibles via une base de données entièrement ouverte et publique. Pour garantir la confidentialité, les données partagées seront dépouillées de toute information susceptible d'identifier les participants.

Version courte :

UTILISATION SECONDAIRE ET DIFFUSION PUBLIQUE DES DONNÉES

Nous rendrons les données disponibles via une base de données entièrement ouverte et publique afin que d'autres chercheurs puissent reproduire et étendre cette recherche (aucune information ne vous identifiera).

Note sur la protection de la confidentialité

Malgré les recommandations ci-haut, *la priorité devrait toujours être de protéger la confidentialité des personnes participantes*. Les données démographiques, si elles ne sont pas nécessaires à vos analyses, n'ont pas besoin d'être téléversées dans la banque de données, puisqu'elles pourraient potentiellement augmenter le risque de réidentification. Ne partagez donc jamais de données identificatoires telles que le code postal, le numéro de téléphone, le courriel, l'adresse, l'adresse IP, ou bien encore le nom ou les initiales de l'individu.

Le même principe s'applique aux autres types de données sensibles : si par exemple des données qualitatives ou de personnes participantes d'une communauté très ciblée permettaient d'identifier les personnes participantes individuellement, celles-ci ne devraient pas être partagées (à moins d'avoir le consentement explicite des personnes participantes à cet égard). Autrement, si elles sont partagées, elles devraient être désidentifiées de sorte que l'identification ne soit pas possible.

Il existe de bonnes lignes directrices sur la question du partage éthique des données, par exemple voir Meyer (2018). De plus, dans l'incertitude, vérifiez toujours avec votre direction ainsi qu'avec le comité éthique, ce que vous pouvez partager de manière sécuritaire, et ce qui s'applique à votre situation. Enfin, il se peut également que pour certains types d'étude (longitudinale ou autre), le comité d'éthique ne vous autorise pas à partager les données, ou bien qu'il soit nécessaire pour les autres équipes de recherche souhaitant accéder aux données de passer par un processus d'évaluation scientifique et éthique très strict.

Les banques de données synthétiques

Notez que dans ces cas il est toujours possible de partager une banque de données *synthétique*, c'est-à-dire des données artificielles qui ont été générées à partir des données originales. Comme ces données générées ont les mêmes propriétés, elles peuvent être utilisées pour répliquer les analyses et résultats de recherche. Cependant, ce sont des données différentes, elles ne permettent donc pas d'identifier un individu en particulier, puisqu'aucun point de donnée ne correspond à une véritable personne.

Cette stratégie est utilisée notamment en médecine pour protéger la confidentialité des données sensibles (Kokosi & Harron, 2022). Il existe un package R permettant de générer des données synthétiques de cette manière, le package *synthpop* (Nowok et al., 2016):

<https://www.synthpop.org.uk/>. Pour ceux n'utilisant pas le logiciel R, il existe aussi une version « Shiny App » disponible sur internet : <https://synthpop.shinyapps.io/synthpop/>.

Les matériels ouverts

Les matériels ouverts sont les items et questionnaires utilisés (s'ils ne sont pas déjà fournis dans l'article publié), ou encore le code requis pour certaines tâches cognitives ou expérimentales. En général, il n'est nécessaire de fournir que les nouvelles échelles qui n'ont pas été publiées ailleurs, à moins que celles-ci aient été modifiées. Si les échelles ont déjà été publiées ailleurs, il ne suffit que de donner la référence pertinente (en particulier dans les cas de droits d'auteurs, il faut s'assurer d'avoir les autorisations nécessaires avant de les publier publiquement).



Les matériels ouverts incluent également de rendre disponible sa syntaxe (le script) d'analyse, afin que d'autres chercheurs puissent répliquer exactement vos résultats en utilisant votre propre script (et vos données ouvertes). Ceci permet notamment de s'assurer que vos résultats sont répliquables en soi et d'identifier de possibles erreurs d'analyse. Si les résultats sont répliquables à cette étape, nous saurons que de possibles difficultés de réplication futures (dans d'autres études) ne seront probablement pas dû à des erreurs dans le script ou dans la manière dont les résultats ont été analysés. Cela nous amène au point suivant : le choix du logiciel pour partager sa syntaxe.

Il est recommandé de partager les matériels (script, questionnaires, tâches) sur OSF, directement dans le projet où vous téléverserez également les données. Si vous avez préenregistré votre étude, votre préenregistrement sera également lié à ce projet. Les journaux offrent généralement une manière de soumettre un document de matériel supplémentaire. Cependant, il est préférable de téléverser ce document sur le projet OSF approprié et de référer au lien pertinent dans l'article puisque cela donne beaucoup plus de flexibilité. Par exemple, les journaux ne permettent généralement pas d'apporter de modifications à leur document maison de matériel supplémentaire, même en cas d'erreur. Pourtant, les erreurs dans ces types de document sont normales et relativement fréquentes, et celles-ci devraient pouvoir être corrigées. En utilisant OSF, vous pouvez rapidement corriger l'erreur dès que vous la constatez, sans devoir attendre une autorisation quelconque, ou même ajouter de l'information, lorsque pertinent. Vous pouvez en quelque sorte considérer le projet OSF comme un espace de stockage en ligne comme Dropbox ou OneDrive, mais gratuit, ce qui le rend très pratique.

L'utilisation de logiciels libres

Il y a plusieurs bénéfices à utiliser des logiciels libres (« open source ») en science, par exemple pour l'analyse ou bien la collecte des données. Bien sûr, le plus important est de partager sa syntaxe, peu importe le logiciel utilisé, que cela soit, par exemple pour l'analyse, SPSS, JASP, SAS, MATLAB, Mplus, AMOS, ou R. Cela étant dit, certains logiciels sont plus compatibles avec les idéaux de la science ouverte, alors que d'autres le sont moins (Quintana, 2020). Premièrement, les logiciels avec des licences payantes rendent le processus difficile voire impossible pour les personnes n'ayant pas les fonds pour payer ces licences qui sont pour la plupart très onéreuses. Au Québec, la plupart des universités paient les licences de plusieurs logiciels statistiques pour l'ensemble de leur communauté (p. ex., SPSS, SAS, parfois Mplus). Par contre, ce n'est pas le cas partout dans le monde (p. ex., en Europe et dans les pays à voie de développement). Ces personnes situées hors du Québec n'ont donc pas licence SPSS, et pourraient ainsi, malgré toute leur bonne volonté, être incapables de répliquer ou répéter vos analyses. Si vous-même vous graduez et n'avez plus accès à la licence de l'UQAM, vous pourriez ne plus être en mesure d'effectuer vos analyses à nouveau. Par exemple, si vous publiez votre article scientifique après avoir gradué, il est possible que les personnes évaluatrices vous demandent de modifier vos analyses. Si votre licence est expirée, cette tâche peut aisément devenir beaucoup plus compliquée que prévu. Cependant, les logiciels libres d'accès comme R (R Core Team, 2022) resteront toujours gratuits et disponibles partout dans le monde.



Deuxièmement, plusieurs des logiciels à licence cachent également leur code source, le phénomène de « la boîte noire », ce qui crée de l'incertitude sur par exemple les formules ou les algorithmes utilisés. Les packages R ont tous leur code disponible au grand public. Ainsi, n'importe qui peut aller vérifier comment une valeur précise a été calculée, ou bien quels paramètres ou algorithmes ont été utilisés. Il y a un avantage supplémentaire à cela. Si un package R n'existe pas encore mais qu'il y a un besoin réel, n'importe qui peut écrire un nouveau package R, amener des améliorations à un package R existant, ou bien simplement écrire une nouvelle fonction R. De plus, il est possible de partager le tout avec les autres, ce qui contribue à créer une communauté vivante et sans cesse grandissante. En ce qui concerne les logiciels nécessitant une licence, il faut souvent patienter pour que la compagnie publie une

nouvelle version avec les nouvelles fonctionnalités, et ce n'est pas garantie que la mise à jour inclura les fonctionnalités désirées.

Troisièmement, le logiciel R permet de réduire les erreurs de copier-coller typiques des autres logiciels. Les erreurs dans le rapport des valeurs statistiques sont en effet assez fréquentes en psychologie; selon certaines estimations, jusqu'à 50% des articles ont au moins une erreur statistique (Nuijten et al., 2016). Le logiciel R permet ainsi d'exporter les résultats (sous forme de texte (e.g., le package « [report](#) » de easystats) ou de tableaux directement dans Microsoft Word ou Microsoft Excel (e.g., le package [rempsyc](#)). Il permet également de vérifier s'il existe des erreurs statistiques évidentes directement dans le PDF de votre article final (via le package [statcheck](#)). Note à part pour les artistes, puisque les possibilités sont infinies, c'est aussi le logiciel qui permet de faire les plus belles figures pour visualiser vos données et vos résultats!

Quatrièmement, les scripts R permettent d'agir un peu comme des tutoriels en eux-mêmes, puisqu'ils permettent notamment de bien intégrer les commentaires de l'analyste au sein de la syntaxe (via des scripts R normaux ou bien via *rmarkdown*). Puisque la communauté R est vivante, engagée, et grandissante, il existe également de plus en plus de tutoriels et de packages en ligne pour réaliser à peu près n'importe quelle opération ou analyse désirée. Ceux-ci permettent notamment de guider les gens qui liront vos scripts à travers le processus complet d'analyse et des bonnes pratiques, à chaque étape du processus.

En effet, même si quelqu'un partage sa syntaxe SPSS, ce sera probablement pour l'analyse en tant que telle, et non pas pour toutes les étapes préalables de nettoyage, de préparation, et de transformation des données. Le logiciel R quant à lui permet de partager l'ensemble des démarches effectuées lors de l'analyse, incluant les tests de puissance, le nettoyage des données, l'exploration et la gestion des données manquantes, le calcul des moyennes, la vérification des postulats statistiques, la transformation et/ou la standardisation des données, le rapport des données descriptives ou démographiques, les analyses statistiques en tant que tel, et même la production des figures et de tableaux.

L'apprentissage et l'utilisation du logiciel R a donc [une place spéciale](#) dans le monde de la science ouverte. Bien sûr, la courbe d'apprentissage de R est plus grande que celle d'autres logiciels, puisque l'interface (RStudio) utilise principalement la syntaxe (programmation). Cependant, il s'agit d'un [investissement à long-terme](#) qui en vaudra certainement la chandelle.

Est-il nécessaire d'apprendre à programmer?

En effet, le langage R est techniquement un langage de programmation, bien que de haut niveau, et la courbe d'apprentissage peut s'avérer ardue au départ. Gardez en tête que comme de nombreuses autres bonnes pratiques de ce guide, ce ne sont pas des obligations, mais des recommandations. En ce sens, cet investissement de temps pourrait être plus profitable et réaliste pour les étudiants du profil recherche (3091) ou scientifique-professionnel (3291), que pour les gens du profil professionnel (3191).

Bon à savoir : les cours de statistiques obligatoires au doctorat en psychologie à l'UQAM sont parfois enseignés exclusivement via l'utilisation du programme R. Ainsi, si cela vous intéresse, assurez-vous de prendre le cours PSY7102 - Techniques d'analyse quantitative de données I lorsqu'il est enseigné par Pier-Olivier Caron, et PSY7105 - Techniques d'analyse quantitative de données II lorsqu'il est enseigné par Marina Doucerain ou Pier-Olivier Caron.

La prépublication

La prépublication (anglais : preprint) est l'idée de rendre la version originale du manuscrit (en PDF, mais formaté dans un logiciel de traitement de texte comme Microsoft Word) disponible dans une archive publique en ligne, un peu comme on le fait pour ses données ou son script d'analyse sur OSF. OSF collabore notamment avec plusieurs plateformes d'archives de prépublications, qui varient en fonction des disciplines. Pour la psychologie, le serveur de prépublication le plus populaire est [PsyArXiv](#) (prononcé en anglais « Psy Archive »), fondé par la Society for the Improvement of Psychological Science, en collaboration avec OSF.

Pourquoi donc devrions-nous publier en ligne une version de notre manuscrit qui n'a toujours pas été révisée par des pairs? Il y a plusieurs raisons et bénéfices à cela. D'un, cela accélère la diffusion des résultats de recherche (Hoy, 2020). En effet, avant d'obtenir la version finale, publiée et formatée par le journal, les chercheurs doivent attendre plusieurs mois, voire années. À chaque étape du processus s'accumule en effet des délais : de l'évaluation initiale du manuscrit par l'éditrice ou l'éditeur au formatage final de l'article, en passant par de possibles



multiples révisions par les pairs. Ce faisant, les équipes de recherche qui s'appuient sur les articles publiés pour par exemple faire une revue de littérature sont nécessairement plusieurs mois ou années en retard sur ce qui se fait de plus récent dans le domaine. En rendant la prépublication par les pairs disponible en ligne, elle est immédiatement disponible, et les autres équipes de recherche peuvent déjà la consulter et potentiellement, la citer. Cela augmente la rapidité, la productivité, et l'harmonisation de la recherche.

De deux, même une fois l'article accepté et publié par un journal, celui-ci n'est peut-être pas disponible en libre accès, ce qui crée de nouvelles barrières pour l'accès aux résultats de ces recherches. En mettant disponible une prépublication en ligne, même si l'article officiel est derrière un mur payant, le public et les autres équipes de recherche peuvent quand même accéder essentiellement au même article, mais sans le formatage propre au journal. Ceci permet de rendre sa recherche plus accessible et découvrable. Ce bénéfice d'accessibilité et de meilleure visibilité se traduit également par un plus grand impact de l'article : plus haut score Altmetric et plus de citations (Fraser et al., 2020; Fu & Hughey, 2019; Serghiou & Ioannidis, 2018).

De trois, si l'article n'est au final pas publié à cause de résultats non-significatifs, l'article reste tout de même public et découvrable en ligne, ce qui permet de réduire l'effet tiroir (« file drawer effect » en anglais), soit le fait de ne publier que les résultats significatifs (alors que les résultats non-significatifs sont également d'une importance critique pour faire progresser la théorie).

Est-ce qu'une prépublication peut prévenir la publication éventuelle?

La vaste majorité des journaux acceptent qu'un article ait été prépublié, car cela n'est pas considéré comme une publication scientifique en soi. Cependant, dans de très rares cas, il se pourrait que certaines revues n'acceptent pas les prépublications. Il est recommandé de vérifier avec les revues directement s'il y a quelque inquiétude à ce niveau-là. Bien que cela soit du cas-par-cas, certaines personnes sont d'avis qu'il est préférable de tout simplement ne pas publier dans les revues qui n'acceptent pas les prépublications. Pour une liste de questions fréquemment posées sur la prépublication, voir : <https://help.osf.io/article/230-preprint-faqs>.

De quatre, la prépublication donne une preuve publique qui crédite l'équipe de recherche originale avec une date bien précise pour la contribution originale, en termes d'idée de recherche, de théorie, de méthode, ou bien d'approche. Les prépublications ont également un identifiant DOI qui leur permet d'être citées comme document à part entière, bien que la pratique courante soit de référer à l'article publié lorsqu'il est accepté dans une revue scientifique. Ainsi, les équipes de

recherche qui pourraient craindre de se faire voler leurs idées en mettant la prépublication publique n'ont pas à s'en faire puisqu'ils auront une preuve tangible de la date à laquelle ils ont généré leurs idées. Sans prépublication, quelqu'un d'autre pourrait publier la même idée avant (considérant le délai de plusieurs mois ou années du processus de publication).

Enfin, la prépublication est également une bonne opportunité d'avoir de la rétroaction supplémentaire sur l'article de d'autres membres de la communauté, et d'améliorer le manuscrit avant la soumission au prochain journal, ce qui augmente globalement la qualité du manuscrit (et potentiellement des analyses, le cas échéant) et peut-être même les chances de publication. D'ailleurs, [selon certains](#), un des avantages supplémentaires pour les étudiants est que cela a plus de poids lors des applications aux bourses, subventions, ou demandes d'emploi, puisque le comité d'évaluation peut directement aller consulter le document en question (grâce au lien doi) pour juger de la qualité du travail, plutôt que de devoir se baser sur une simple mention sur le CV (typiquement : « manuscrit en cours » ou bien « manuscrit soumis pour publication » sans grande information supplémentaire).

Est-ce que la prépublication remplace le processus de publication traditionnel?

Non, la prépublication ne remplace pas le processus de publication traditionnel. La révision scientifique par les pairs est fondamentale au processus scientifique. Il existe souvent des différences considérables en termes d'introduction, de méthodes, d'analyses, de résultats, de discussion et de conclusion entre un manuscrit avant l'examen par les pairs et sa version finale publiée.

De plus, il existe également un autre type de document qu'il est possible de partager, la soi-disant postpublication (« postprint »), qui consiste à mettre la version du manuscrit qui intègre les révisions par les pairs, mais avant le formatage en PDF par le journal (si le journal l'autorise). Pour la postpublication néanmoins, il faut se renseigner auprès du journal, car il y a parfois un embargo sur la publication de cette version, c'est-à-dire qu'il faut attendre un certain temps (par exemple, un an) avant de pouvoir la partager publiquement et gratuitement. Voir la section suivante à cet effet.

La publication en libre accès

La dernière étape dans le processus de la recherche est typiquement la publication de l'article final révisé par les pairs. C'est à cette étape que la publication *en libre accès* entre en jeu. Celle-ci consiste simplement à publier un article dans une revue scientifique de sorte que le grand public n'ait aucun frais à payer pour pouvoir lire ou consulter l'article. Comme pour la prépublication, cela augmente l'accessibilité et la découvrabilité de la recherche. Cela peut parfois aussi se traduire en un plus grand nombre de citations ou impact (Holmberg et al., 2020; Langham-Putrow et al., 2021). Les Fonds de recherche du Québec (FRQ) [soutiennent aussi la science ouverte](#). En 2021, ils ont notamment adhéré au [Plan S](#), qui vise à favoriser un accès libre et immédiat des publications scientifiques. Par conséquent, depuis 2021, toute personne recevant une bourse ou une subvention des FRQ doit rendre immédiatement disponible publiquement toute publication révisée par les pairs découlant de ce financement (Fonds de recherche du Québec, 2022).



Quoi faire en cas de politiques conflictuelles?

Que faire si les politiques de la revue ne permettent pas le partage de l'article, donc en cas de conflit entre les politiques de la revue et celles de l'organisme subventionnaire? Vous avez alors principalement deux options. Option 1 : vous soumettez l'article à une revue différente, qui permet le partage de l'article, en respect des politiques de l'organisme subventionnaire. Option 2 : vous faites fi des politiques de l'organisme subventionnaire et soumettez quand même à une revue qui ne permet pas le partage, et suivez les politiques de la revue, donc ne partagez pas l'article. Le choix est le vôtre, mais nous recommandons la première option pour être conforme aux politiques des organismes subventionnaires sans lesquelles cette recherche ne serait pas possible.

Il existe [plusieurs types de publication en libre accès](#), catégorisés selon un code de couleur—les plus connus : vert, doré, et diamant. La publication en libre accès « vert » permet le partage libre du manuscrit final accepté (mais pas formaté), sans frais. La publication en libre accès « doré » permet le partage libre du manuscrit final accepté (et formaté), moyennant des frais de publication. Ces deux variantes de l'accès libre sont les plus populaires, mais une troisième variante gagne de plus en plus en popularité également ([notamment grâce au Plan S](#)) : la publication en libre accès « diamant », qui permet le partage libre du manuscrit final accepté (et formaté), sans frais pour les auteurs ni pour les institutions. Il n'y a donc pas de frais d'abonnement à la revue pour ce type de libre accès : les frais de fonctionnement de ces revues sont couverts directement par des organismes subventionnaires, des universités, ou des sociétés professionnelles par exemple. Il est généralement possible de négocier avec le journal pour obtenir l'autorisation de publier le manuscrit final accepté, mais pas formaté (libre accès vert).

Attention aux revues prédatrices!

Les revues prédatrices adoptent un modèle de publication en libre accès, mais seulement en apparence, car le processus d'évaluation scientifique est faible voir nul. En effet, celles-ci chargent généralement des frais de publication élevés, mais il n'y a pas de réel processus de révision par les pairs, ou bien celui-ci est tout à fait inadéquat.

Les Fonds de recherche du Québec nous rappellent qu'il faut [faire attention à ne pas publier dans celles-ci](#). Alors que certaines personnes pourraient être tentées de le faire, concernant un article qui peine à être accepté, publier dans des revues prédatrices risque plus de nuire à sa propre carrière qu'autre chose. Dans un tel cas, considérez plutôt la prépublication en attendant de trouver un medium de publication approprié.

Ainsi, si vous recevez une invitation par courriel à publier dans un journal, c'est presque certainement une revue prédatrice. Sauf exceptions, ce sont les auteurs de l'article qui doivent eux-mêmes soumettre leur manuscrit dans une revue (un exemple d'exception serait une invitation à contribuer à un numéro spécial, mais ces cas peuvent être reconnus facilement). Un principe similaire s'applique aux invitations à présenter à des congrès, souvent à des endroits exotiques. Ces invitations sont presque toujours des arnaques et ce ne sont pas des événements scientifiques sérieux. Dans le doute, les bibliothécaires sont certainement bien placées pour vous informer sur la légitimité des différentes revues scientifiques.

L'UQAM possède une documentation fort pertinente pour reconnaître les revues prédatrices : <https://uqam-ca.libguides.com/editeurs-predateurs/>.

Les bénéfices (et limites) d’embrasser la science ouverte

Il est à noter que toutes les équipes de recherche ne sont pas d’avis que la science ouverte et une plus grande rigueur sont la solution à tout; certaines pensent même que cela vient nécessairement avec certains désavantages ou compromis (Fiedler et al., 2012; Rubin, 2020; Rubin & Donkin, 2022; Schwartz, 2023; Wiggins & Christopherson, 2019). À ce titre, il est clair que la recherche exploratoire a une place fondamentale à jouer en science, à condition qu’elle soit conduite avec un esprit de transparence.

Somme toute, la science ouverte et une plus grande rigueur viennent également avec de nombreux bénéfices pour le progrès scientifique et pour vous personnellement. Pour votre direction : votre recherche aura beaucoup plus de crédibilité, et possiblement plus de citations et d’impact; vos candidatures lors des concours de bourse ou de subvention pourraient être avantagées si votre engagement dans la science ouverte est reconnu. Les compétences que vous développerez vous rendront également une ou un candidat beaucoup plus intéressant sur le marché du travail puisqu’elles sont de plus en plus valorisées et reconnues. Pour votre futur : il sera plus facile de retrouver et partager des documents importants comme vos fichiers finaux de données et d’analyse lors du processus de publication de vos articles ou lors de demandes extérieures (de vos collègues ou collaboratrices et collaborateurs) puisque vous aurez tout décrit et organisé de manière transparente. Pour la science : avec une science plus rigoureuse, répliquable, et transparente, tout le monde y gagne! Cela aide à réduire les faux positifs ainsi que les faux négatifs, à faire progresser la théorie, réduit le risque que les autres équipes de recherche perdent leur temps et leur argent à tenter de répliquer des études qui ont des résultats qui ne sont pas fiables, augmente la productivité des équipes qui aimeraient faire des réplifications exactes ou conceptuelles de vos recherches, et pourrait même servir de tutoriels pour vos futures collègues ou étudiantes et étudiants!

OK, et maintenant?

Vous venez de terminer la lecture de ce document et vous vous demandez par où commencer? Commencez simplement. Vous ne devez pas nécessairement appliquer toutes les recommandations proposées dans ce guide dès maintenant, puisqu'il s'agit d'un processus d'apprentissage continu. L'important, c'est d'y aller à son rythme et de commencer à intégrer progressivement ces idées dans sa recherche. En y allant étape par étape, cela sera moins intimidant et vous verrez éventuellement un progrès réel. Et le plus important : n'hésitez pas à demander de l'aide! Cela rendra la tâche plus facile et motivante. Entretemps, voici quelques suggestions :

1. Créez-vous un compte OSF et remplissez votre profil! Vous pourrez ensuite commencer à accumuler des « points » OSF pour chacune des actions que vous poserez sur la plateforme. <https://osf.io/>
2. Si applicable à votre situation, vous pourrez ensuite (a) y créer un projet associé à votre projet de recherche, (b) inviter vos collaboratrices et collaborateurs à se joindre au projet; (c) téléverser vos données désidentifiées dans ce projet, (d) préenregistrer votre étude à venir, ou même (e) y téléverser votre prépublication (preprint).
3. Considérez aussi rejoindre le *Psychological Science Accelerator*, ce réseau collaboratif à grande échelle de recherche en psychologie! <https://psysciacc.org/>
4. Pour des diapositives brèves, mais très informatives et stimulantes, consultez ce document « Five things about open and reproducible science that every early career researcher should know » : <https://osf.io/2jt9u/>
5. Pour un guide détaillé de bonnes pratiques en recherche de l'Université de Guelph, veuillez consultez le document suivant : <https://www.uoguelph.ca/psychology/graduate/thesis-statistics>

Bonne recherche!

« Checklist » à partager à votre direction, comité, ou collègues

L'étudiante ou l'étudiant confirme qu'il va s'engager dans les pratiques suivantes :

Bonne pratique intégrée?	Oui	Non	(Si non, justification)
Justifier sa taille d'échantillon (p. ex., via une analyse de puissance pour évaluer la taille d'échantillon requise a priori)			
Déterminer a priori : comment gérer les inclusions, exclusions, données manquantes, données aberrantes, assomptions statistiques, standardisation, opérationnalisation des variables, choix des tests statistiques et de leurs paramètres, corrections pour comparaisons multiples, seuil de signification			
Préenregistrement de son étude ou soumission d'un rapport enregistré			
Mise en ligne de ses données désidentifiées (p. ex., sur OSF)			
Mise en ligne de ses matériels et de son script d'analyse (p. ex., sur OSF)			
Inclusion des tailles d'effet et de leurs intervalles de confiance			
Vérification des statistiques de l'article avec http://statcheck.io/			
Soumission d'une prépublication (p. ex., sur OSF ou PsyArXiv)			
Soumission à des revues en libre accès			

L'étudiante ou l'étudiant confirme qu'il ne va pas s'engager dans les pratiques suivantes :

Pratique questionnable	Oui	Non	(Si oui, justification)
« <i>p-hacking</i> » (en général) en déterminant <i>ad hoc</i> (plutôt qu' <i>a priori</i>) les décisions d'analyse à venir (voir tableau précédent pour les exemples)			
Changer le rôle des variables <i>ad hoc</i> (covariable, médiateur, modérateur, etc.)			
Rapporter que certaines variables (plutôt que toutes)			
Rapporter que certains groupes ou sous-groupes expérimentaux (plutôt que tous)			
Rapporter que certaines analyses effectuées (plutôt que toutes)			
Rapporter que certaines études complétées (plutôt que toutes)			
Changer ses hypothèses en fonction des résultats (« HARKing »)			
Utiliser des échantillons trop petits ou inappropriés pour les méthodes d'analyses choisies (ou vice-versa; trop faible puissance statistique)			
Rester vague sur si les analyses sont confirmatoires ou exploratoires (plutôt que le communiquer explicitement)			

Remerciements

L’auteur aimerait remercier Liesette Brunson, PhD, Mara Brendgen, PhD, Catherine Herba, PhD, ainsi qu’Andreea Gavrilă, Jany St-Cyr, Catherine Cimon-Paquet, Benjamin Chabot, Gabriel Pelletier, Valérie Lapointe, et Charles-Étienne Lavoie pour des rétroactions fort utiles sur ce document.

Références

- Al-Aidroos, N., Fiacconi, C., Powell, D., Marmurek, H., Newby-Clark, I., Spence, J., Stanley, D., & Tric, L. (2018). *Statistical methods in theses: Guidelines and explanations*.
<https://www.uoguelph.ca/psychology/graduate/thesis-statistics>
- Baldwin, J. R., Pingault, J.-B., Schoeler, T., Sallis, H. M., & Munafò, M. R. (2022). Protecting against researcher bias in secondary data analysis: Challenges and potential solutions. *European Journal of Epidemiology*, 37(1), 1-10. <https://doi.org/10.1007/s10654-021-00839-0>
- Camerer, C. F., Dreber, A., Holzmeister, F., Ho, T.-H., Huber, J., Johannesson, M., Kirchler, M., Nave, G., Nosek, B. A., Pfeiffer, T., Altmejd, A., Buttrick, N., Chan, T., Chen, Y., Forsell, E., Gampa, A., Heikensten, E., Hummer, L., Imai, T., Isaksson, S., Manfredi, D., Rose, J., Wagenmakers, E.-J., & Wu, H. (2018). Evaluating the replicability of social science experiments in nature and science between 2010 and 2015. *Nature Human Behaviour*, 2(9), 637-644. <https://doi.org/10.1038/s41562-018-0399-z>
- Chambers, C. (2018). *Strategies for improving study robustness in controversial research*.
<https://osf.io/v7ndz>

- Département de psychologie de l'Université du Québec à Montréal. (2020). *Politique d'encadrement du projet de recherche doctorale (PRD)*. https://psychologie.uqam.ca/wp-content/uploads/sites/73/2021/01/Prd_-final-adopte%CC%81-copie.pdf
- Ferguson, C. J., & Heene, M. (2012). A vast graveyard of undead theories: Publication bias and psychological science's aversion to the null. *Perspectives on Psychological Science*, 7(6), 555-561. <https://doi.org/10.1177/1745691612459059>
- Fiedler, K., Kutzner, F., & Krueger, J. I. (2012). The long way from α -error control to validity proper: Problems with a short-sighted false-positive debate. *Perspectives on Psychological Science*, 7(6), 661-669. <https://doi.org/10.1177/1745691612462587>
- Fonds de recherche du Québec. (2022). *Politique de diffusion en libre accès (version révisée)*. https://frq.gouv.qc.ca/app/uploads/2022/06/politique-libre-acces-revisee_vf.pdf
- Fraser, N., Momeni, F., Mayr, P., & Peters, I. (2020). The relationship between biorxiv preprints, citations and altmetrics. *Quantitative Science Studies*, 1(2), 618-638. https://doi.org/10.1162/qss_a_00043
- Fu, D. Y., & Hughey, J. J. (2019). Releasing a preprint is associated with more attention and citations for the peer-reviewed article. *eLife*, 8, e52646. <https://doi.org/10.7554/eLife.52646>
- Goldacre, B. (2009). *Bad science: Quacks, hacks, and big pharma flacks*. Fourth Estate.
- Gouvernement du Canada. *Déclaration de principes des trois organismes sur la gestion des données numériques*. <https://science.gc.ca/site/science/fr/financement-interorganismes-recherche/politiques-lignes-directrices/gestion-donnees-recherche/declaration-principes-trois-organismes-gestion-donnees-numeriques>

- Haefel, G. J. (2022). Psychology needs to get tired of winning. *Royal Society Open Science*, 9(6), 220099. <https://doi.org/doi:10.1098/rsos.220099>
- Holmberg, K., Hedman, J., Bowman, T. D., Didegah, F., & Laakso, M. (2020). Do articles in open access journals have more frequent altmetric activity than articles in subscription-based journals? An investigation of the research output of Finnish universities. *Scientometrics*, 122(1), 645-659. <https://doi.org/10.1007/s11192-019-03301-x>
- Hoy, M. B. (2020). Rise of the Rxivs: How preprint servers are changing the publishing process. *Medical Reference Services Quarterly*, 39(1), 84-89. <https://doi.org/10.1080/02763869.2020.1704597>
- Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS Medicine*, 2(8), e124. <https://doi.org/10.1371/journal.pmed.0020124>
- Kokosi, T., & Harron, K. (2022). Synthetic data in medical research. *BMJ Medicine*, 1(1), e000167. <https://doi.org/10.1136/bmjmed-2022-000167>
- Langham-Putrow, A., Bakker, C., & Riegelman, A. (2021). Is the open access citation advantage real? A systematic review of the citation of open access and subscription-based articles. *PloS One*, 16(6), e0253129. <https://doi.org/10.1371/journal.pone.0253129>
- Meyer, M. N. (2018). Practical tips for ethical data sharing. *Advances in Methods and Practices in Psychological Science*, 1(1), 131-144. <https://doi.org/10.1177/2515245917747656>
- Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences*, 115(11), 2600-2606. <https://doi.org/doi:10.1073/pnas.1708274114>

- Nosek, B. A., & Lakens, D. (2014). Registered reports: A method to increase the credibility of published results. *Social Psychology*, 45, 137-141. <https://doi.org/10.1027/1864-9335/a000192>
- Nowok, B., Raab, G. M., & Dibben, C. (2016). Synthpop: Bespoke creation of synthetic data in R. *Journal of Statistical Software*, 74(11), 1 - 26. <https://doi.org/10.18637/jss.v074.i11>
- Nuijten, M. B., Hartgerink, C. H. J., van Assen, M. A. L. M., Epskamp, S., & Wicherts, J. M. (2016). The prevalence of statistical reporting errors in psychology (1985–2013). *Behavior Research Methods*, 48(4), 1205-1226. <https://doi.org/10.3758/s13428-015-0664-2>
- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251), aac4716. <https://doi.org/doi:10.1126/science.aac4716>
- Quintana, D. S. (2020). *Five things about open and reproducible science that every early career researcher should know*. <https://osf.io/2jt9u>
- R Core Team. (2022). R: A language and environment for statistical computing (version 4.2.1) [Computer software]. *R Foundation for Statistical Computing*, Vienna, Austria. <https://www.R-project.org/>
- Rawat, S., & Meena, S. (2014). Publish or perish: Where are we heading? *Journal of Research in Medical Sciences*, 19(2), 87-89.
- Rubin, M. (2020). Does preregistration improve the credibility of research findings? *The Quantitative Methods for Psychology*, 16(4), 376–390. <https://doi.org/10.20982/tqmp.16.4.p376>

- Rubin, M., & Donkin, C. (2022). Exploratory hypothesis tests can be more compelling than confirmatory hypothesis tests. *Philosophical Psychology*, 1-29.
<https://doi.org/10.1080/09515089.2022.2113771>
- Sarewitz, D. (2016). The pressure to publish pushes down quality. *Nature*, 533(7602), 147-147.
<https://doi.org/10.1038/533147a>
- Schwab, S., Janiaud, P., Dayan, M., Amrhein, V., Panczak, R., Palagi, P. M., Hemkens, L. G., Ramon, M., Rothen, N., Senn, S., Furrer, E., & Held, L. (2022). Ten simple rules for good research practice. *PLoS Computational Biology*, 18(6), e1010139.
<https://doi.org/10.1371/journal.pcbi.1010139>
- Schwartz, B. (2023). Psychology's increased rigor is good news. But is it only good news? *Behavioral Scientist*. <https://behavioralscientist.org/psychologys-increased-rigor-is-good-news-but-is-it-only-good-news/>
- Serghiou, S., & Ioannidis, J. P. A. (2018). Altmetric scores, citations, and publication of studies posted as preprints. *JAMA*, 319(4), 402-404. <https://doi.org/10.1001/jama.2017.21168>
- Thibault, R. T., Kovacs, M., Hardwicke, T. E., Sarafoglou, A., & Munafo, M. (2023). Reducing bias in secondary data analysis via an Explore and Confirm Analysis Workflow (ECAW): A proposal and survey of observational researchers.
<https://doi.org/10.31222/osf.io/md2xz>
- Wagenmakers, E.-J., Wetzels, R., Borsboom, D., van der Maas, H. L. J., & Kievit, R. A. (2012). An agenda for purely confirmatory research. *Perspectives on Psychological Science*, 7(6), 632-638. <https://doi.org/10.1177/1745691612463078>
- Wicherts, J. M., Veldkamp, C. L. S., Augusteijn, H. E. M., Bakker, M., van Aert, R. C. M., & van Assen, M. A. L. M. (2016). Degrees of freedom in planning, running, analyzing, and

reporting psychological studies: A checklist to avoid p-hacking. *Frontiers in Psychology*,

7. <https://doi.org/10.3389/fpsyg.2016.01832>

Wiggins, B. J., & Christopherson, C. D. (2019). The replication crisis in psychology: An overview for theoretical and philosophical psychology. *Journal of Theoretical and Philosophical Psychology*, 39, 202-217. <https://doi.org/10.1037/teo0000137>

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., Gonzalez-Beltran, A., Gray, A. J. G., Groth, P., Goble, C., Grethe, J. S., Heringa, J., 't Hoen, P. A. C., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S. J., Martone, M. E., Mons, A., Packer, A. L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M. A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., & Mons, B. (2016). The fair guiding principles for scientific data management and stewardship. *Scientific Data*, 3(1), 160018. <https://doi.org/10.1038/sdata.2016.18>