Pour un partage des données d'essais cliniques à des fins de recherche

Sharing clinical trial data for research purposes

Marc Buyse[1,2]
Frank Rockhold[3]

[1] International Drug Development Institute (IDDI)
775, Francisco Street
CA 94133 San Francisco
États-Unis
<marc.buyse@iddi.com>

[2] Université de Hasselt
Département de biostatistique
Martelarenlaan 42
BE3500 Hasselt
Belgique

[3] Duke Clinical Research Institute
Duke University Medical School
Durham, NC
PO Box 17969
États-Unis
<frank.rockhold@duke.edu>

Remerciements et autres mentions :

Financement : aucun.

Liens d'intérêts : les auteurs déclarent ne pas avoir de lien d'intérêt.

RÉSUMÉ

Le partage des données cliniques fait partie d'un mouvement plus général vers la science ouverte. Nous discutons les avantages et les inconvénients du partage des données, nous décrivons les plates-formes existantes d'accès aux données, et nous suggérons quelques principes simples pour faciliter le partage des données cliniques à l'avenir.

● Mots clés : essais cliniques ; partage des données ; science ouverte.

ABSTRACT

Sharing clinical data is part of a broader movement towards open science. We discuss the pros and cons of data sharing, describe existing data access platforms, and suggest some simple principles to facilitate clinical data sharing in the future.

● *Key words: clinical trials; data sharing; open science.*

Notre époque est celle des paradoxes : alors que des données personnelles, parfois extrêmement sensibles, sont partagées (et vendues à des fins mercantiles) par les réseaux sociaux, les données d'essais cliniques, qui constituent une potentielle mine d'or pour la recherche médicale, sont en revanche mieux protégées que les lingots de Fort Knox. Jusqu'à il y a peu, il était pratiquement impossible d'y avoir accès, en tout cas pour les essais conduits par les groupes pharmaceutiques, car ces données étaient considérées comme « propriétaires » – c'est-à-dire appartenant aux promoteurs des essais. Les patients qui signaient le consentement éclairé d'un essai clinique visant à démontrer l'efficacité et l'innocuité d'un médicament en cours de développement auraient été bien surpris de savoir que leurs données ne pourraient en principe être utilisées que dans le seul cadre de l'essai clinique auquel ils avaient accepté de participer, même si dans de très nombreux cas l'efficacité du médicament ainsi que son profil de toxicité ne peuvent être établis de manière définitive que par une méta-analyse des données de nombreux essais. Or, ces méta-analyses sont notoirement difficiles à conduire du fait de la difficulté d'accéder aux données individuelles des patients de tous les essais. Par exemple, la méta-analyse des essais cliniques du tamoxifène dans le cancer du sein, initiée par Richard Peto et ses collègues à Oxford dans les années 1980, a exigé des années de travail avant que les données de (quasiment) tous les essais soient prêtes à l'analyse. Cette méta-analyse fait date car elle a montré comment la combinaison des données individuelles de nombreux essais établissait de manière certaine l'efficacité d'un médicament, là où la plupart des essais cliniques, d'effectifs trop restreints, ne permettaient pas de conclure. Rien ne justifie que cette méta-analyse ait pris tant de temps et qu'elle n'ait pu voir le jour que grâce à l'obstination de quelques chercheurs clairvoyants [1]. Un autre exemple, plus malheureux,

Tirés à part : F. Rockhold

Pour citer cet article : Buyse M, Rockhold F. Pour un partage des données d'essais cliniques à des fins de recherche. *Innov Ther Oncol* 2019 ; 5 : 208-212. doi : 10.1684/ito.2019.0166

doi: 10.1684/ito.2019.0166

© John Libbey Eurotext, 2019

Tirés à part

concerne la rosiglitazone, un antidiabétique oral retiré du marché suite à une série d'analyses contestées et d'une méta-analyse limitée aux données publiées d'essais sélectionnés, qui tendaient à montrer une augmentation du risque cardiovasculaire chez les patients prenant la rosiglitazone. Ces multiples analyses, utilisant des données disparates et non accessibles à tous, ont créé la confusion plutôt que d'apporter une réponse définitive au problème [2]. Là aussi, l'accès libre aux données de tous les essais aurait très certainement permis de conclure plus vite et de manière définitive. Le but du présent article est de discuter pourquoi l'accès aux données des essais cliniques est éthiquement et scientifiquement indispensable, et comment il deviendra possible dans un avenir proche.

Le mouvement vers une « science ouverte »

L'accès aux données, ou le partage des données (nous utiliserons ces termes sans distinction), doit se comprendre dans le mouvement vers une « science ouverte », qui tend à transformer la pratique scientifique contemporaine en prônant le libre accès à toutes les publications scientifiques revues par les pairs, le partage des données de recherche, la transparence du processus d'évaluation scientifique, l'intégration de non-professionnels dans les projets de recherche scientifique (par exemple, les associations de patients dans les projets de recherche clinique), et le développement de réseaux sociaux où sont discutées librement les hypothèses des chercheurs. En bref, le but est de générer « *un savoir transparent et accessible, partagé et développé par le biais de réseaux de collaboration* » [3-5]. Le partage des données de recherche a le potentiel d'encourager la collaboration, d'offrir une alternative aux données simulées, de dépasser les contraintes liées aux petits effectifs (par exemple, dans les maladies rares ou « orphelines »), de réduire les coûts de la recherche en évitant des duplications, de promouvoir la transparence et la reproductibilité des résultats, et de contribuer à l'éducation et à la formation des jeunes chercheurs. Dans le domaine de la recherche clinique, l'accès, le partage et la transparence des données des essais pourraient renforcer la crédibilité des résultats des essais, permettre des analyses complémentaires, par exemple des analyses de sous-groupes, améliorer la conception des futurs essais, et, de manière plus générale, accélérer le développement de nouveaux traitements, ce qui est aujourd'hui indispensable pour tester le nombre croissant de traitements prometteurs issus tant de la biologie moléculaire que des progrès technologiques. Les modalités actuelles d'une recherche clinique longue, coûteuse et cloisonnée sont totalement dépassées.

Pour ou contre le partage des données cliniques ?

Même si les arguments en faveur d'une science ouverte paraissent convaincants, le partage des données cliniques reste particulièrement délicat (les données biologiques, par exemple de type génomique, sont en revanche partagées et accessibles sur Internet). Le journal *PLOS* relate un cas récent et exemplaire de refus d'accès aux données d'un essai clinique par un promoteur académique soucieux de protéger l'identité des patients inclus dans l'essai [6]. Ce cas illustre toute la complexité du problème, depuis une décision de justice ayant obligé le promoteur à partager certaines données de l'essai, au désaccord profond des experts quant à la nécessité de partager toutes les données et, le cas échéant, quant à la manière d'effectuer ce partage. Le *tableau 1* résume les principaux arguments qui ont été avancés en faveur, ou à l'encontre, du partage des données cliniques.

Les arguments scientifiques en faveur du partage tombent sous le sens, ainsi d'ailleurs que l'argument éthique : les données cliniques sont générées grâce à la participation volontaire de patients qui souhaitent non seulement recevoir le meilleur traitement possible pour leur maladie mais aussi, par la même occasion, contribuer au progrès thérapeutique dans un but humanitaire [7, 8]. Les inconvénients du partage ne sont cependant pas négligeables, en particulier celui du risque d'identification des participants aux essais. Le consensus est actuellement que tout partage des données cliniques doit se faire de manière responsable [9].

Tableau 1. Avantages et inconvénients du partage des données cliniques.
Table 1. *Pros and cons of sharing clinical data.*

Avantages	Inconvénients
Confirmation des résultats par des tiers	Analyses erronées, biaisées ou trompeuses
Méta-analyse d'essais semblables	Données peu interprétables par des tiers
Analyses plus poussées des données	Risque d'identification des participants aux essais
Planification d'essais cliniques ultérieurs	Propriété des données
Utilisation maximale de toutes les données	Pas de crédit pour les producteurs des données
But humanitaire des essais	Coûts

© John Libbey Eurotext, 2019

Premiers pas vers la science ouverte en recherche clinique

De grands progrès ont récemment été accomplis dans l'accès aux données de recherche clinique. Ces progrès sont l'aboutissement d'un mouvement amorcé il y a une quinzaine d'années avec une initiative des éditeurs de revues médicales visant à réduire le « biais de publication ». En 2004, le Comité international des rédacteurs de revues médicales (ICMJE) décidait que les rapports d'essais cliniques ne seraient désormais publiés dans leurs journaux que si leurs protocoles avaient été enregistrés de manière prospective dans une base de données telle que ClinicalTrials.gov [10]. L'enregistrement des essais cliniques a été depuis un succès retentissant, avec plus de 300 000 essais enregistrés dans ClinicalTrials.gov aujourd'hui[1]. L'ICMJE a fait un pas de plus en 2017, en décidant que les essais cliniques soumis pour publication devraient inclure une déclaration explicite de partage des données [11]. Ce mouvement vers un partage des données individuelles des participants correspond aux exigences des instituts nationaux de la santé (NIH) aux États-Unis [12, 13]. En Europe, l'Agence européenne des médicaments (EMA) a publié une audacieuse déclaration de politique générale en 2015, dans le but de « *rendre les données cliniques disponibles de manière proactive, afin de permettre au public de surveiller et d'appliquer les nouvelles connaissances dans les recherches futures, le tout dans l'intérêt de la santé publique* » [14]. Cette politique devra être mise en œuvre par étapes, avec la publication des rapports cliniques dans une première phase (comme c'est déjà le cas aux États-Unis) et l'accès aux données individuelles des patients dans une seconde phase. Ni le calendrier ni les modalités de cet accès aux données ne sont encore précisés. Il est clair cependant que l'accès devra se faire conformément au règlement général sur la protection des données (RGPD), adopté en 2016 par le Parlement européen pour garantir aux citoyens européens la propriété de leurs données et le libre choix de leur utilisation [15].

Plates-formes d'accès aux données

Les agences fédérales de recherche aux États-Unis ont mis en place des plates-formes remarquables d'accès aux données : par exemple, en oncologie, le *Cancer Data Access System* du *National Cancer Institute*[2]. Plus récemment, l'industrie pharmaceutique a mis en place plusieurs initiatives distinctes pour faire progresser le partage de données dans le secteur de la recherche clinique : les plus en vue sont le *Project Data Sphere*[3], *Clinical Study Data Request*[4], *Yale Open Data Access*[5], *Duke Clinical Research Institute-Bristol-Myers Squibb Data Sharing Initiative*[6] et le *Vivli Center for Global Clinical Research Data*[7], ce dernier ayant pour ambition de fédérer les autres pour offrir, à terme, un portail unique d'accès aux données.

Quel type d'accès aux données ces plates-formes permettent-elles ?

Le *tableau 2* résume les différents types d'accès possible, du plus libéral (accès ouvert) au plus restreint (accès nul). La norme, jusqu'à il y a peu, était l'accès nul ou exceptionnel aux données, toute autre forme d'accès est donc à considérer comme un progrès vers la science ouverte. Cependant, tous les systèmes d'accès aux données existants imposent une série de contraintes reprises dans le *tableau 2*. Certaines de ces contraintes sont justifiées pour assurer le partage « responsable » des données (par exemple, l'engagement moral de ne pas identifier les participants aux essais), d'autres semblent contestables (droit *de veto* à l'accès des données par le propriétaire), d'autres enfin sont scientifiquement et éthiquement inacceptables (le projet *Data Sphere*, par exemple, ne mettait initialement à disposition des chercheurs que les données des bras contrôles des essais randomisés).

Et les essais académiques ?

L'existence de ces plates-formes constitue une formidable preuve du concept que l'accès aux données est possible, sous certaines conditions. Le nombre d'essais repris par ces plates-formes reste limité. En particulier, les essais académiques n'y figurent pas, or ils constituent plus de la moitié des essais recensés par ClinicalTrials.gov. En effet, aucune des plates-formes de partage mises en place par l'industrie pharmaceutique n'accorde de crédit aux générateurs des données, ce qui constitue un obstacle majeur à l'extension de ces plates-formes aux essais universitaires [16]. Outre l'implication souvent nécessaire des investigateurs des essais dans une réanalyse de leurs données [1], il serait hautement souhaitable que ces investigateurs soient coauteurs de publications ultérieures basées sur ces données [17]. Une politique de publication qui reconnaîtrait le rôle des générateurs de données éliminerait à terme la tendance actuelle des chercheurs à rester « assis » le plus longtemps possible sur leurs données, ce qui n'est souhaitable ni d'un point de vue scientifique, ni d'un point de vue éthique ; après tout, les données appartiennent aux patients qui ont accepté de participer à la recherche, et non aux chercheurs qui ont coordonné les essais, comme le souligne le nouveau règlement général européen sur la protection des données [15].

[1] https://www.clinicaltrials.gov
[2] https://biometry.nci.nih.gov/cdas/
[3] https://projectdatasphere.org/projectdatasphere/html/home
[4] https://clinicalstudydatarequest.com/Default.aspx
[5] https://yoda.yale.edu/
[6] https://dcri.org/our-work/analytics-anddata-science/data-sharing/bms-studies/
[7] https://vivli.org/

© John Libbey Eurotext, 2019

Tableau 2. Types d'accès aux données.
Table 2. *Types of access to data.*

Type d'accès	Caractéristiques
Ouvert	Données publiquement accessibles sans restriction
Protégé	Interface sécurisée permettant l'accès avec mot de passe aux utilisateurs autorisés
Partiel	Certaines données accessibles seulement
Limité	Données non copiables, accessibles sur une plate-forme Internet ne permettant que certaines analyses
Conditionnel	Accès aux données restreint aux utilisateurs satisfaisant à certaines conditions • approbation du projet de recherche par un comité d'éthique ; • pertinence du projet certifiée par un panel d'experts indépendants ; • déclaration d'absence de conflits d'intérêts ; • soumission du curriculum vitae de tous les investigateurs ; • soumission d'un projet détaillé, y compris un plan d'analyse statistique ; • engagement moral de ne pas identifier les participants aux essais ; • engagement moral de ne pas partager les données ; • engagement moral de publier les résultats de la recherche.
Opposable	Droit *de veto* à l'accès des données par le propriétaire
Nul	Données propriétaires non partagées (hors agences d'enregistrement)

Qui finance le partage des données ?

Vient enfin la question du coût du partage des données, tant pour le générateur des données que pour l'utilisateur. Ce coût peut paraître trivial par rapport aux autres coûts de la recherche clinique, il n'en demeure pas moins un obstacle majeur pour les essais académiques. En effet, pour que le partage des données soit effectif, il faut qu'elles soient compréhensibles et réutilisables par des tiers. Les formats de données préconisés par *Clinical Data Interchange Standards Consortium* (CDISC)[8], et exigés par les agences pour les essais à visée d'enregistrement, sont peu utilisés dans les essais académiques. Les données de ces essais sont souvent difficiles à manipuler, à interpréter, et à intégrer dans une méta-analyse d'essais utilisant des conventions différentes pour le codage et le calcul des mêmes données. Il semble difficile d'exiger que les générateurs des données paient également pour les partager avec d'autres, pas plus qu'il ne serait souhaitable que les coûts du partage soient entièrement couverts par ceux qui veulent utiliser ces données à des fins de recherche. Les plates-formes existantes sont financées par les agences nationales et l'industrie pharmaceutique, ainsi que par des organisations caritatives (par exemple le *Wellcome Trust* en Grande-Bretagne[9] et la fondation Bill et Melinda Gates aux États-Unis[10]). La question du financement du partage des données est un problème de société qui devra être résolu pour pouvoir honorer pleinement la participation des patients aux essais cliniques.

Quelques principes simples pour le partage des données d'essais cliniques

Il est probable qu'à terme les données de tous les essais cliniques, qu'ils soient conduits par les pouvoirs publics, le secteur privé ou le secteur académique, devront être rendues publiques. Les principes suivants, simples et suffisants, peuvent aider les promoteurs d'essais cliniques à préparer l'indispensable transition vers une science ouverte [18] :

- préparer un plan pour le partage des données, même si les moyens de l'essai ne permettent pas de financer un tel partage. La planification d'un éventuel partage des données avant le début d'une étude permet d'éviter bien des complications concernant le consentement des patients, la collecte et l'anonymisation des données, le respect des règlementations, le formatage des données et l'efficacité de leur partage, et la mise à disposition des informations permettant aux tiers de réanalyser les données (protocole, cahier d'observation, programmes de recodage et d'analyse des données pour assurer la reproductibilité des résultats[11]) ;
- protéger les droits des patients et responsabiliser les utilisateurs des données. Il ne s'agit pas seulement de préserver l'anonymat des patients et de sécuriser le stockage des données, mais de transférer cette responsabilité aux utilisateurs suivants. Comme dans tout projet de recherche clinique, l'implication des associations de patients est essentielle pour assurer

[8] https://www.cdisc.org/
[9] https://wellcome.ac.uk/
[10] https://www.gatesfoundation.org/
[11] https://reproducibleresearch.net/

© John Libbey Eurotext, 2019

une utilisation optimale et responsable des données des essais ;

- citer la source des données, au minimum, mais de préférence impliquer les investigateurs principaux de l'essai dans toute réanalyse des données, et les associer comme coauteurs des éventuelles présentations/publications. À noter que ce principe n'est respecté par aucune des plates-formes existantes, qui perdent dès lors tout attrait pour les essais académiques ;
- citer la source de financement qui a permis le partage des données. Les agences fédérales américaines telles que le NIH et certaines fondations privées *(Bill and Melinda Gates Foundation)* exigent que les données des études qu'ils financent soient partagées.

Take home messages

- Le partage des données d'essais cliniques devient la norme dans le mouvement vers une « science ouverte ».
- L'industrie pharmaceutique finance des plates-formes d'accès aux données.
- Les données des essais académiques sont insuffisamment partagées.

RÉFÉRENCES

1. CTSU. *Early Breast Cancer Trialists' Collaborative Group*. https://www.ctsu.ox.ac.uk/research/ebctcg

2. Cohen D. Rosiglitazone – what went wrong. *BMJ* 2010 ; 341 : 530-4.

3. Vicente-Saez R, Martinez-Fuentes C. Open Science now: a systematic literature review for an integrated definition. *J Bus Res* 2018 ; 88 : 428-36.

4. Nosek BA, Alter G, Banks GC, *et al.* Promoting an open research culture. *Science* 2015 ; 348 : 1422-5.

5. Center for Open Science. *TOP guidelines*. 2019. http://cos.io/TOP

6. Puebla I, Heber J. Data sharing in clinical research: challenges and open opportunities. 2017. https://blogs.plos.org/everyone/2017/05/02/data-sharing-in-clinical-research/

7. Ellis PM. Attitudes towards and participation in randomised clinical trials in oncology: a review of the literature. *Ann Oncol* 2000 ; 11 : 939-45.

8. Mello MM, Lieou V, Goodman SN. Clinical trial participants' views of the risks and benefits of data sharing. *N Engl J Med* 2018 ; 378 : 2202-11.

9. Institute of Medicine. *Sharing clinical trial data: maximizing benefits, minimizing risk*. Washington (DC) : National Academies Press, 2015.

10. Zarin DA, Tse T, Williams RJ, *et al.* Trial reporting in ClinicalTrials.gov – The final rule. *N Engl J Med* 2016 ; 375 : 1998-2004.

11. Taichman DB, Sahni P, Pinborg A, *et al.* Data sharing statements for clinical trials: a requirement of the International Committee of Medical Journal Editors. *Ann Intern Med* 2017 ; 167 : 63-5.

12. Hudson KL, Lauer MS, Collins FS. Toward a new era of trust and transparency in clinical trials. *JAMA* 2016 ; 316 : 1353-4.

13. NIH. *Proposed provisions for a draft NIH data management and sharing policy*. https://osp.od.nih.gov/wp-content/uploads/Data_Sharing_Policy_Proposed_Provisions.pdf

14. European Medicines Agency. *Clinical data publication*. https://www.ema.europa.eu/en/human-regulatory/marketing-authorisation/clinical-data-publication

15. Parlement européen. *Règlement général sur la protection des données (RGPD)*. Foster Open Science, 2016. https://www.fosteropenscience.eu/resources

16. Lo B, DeMets DL. Incentives for clinical trialists to share data. *N Engl J Med* 2016 ; 375 : 1112-5.

17. Bierer BE, Crosas M, Pierce HH. Data authorship as an incentive to data sharing. *N Engl J Med* 2017 ; 376 : 1684-7.

18. Rockhold F, Bromley E, Wagner E, Buyse M. Open Science: the open clinical trials data journey. *Clin Trials* 2019 (in press).

Tiré à part d'auteur

© John Libbey Eurotext, 2019