

STRUCTURER ET HARMONISER LE DESIGN DES ÉTUDES EN VIE RÉELLE SUR DONNÉES SECONDAIRES : L'APPORT DE L'INITIATIVE REPEAT



REPEAT

Reproducible Evidence: Practices to Enhance and Achieve Transparency

Note écrite par

- Audrey Lajoinie, RCTs
- Asma Hamid Melaine, Oracle
- Françoise Brugnard, stève consultants
- Manon Belhassen, Epimentis

Membres du groupe de travail RWD de l'AFCROs

11/2025

Ces dernières décennies, les études en vie réelle réalisées à partir de bases de données existantes, appelées données secondaires – e.g. bases médico-administratives (Système National des Données de Santé (SNDS)), registres, dossiers médicaux électroniques, objets connectés, Entrepôt de Données de Santé (EDS) – ont connu un essor considérable. L'intérêt croissant pour ces sources de données se traduit par une augmentation significative du nombre de projets et de publications scientifiques. Cette dynamique s'accompagne toutefois d'une grande hétérogénéité dans la manière de concevoir, de documenter et de présenter le plan expérimental (design) de ces études, ce qui peut nuire à la transparence, à la reproductibilité et à la comparaison des résultats. L'initiative REPEAT a conçu des outils pour répondre à cette problématique, que nous nous proposons de présenter ici dans ses grandes lignes.

1. Pourquoi structurer la présentation des designs des études en vie réelle réalisées avec des données secondaires ?

L'essor des études en vie réelle s'appuyant sur des données secondaires issues de sources variées rend indispensable une description claire, transparente et standardisée du design de l'étude. Cet essor s'accompagne toutefois d'une grande hétérogénéité dans la conception, la documentation et la description du design des études. Cette hétérogénéité dans la présentation des études entraîne plusieurs difficultés, et notamment :

- elle peut nuire à la compréhension, la reproductibilité, l'interprétation et la comparaison des résultats rapportés [1],
- elle limite la comparabilité entre les études, d'autant plus que les données secondaires mobilisées sont issues de sources très diverses.

Cette hétérogénéité est en réalité le plus souvent présente dès la rédaction du protocole d'étude. Des chercheurs ont ainsi conclu à la nécessité de partager des templates de protocoles avec des définitions codifiées (termes usuels, algorithmes, codes), et de mettre en place des standards de reporting (check-lists) adaptés aux études en vie réelle sur données secondaires.

2. L'initiative REPEAT : Origines et ambitions

L'initiative internationale **REPEAT** (*REProducible Epidemiology And Treatment Effectiveness Studies*) a été développée par une équipe de chercheurs en pharmacoépidémiologie et en science de la reproductibilité, affiliés notamment à l'université de Harvard (Brigham and Women's Hospital), en collaboration avec la communauté OHDSI (*Observational Health Data Sciences and Informatics*). Elle vise à promouvoir des standards de transparence, de structuration du design des études et de reporting pour les recherches en vie réelle, notamment celles menées à partir de données secondaires.

A noter que cette initiative s'inscrit en complément de celle des groupes **RECORD** ou **STROBE** : alors que ces dernières s'attachent à améliorer la clarté du reporting dans les publications scientifiques, **REPEAT** se concentre sur la **reproductibilité technique et conceptuelle des études**, en proposant des outils pour formaliser, dès le protocole, le design de l'étude : fenêtres temporelles, date index, critères de sélection et de suivi, etc.

Initiative	Objectif principal	Focus principal	Stade d'application
REPEAT	Garantir la reproductibilité technique et conceptuelle des études en vie réelle	Standardisation de la conception et des méthodes dès le protocole d'étude	Conception + protocole + analyse
STROBE	Améliorer la qualité du reporting des études observationnelles en général	Clarté et exhaustivité du reporting dans les publications	Publication finale
RECORD	Améliorer la clarté du reporting spécifiquement pour les études utilisant des données secondaires (bases de données de santé)	Reporting transparent des données secondaires utilisées	Publication finale

Les outils de l'initiative REPEAT sont d'ores et déjà utilisés par la communauté OHDSI dans des projets collaboratifs multibases, par des auteurs dans des journaux de référence en pharmacoépidémiologie, dans les guides récents de la ISPE et d'autres institutions internationales. L'EMA ou le Health Data Hub s'y réfèrent également ponctuellement, et certains consortiums (ex. EHDEN) encouragent son usage.

3. Deux apports majeurs de l'initiative REPEAT : un vocabulaire harmonisé et une représentation graphique

REPEAT a développé deux outils clés pour structurer la conception d'étude :

- Un **glossaire normalisé** : il définit des concepts tels que index date, cohort entry event, observation window, time-at-risk ou encore washout period, souvent employés de manière floue ou variable.
- Des **schémas graphiques standardisés** (*graphical depiction*) qui permettent de représenter visuellement le design d'une étude, ses fenêtres d'observation et ses critères d'inclusion et de suivi.

L'objectif est de **parler un langage commun** entre épidémiologistes, biostatisticiens, data managers, et évaluateurs (relecteurs, autorités, financeurs...).

• Quelques définitions utiles issues de REPEAT

Terme	Définition REPEAT	Pourquoi c'est important	Exemple
Date Index (ou date d'ancrage)	Date de référence (inclusion) d'un individu dans la cohorte, généralement notée T0	Point de départ pour le suivi ; choix critique pour éviter les biais temporels	Date de première dispensation d'un traitement étudié
Evènement d'entrée dans la cohorte	L'évènement qui déclenche l'inclusion dans la cohorte	Nécessite une définition claire et reproductible	Première dispensation du traitement étudié au cours d'une période de temps donnée

Fenêtre d'observation	Période pendant laquelle les données du patient sont observables	Peut inclure une période rétrospective pour mesurer des antécédents	Sur 3 années avant la date index et jusqu'au 31 décembre 2025. Sur 3 années avant la date index et jusqu'à 2 années après la date index
Période à risque	Fenêtre pendant laquelle on considère que l'événement d'intérêt peut se produire	Sa définition affecte directement les mesures d'incidence	De la date index jusqu'à 90 jours après la date de dernière dispensation du traitement étudié
Période de Washout	Période sans exposition ou événement, utilisée pour sélectionner des "nouveaux utilisateurs"	Évite de sélectionner initiations et poursuites de traitement dans une même cohorte	Période de 3 ans sans dispensation du traitement étudié

Application au SNDS

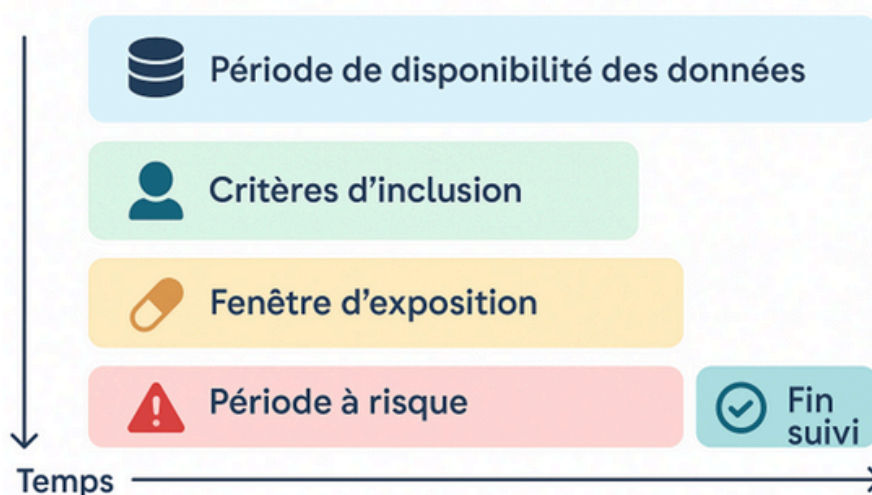
Dans les études SNDS, **la date index** est souvent définie comme la date de première délivrance d'un traitement, la date d'un acte, de l'enregistrement d'une ALD, ou d'une hospitalisation. Elle peut correspondre à la date la plus ancienne de plusieurs de ces critères. Il est essentiel de bien justifier cette date, surtout en cas de traitements chroniques ou de diagnostics tardifs (souvent décalés dans le temps), et d'anticiper les biais sur l'interprétation des résultats. L'**observation window** est souvent bornée par la date de dernières données disponibles dans le SNDS au moment de l'extraction.

- **Un exemple de schéma graphique : plus clair qu'un long paragraphe**

Les schémas REPEAT permettent de synthétiser un protocole complexe sous forme de frise temporelle. On y visualise :

- les critères d'inclusion,
- la période de disponibilité des données,
- la fenêtre d'exposition,
- la période à risque,
- la période de washout,
- la définition de la fin de suivi.

Cet outil est particulièrement utile pour faciliter les discussions méthodologiques en phase de conception, en comités de pilotage ou pour le dépôt réglementaire.



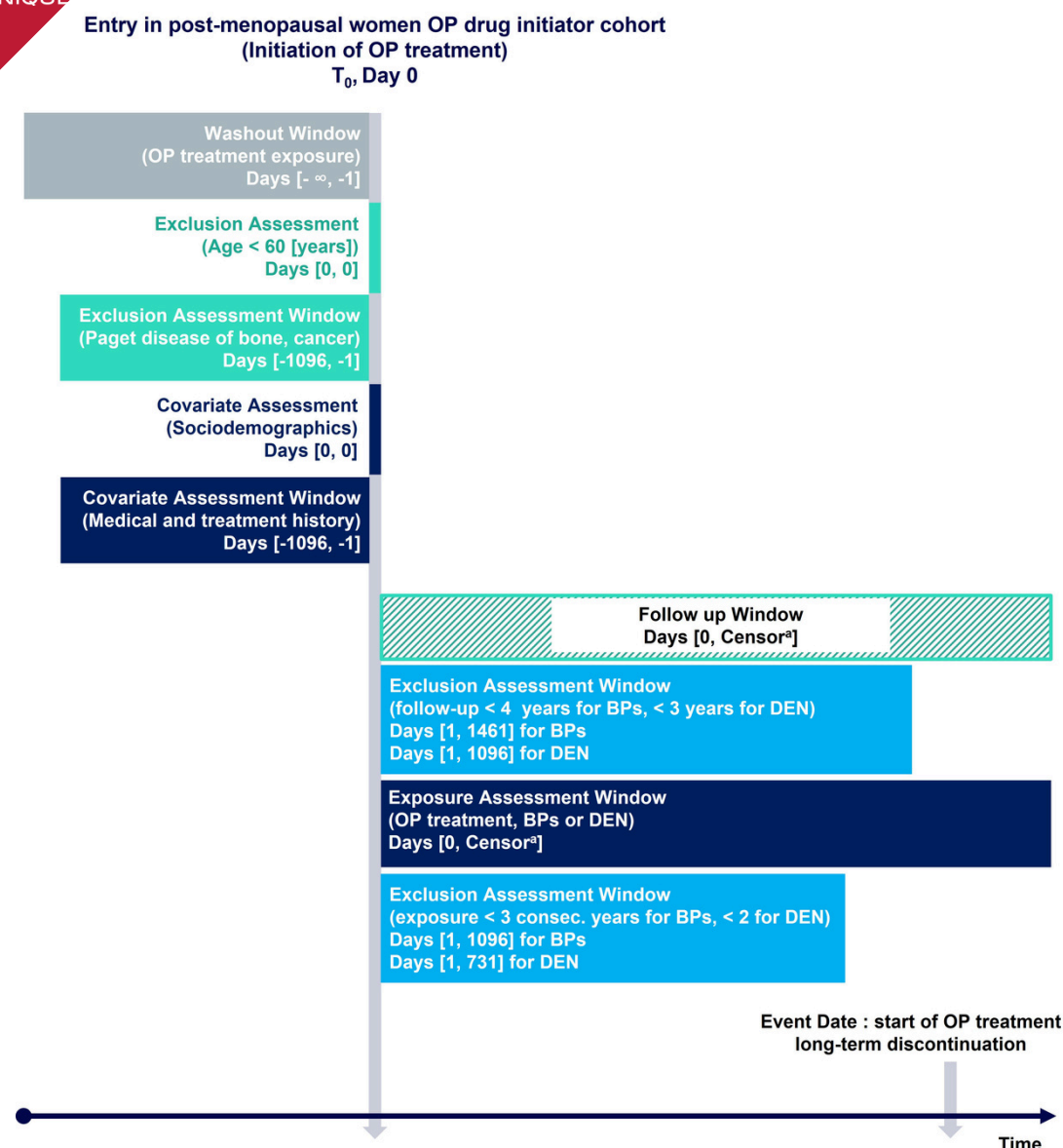
Et si on appliquait ce schéma à une étude SNDS?

Prenons l'exemple d'une étude portant sur l'arrêt au long cours des traitements anti-ostéoporotiques chez les femmes atteintes d'ostéoporose post-ménopausique [2].

Dans cette étude utilisant le SNDS :

- La date index correspond à la date d'initiation d'un traitement anti-ostéoporotique (bisphosphonates ou denosumab), entre 2012 et 2017.
- Une période de washout est appliquée pour identifier les nouvelles utilisatrices (absence de traitement sur plusieurs années).
- Des fenêtres d'exclusion et d'évaluation des covariables précèdent cette date (âge < 60 ans, maladie de Paget, antécédents de cancer).
- La période de suivi (follow-up window) commence à la date index et se termine à la date de censure (décès, fin de couverture, ou 31/12/2021).
- L'événement d'intérêt est l'arrêt prolongé du traitement (non-reprise sur plusieurs années).

Un schéma graphique inspiré de REPEAT (voir ci-dessous) permet de visualiser les différentes fenêtres temporelles, facilitant la compréhension du protocole et la reproductibilité de l'étude.



a. Censored at (i) 31 December 2021, or (ii) death, or (iii) the end of coverage by health insurance schemes available in the SNDS; $-\infty$ corresponds to the maximum history period available, i.e., 1 January 2009

BPs: Bisphosphonates; DEN: Denosumab

Figure 1: **Représentation graphique du design de l'étude SNDS-PMO selon les concepts de l'initiative REPEAT, avec l'initiation du traitement anti-ostéoporotique comme date index.** [2]

Et si on appliquait ce schéma à un registre ?

Prenons un registre de patients atteints de cancer. L'index date pourrait être la date de confirmation histologique. La time-at-risk commence après un traitement curatif. Le schéma graphique clarifie les critères de censure (récidive, décès, fin de suivi). Dans le SNDS, en revanche, la date de diagnostic n'est pas directement observable, il faut construire des algorithmes à partir des séquences de soins (actes CCAM, séjours PMSI, etc.). La documentation rigoureuse de ces choix méthodologiques constitue donc un prérequis indispensable pour garantir la comparabilité et la validité des résultats.

4. Études multisources : REPEAT comme facilitateur d'interopérabilité

Avec la multiplication des études à partir de plusieurs sources de données (bases médico-administratives, registres, dossiers médicaux électroniques, cohortes de recherche), les difficultés méthodologiques s'accroissent : différences de structure, de fenêtres temporelles, de définitions d'exposition et d'événements, d'unités de temps...

Dans ce contexte, les outils REPEAT permettent de formaliser une structure commune de l'étude, même si les données ne sont pas directement comparables. Ils offrent :

- une représentation visuelle commune du design, adaptable à chaque source ;
- un vocabulaire partagé pour les fenêtres d'observation, les dates index, les covariables ;
- un cadre facilitant les discussions préalables entre partenaires (recherche, réglementaire, industriels).

Cela s'inscrit parfaitement dans les logiques de réseaux fédérés (type OHDSI ou HDH), et dans la montée en puissance des projets multi-sources (comme ceux du programme SNDS, ou les consortiums européens type EHDEN, DARWIN EU, etc.).

5. Ce que REPEAT change dans nos pratiques

Adopter les principes REPEAT, c'est :

- **Expliciter clairement les choix méthodologiques** (au lieu de les sous-entendre) ;
- **Faciliter les relectures et validations** (internes, externes, réglementaires) ;
- **Améliorer la transparence scientifique**, élément de plus en plus scruté par les revues et les autorités.
- Facilite la discussion méthodologique entre partenaires de disciplines différentes.

REPEAT s'inscrit dans une dynamique plus large, portée notamment par la communauté OHDSI, pour améliorer la reproductibilité des études observationnelles.

Conclusion : une aide précieuse pour structurer les études en vie réelle

Que vous travailliez sur un registre institutionnel, un entrepôt hospitalier ou une base comme le SNDS, les outils de REPEAT vous permettent de **structurer votre raisonnement, documenter vos hypothèses, et mieux communiquer vos choix**.

Dans un contexte où la rigueur méthodologique est scrutée, ils sont un atout précieux. Et si, pour votre prochaine étude, vous utilisiez un schéma REPEAT dès le protocole ?

Ressources utiles :

- REPEAT initiative : repeatinitiative.org
- Book of OHDSI (chapitres sur la conception d'études observationnelles) : <https://ohdsi.github.io/TheBookOfOhdsi/>

Références

1. Wang SV, Sreedhara SK, Schneeweiss S, REPEAT Initiative. Reproducibility of real-world evidence studies using clinical practice data to inform regulatory and coverage decisions. Nat Commun. 2022;13:5126.
2. Laborey M, Maugard C, Benmerad M, Lajoinie A, Desamericq G, Chelbani L, et al. Arrêt au long cours des anti-ostéoporotiques chez les femmes ménopausées en France (SNDS-PMO). Journal of Epidemiology and Population Health. 2024;72:202329.