

Introduction au pré-enregistrement d’une recherche expérimentale

Mattia A. Fritz

27/03/2023

Résumé

Ce document introduit la pratique du pré-enregistrement d’une expérience à travers un template adapté de la plateforme *Open Science Framework* (OSF). Le document guide des étudiant-es sans expérience préalable en méthode expérimentale dans la spécification des informations saillantes pour planifier une expérience. Les implications pédagogiques et relatives à la philosophie *Open Science* sont également traitées brièvement.

Introduction

Le pré-enregistrement d’une expérience est une pratique liée à la philosophie de la Science Ouverte – ou *Open Science* en anglais (Nosek et al., 2015; Open Science Collaboration, 2015) – qui consiste à définir et publier dans une plateforme indépendante les détails d’une expérience **avant que les données empiriques soient récoltées**.

Cette pratique augmente la transparence de la recherche et permet de limiter ce qu’on appelle des *Questionable Research Practices* (QRP), c’est-à-dire des pratiques qui se trouvent dans une *zone grise* entre les fraudes et les mauvaises pratiques scientifiques (John et al., 2012; Kerr, 1998; Makel et al., 2019). Parmi les QRP figurent notamment :

- Modifier ou supprimer les hypothèses théoriques et opérationnelles d’une recherche pour *aller* dans la direction des résultats. Ce phénomène est souvent identifié avec l’acronyme HARKing (Kerr, 1998) : *Hypothesizing After the Results are Known*. Par exemple, les chercheurs s’attendaient à l’origine à un effet plus large de l’intervention A par rapport à l’intervention B, mais les données montrent exactement le contraire. À ce moment, les chercheurs modifient leurs hypothèses (y compris la justification scientifique/cadre théorique) pour s’aligner aux résultats ;

-
- Ne pas reporter la présence d’une ou plusieurs manipulations expérimentales qui n’ont pas abouties aux résultats attendus ;
 - Mener plusieurs tests statistiques et reporter dans l’article seulement les plus intéressants, par exemple seulement ceux qui atteignent le seuil conventionnel de *significativité statistique* à $p < 0.05$

Le pré-enregistrement n’exclut cependant pas la possibilité de mener d’autres analyses une fois les données récoltées. La différence consiste dans le fait que le pré-enregistrement *oblige* les chercheurs à signaler la distinction de ce qui a été prévu à priori et ce qui a été fait à posteriori, suite à de l’information extraite directement par les données ou l’expérience elle-même. D’un point de vue méthodologique, le pré-enregistrement force aussi les chercheurs à prévoir à l’avance la plupart des détails saillants pour mener une expérience, même si le *simple* fait de pré-enregistrer une expérience ne définit pas automatiquement une meilleure qualité scientifique. Il existe des expériences de qualité qui n’ont pas fait recours à des QRP même sans pré-enregistrement. Il s’agit d’un mécanisme de transparence qui a une double perspective. D’une part, il vise à protéger les chercheurs de ses propres biais ou de la pression de publier des résultats qui correspondent forcément à des caractéristiques standard, indépendamment de ce qui a été effectivement observé dans l’expérience. De l’autre, il donne la possibilité aux expériences d’être scrutées attentivement, avec un niveau de détail qui n’est souvent pas disponible dans les publications scientifiques. Nelson et collègues résument bien la philosophie du pré-enregistrement de la manière suivante : “preregistrations do not tie researchers’ hands, but merely uncover readers’ eyes” (Nelson et al., 2018, p. 519).

Le pré-enregistrement peut être effectué de manière indépendante ou combiné avec les Registered Reports. Dans ces cas, un journal scientifique s’engage à publier un article sur la base de la qualité du pré-enregistrement et donc indépendamment des résultats qui seront effectivement observés après la phase de récolte et analyse des données.

Il existe différents formats de pré-enregistrement et le pré-enregistrement ne s’applique pas seulement à la méthode expérimentale, mais également aux recherches basées sur des observations ou qui utilisent une méthode *qualitative*.

Dans cette section, nous proposons un format de pré-enregistrement basé principalement sur le template proposé par le Open Science Framework (OSF), dont la finalité pour le niveau introductif de ce cours est d’inciter les étudiant-es à réfléchir sur les différents éléments à considérer avant de mener une expérience.

1 Métadonnées

Les métadonnées permettent de fournir des informations sur le pré-enregistrement.

1.1 Titre

Titre du pré-enregistrement, ceci correspond normalement à un titre au moins similaire à celui envisagé pour l'article/contribution scientifique finale.

1.2 Description

Description du pré-enregistrement, par exemple au niveau de la justification scientifique de l'expérience.

2 Information sur l'étude/expérience

Les informations sur l'étude/expérience concernent principalement les hypothèses. La description de cette partie sur le template OSF est le suivant (traduction libre) :

Énumérez des hypothèses spécifiques, concises et vérifiables. Veuillez indiquer si les hypothèses sont directionnelles ou non directionnelles. Si elles sont directionnelles, indiquez la direction. Un effet prédit est également approprié ici. Si une interaction ou une modération spécifique est importante pour votre recherche, vous pouvez l'indiquer comme une hypothèse distincte.

Exemple : Si la durée conseillée d'une capsule vidéo pédagogique est inférieure à 10 minutes, on peut s'attendre à des meilleures scores de rétention de l'information si les participant-es sont exposé-es à deux vidéos de 4 minutes, avec une pause de 2 minutes au milieu, plutôt que à une vidéo de 8 minutes sans pause.

3 Plan de l'étude/expérience

Le Plan de l'étude/expérience donne des informations sur la manière dont les données sont générées/récoltées.

3.1 Type d'étude

Dans le template OSF on peut choisir parmi les options suivantes (traduction libre) :

- Expérience - Les chercheurs attribuent de manière aléatoire des traitements aux entités de l'étude, ce qui inclut les expériences sur le terrain ou en laboratoire. Elle est également connue sous le nom d'expérience d'intervention et comprend les essais contrôlés randomisés.
- Étude d'observation - Les données sont recueillies auprès d'entités d'étude qui ne sont pas assignés au hasard à un traitement. Cela inclut les enquêtes, les *expériences naturelles* et les modèles de discontinuité de régression.
- Méta-analyse - Revue systématique de résultats de contributions scientifiques (publiées)
- Autre

3.2 Connaissance des attributions/interventions

Dans une expérience, le type d'intervention (e.g. facteurs expérimentaux et modalités) peut être transparente ou cachée selon les différents acteurs impliqués :

- Les participant-es seront au courant de l'attribution/intervention
- Les participant-es ne seront pas au courant de l'attribution/intervention
- Les chercheurs/facilitateurs en contact avec les participant-es ne seront pas au courant de l'attribution/intervention
- Les chercheurs qui analysent les données ne seront pas au courant de l'attribution/intervention

L'idée de ces critères est de limiter les moins possibles des potentiels biais des participant-es et/ou des chercheurs. Le moins on connaît sur l'expérience, le moins le comportement risque d'être influencé, par exemple en direction des hypothèses.

3.3 Éléments cachés

Dans certaines expériences les chercheurs doivent cacher des éléments de procédure ou lié au matériel expérimental afin de pouvoir mieux contrôler d'autres facteurs d'influence. Par exemple on peut induire un état affectif des participants participant (le faire attendre longtemps dans une salle d'attente) ou leur cacher qu'une personne impliquée dans l'expérience est en réalité un chercheur qui doit *réciter* le même rôle dans toutes les passations.

3.4 Design/plan expérimental

Le design/plan expérimental illustre la manière dont l'intervention est séparée/comparée en facteurs avec différentes modalités. Le template OSF décrit cette partie de la manière suivante (traduction et adaption libres) :

Décrivez la conception de votre étude. L'essentiel est d'être aussi détaillé que nécessaire compte tenu des paramètres spécifiques de la conception. Il peut y avoir un certain chevauchement entre cette question et les questions suivantes. Ce n'est pas grave, tant que vous donnez suffisamment de détails dans l'un des critères pour fournir toutes les informations demandées. Les exemples incluent les plans inter-sujets, intra-sujets et mixtes ; différentes types de randomization (avec ou sans équilibration du nombre des participants) ; et enfin des plans à mesure répétées.

Exemple : L'expérience utilise un plan inter-sujets avec un facteur expérimental (support d'étude) avec trois modalités (smartphone, tablette, ou notebook).

3.5 Randomization

Si l'expérience prévoit l'attribution aléatoire aux modalités des facteurs expérimentaux, il faut indiquer comment cette attribution aléatoire est effectué. Parmi les choix possibles figurent :

- Attribution totalement aléatoire
- Attribution aléatoire avec équilibre du nombre des participants par condition expérimentale
- Attribution par blocs

Exemple : l'attribution aux trois conditions expérimentales sera tirée au sort avant chaque passation en utilisant des billes colorées cachées dans un sac

4 Plan d'échantillonnage

Le plan d'échantillonnage détermine et justifie comment les données sont récoltées.

4.1 Données existantes

Un élément fondamentale dans le pré-enregistrement consiste à quel moment les chercheurs décident de pré-enregistrer leur expérience. Idéalement, ceci se fait avant que toute donnée

soit collecté, mais parfois ceci n'est pas possible. À ce moment, il faut justifier les raisons pour lesquelles des données ont déjà été récoltées/analysées.

Le template OSF propose les options suivantes (traduction libre) dans un ordre qui rend progressivement le pré-enregistrement moins *efficace* :

- Enregistrement avant la création des données
- Enregistrement préalable à toute observation humaine des données
- Enregistrement avant l'accès aux données
- Enregistrement avant l'analyse des données
- Enregistrement après l'analyse des données

4.2 Procédures de collecte de données

La manière dont les données sont collectées au niveau concret. Le template OSF décrit cette section de la manière suivante (traduction et adaptation libre) :

Veuillez décrire le processus par lequel vous allez collecter vos données et vos critères d'inclusion et d'exclusion. Si vous utilisez des participant-es, cela doit inclure le groupe/population de référence, les modalités de recrutement, les critères de sélection si plus de participant-es que nécessaire répondent à l'appel, etc.

Exemple : les participant-es seront recruté-es à travers une plateforme en ligne spécialisée dans la passation de test utilisateurs sur des interfaces mobiles. Les participant-es devront être francophones, âgées de plus de 18 ans, et étudiant-es au niveau bachelor ou master. Une rémunération de 20\$ est dispensée directement par la plateforme à la fin de l'expérience.

4.3 Taille de l'échantillon

Définir le nombre d'entités (e.g participant-es) qui participeront à l'expérience. Le template OSF décrit cette section ainsi (traduction et adaptation libres) :

Décrivez la taille de l'échantillon de votre étude. Combien d'entités seront analysées dans l'étude ? Il peut s'agir du nombre de personnes, d'oiseaux, de salles de classe, de parcelles ou de pays inclus. Si les unités ne sont pas des individus, décrivez alors la taille requise pour chaque unité. Si vous utilisez une conception en *clusters* ou à plusieurs niveaux, décrivez le nombre d'unités que vous collectez à chaque niveau de l'analyse. Il peut s'agir du nombre d'échantillons ou d'une fourchette, d'un minimum ou d'un maximum.

Exemple : l'échantillon envisagé pour cette expérience est de 80 participant-es. Pour disposer d'un certain marge, nous allons recruter 100 participant-es pour limiter le risque de données manquantes.

4.4 Justification de la taille de l'échantillon

La taille de l'échantillon peut être définie selon un calcul de puissance statistique ou en raison de disponibilités limitées en termes de temps, d'argent ou de chercheurs/facilitateurs.

Exemple : nous estimons pouvoir recruter environ 50 participant-es pendant la période d'examens universitaires, moment pendant lequel les étudiant-es ne sont pas très disponibles.

4.5 Règles d'arrêt

Parfois il est difficile de définir à l'avance le nombre de participant-es, par exemple lors du recrutement volontaire. À ce moment, on peut définir des critères d'arrêt de sélection.

Exemple : nous allons répéter l'expérience sur un maximum de 4 semestre universitaires afin de garantir une uniformité dans les formats d'études/cours proposés.

5 Variables

Liste des variables définies/choisies par les chercheurs (VIs) ou mesurées (VDs, variables contrôlées, ...).

5.1 Variables indépendantes

Illustrer les facteurs expérimentaux définis/choisis par les chercheurs. Leur nombre et le nombre de modalités prévu pour chaque facteur.

Exemple : Les deux variables indépendantes présentes dans cette étude consiste dans :

1. Le niveau de difficulté d'une page HTML5 à coder, avec 3 modalités : *facile*, *moyenne* ou *difficile*
2. Le nombre d'écrans à disposition, avec 2 modalités : *un seul écran* ou *deux écrans*

5.2 Variables mesurées

Illustrer toutes les variables qui vont être mesurées, ainsi que la manière de le mesurer, avant ou pendant la tâche, comme par exemple les variables dépendantes utilisées pour mesurer le(s) phénomène(s) d'intérêt ou des éventuels facteurs à contrôler.

Exemple : L'expérience prévoit une variable dépendante et une variable contrôle. La variable dépendante consiste dans la perception d'usabilité d'un système/logiciel informatique mesuré à travers une traduction française de la *System Usability Scale* (Brooke, 1996), une échelle à 10 items, chacun évalué sur 5 points allant de *fortement en désaccord* à *fortement en accord*. La variable contrôle consiste dans la perception de fatigue avant d'entamer l'expérience, auto-reportée par les participant-es sur une échelle à 7 points allant de *pas du tout fatigué* à *très fatigué*.

5.3 Indices

Si une VD ou une variable contrôle est calculée à l'aide d'un score composé ou d'une sélection entre plusieurs possibilités, indiquer comment cet indice est déterminé.

Exemple : Le temps de réaction du participant est déterminé en moyennant les résultats de 8 stimuli sur 10, avec le temps de réponse plus long et plus court qui sont écartés de l'indice.

6 Type d'analyse

Le type d'analyse définit quel test/procédure d'analyse statistique est utilisée pour détecter/tester les effets/hypothèses.

6.1 Modèle statistique

Le type de modélisation de données utilisées pour l'analyse est un élément déterminant, car il dépend de plusieurs choix épistémologiques et techniques. Le template OSF décrit cette section de la manière suivante (traduction et adaptation libres) :

Quel modèle statistique allez-vous utiliser pour tester chaque hypothèse ? Veuillez inclure le type de modèle (par exemple *t* test de Welch, ANOVA, régression multiple, modèle multi-niveaux, SEM, etc) et la spécification du modèle. Cela inclut chaque variable qui sera incluse, toutes les interactions, les analyses de sous-groupes, les contrastes par paires ou complexes, et tous les tests de suivi des tests omnibus. Fournissez suffisamment de détails pour qu'une autre personne puisse effectuer la même analyse avec les informations fournies.

Exemple : Une analyse de variance simple (omnibus ANOVA) sera effectué avec les trois modalités de la variable indépendante *support d'étude* (smartphone, tablette, ou ordinateur portable) et la variable dépendante *perception de satisfaction*. Des comparaisons multiples sont prévus entre les trois modalités.

6.2 Transformations

Sous certaines conditions qui dépassent le niveau introductif de ce cours, les données doivent/peuvent subir des transformations mathématiques.

6.3 Critères d'inférence statistique

Bien que l'habitude général en sciences sociales est celle d'interpréter les résultats selon le seuil de significativité d'une p -valeur inférieur à 0.05, l'inférence statistique peut être plus complexe.

L'utilisation de statistiques dites *bayésiennes* commencent à être de plus en plus présentes dans les articles en sciences sociales. On commence également à déployer des tests qui cherchent à déterminer l'absence d'un effet plutôt que la présence d'un effet. Il est donc important de commencer à réfléchir sur la manière avec laquelle on établit la *qualité* du modèle en termes statistiques, car ceci a des conséquences directes sur l'inférence *pratique*.

Exemple : l'inférence statistique sera effectué en utilisant un niveau de contrôle de l'erreur de Type I fixé à $\alpha = 0.05$ pour l'ANOVA simple. Les comparaisons entre modalités utiliserons le même seuil, mais avec une correction de Tukey pour contrôler les comparaisons multiples.

6.4 Exclusion de données

Un autre élément important concerne l'exclusion *statistique* de données, comme par exemple les valeurs extrêmes.

6.5 Données manquantes

Même principe pour les données manquantes ou incomplètes, il serait utile de manifester à l'avance comment vous comptez les gérer.

6.6 Analyses exploratoires

Enfin, comme indiqué en introduction, le pré-enregistrement des expériences n'exclut pas de mener des analyses exploratoires qu'on peut aussi spécifier à l'avance.

Références

Brooke, J. (1996). *SUS: A quick and dirty usability scale* (B. A. Weerdmester & I. L. McClelland, Éd.). Taylor & Francis.

- John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the Prevalence of Questionable Research Practices With Incentives for Truth Telling. *Psychological Science*, 23(5), 524532. <https://doi.org/10.1177/0956797611430953>
- Kerr, N. L. (1998). HARKing: Hypothesizing After the Results are Known. *Personality and Social Psychology Review*, 2(3), 196217. https://doi.org/10.1207/s15327957pspr0203_4
- Makel, M. C., Hodges, J., Cook, B. G., & Plucker, J. (2019). *Questionable and Open Research Practices in Education Research*.
- Nelson, L. D., Simmons, J., & Simonsohn, U. (2018). Psychology’s Renaissance. *Annual Review of Psychology*, 69(1), 511-534. <https://doi.org/10.1146/annurev-psych-122216-011836>
- Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., Buck, S., Chambers, C. D., Chin, G., Christensen, G., Contestabile, M., Dafoe, A., Eich, E., Freese, J., Glennerster, R., Goroff, D., Green, D. P., Hesse, B., Humphreys, M., ... Yarkoni, T. (2015). Promoting an open research culture. *Science*, 348(6242), 14221425. <https://doi.org/10.1126/science.aab2374>
- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251). <https://doi.org/10.1126/science.aac4716>