



Bon usage des bases SAE et spécificités 2022

Note méthodologique

La **Statistique annuelle des établissements de santé (SAE)** est une enquête administrative, exhaustive et obligatoire, réalisée par la DREES auprès des établissements de santé installés en France (métropole et DROM). Elle produit deux types de données, intégralement mises en ligne :

- Une base « administrative » restituant les déclarations des établissements (publiée fin juillet N+1) ;
- Une base « statistique » destinée aux études (publiée fin octobre N+1).

Cette note méthodologique vise à **rappeler les règles d'usage de ces deux types de bases**, et les précautions indispensables à prendre lorsque les données SAE sont mobilisées « en évolution ».

En outre, durant l'année 2023, la DREES a poursuivi des échanges avec les établissements de santé sur la qualité de leurs réponses sur les « lits en état d'accueillir des patients », notamment pendant la crise sanitaire. Ces travaux conduisent à identifier des **révisions sensibles des bases statistiques sur les années 2020 et 2021**, dont il faut tenir compte pour toute analyse de l'évolution des principaux agrégats qui peuvent être inférés de la SAE.

C'est en particulier le cas pour l'évolution du nombre de lits hospitaliers : elle connaît des révisions notables en 2020, mais plus encore en 2021.

Table des matières

A-	Bases administratives : données provisoires, en coupe, non révisables (« scellées »)	2
B-	SAE-Diffusion : mêmes précautions d'usage que les bases administratives	2
C-	Bases statistiques : fiabilisées pour les études, révisables.....	3
D-	Des révisions à apporter aux bases statistiques 2020 et 2021	4
E-	Mise à disposition d'un « fichier de remords » des bases statistiques.....	5

A- Bases administratives : données provisoires, en coupe, non révisables (« scellées »)

Les bases administratives de la SAE portant sur les données de l'année N sont intégralement mises en ligne chaque année au plus tard le 31 juillet de l'année N+1. Elles sont régies par l'arrêté SAE du 9 septembre 2010 qui conduit la DREES à mettre à disposition les données « telles que validées par les établissements ». Du fait de ce caractère « administratif », ces données sont considérées comme « scellées » une fois publiées. Elles ne sont jamais révisées.

D'avril à juin N+1, la DREES conduit une courte phase dite de « validation » des données saisies avec les établissements, en leur proposant d'étudier un certain nombre d'anomalies repérées automatiquement par un outil statistique. Mais cet échange est bref et ne permet pas d'expertiser et fiabiliser complètement les réponses des établissements. L'objectif reste de mettre à disposition les données brutes le plus rapidement possible. En particulier, à l'issue de la « validation », il demeure de la non réponse totale (certains répondants habituels ne valident pas leur SAE) ou partielle (certaines données habituellement renseignées ne le sont pas faute de temps).

La contrepartie de cette recherche de précocité de la donnée est donc qu'il demeure des fragilités dans les données, et notamment de la non-réponse biaisant la lecture des données « en évolution », notamment lorsque les données d'un ensemble d'établissements sont agrégées. Il convient donc d'utiliser avec prudence les données des bases administratives.

Les recommandations de bon usage de la DREES pour les bases administratives :

- Considérer ces données comme provisoires, dans l'attente des bases statistiques : dès que celles-ci sont disponibles, les bases administratives peuvent être considérées comme « caduques » ;
- Ne pas utiliser ces données pour calculer des évolutions avec les années précédentes, mais les considérer plutôt comme une « photographie » de l'année considérée.

Rappelons aussi le principe de « subsidiarité des sources » : les données d'activité de la SAE pré-remplies avec le PMSI ne doivent pas être utilisées pour suivre l'activité des établissements de santé, il convient de consulter directement les bases PMSI. La DREES procède ainsi pour ses propres publications : elle ne mobilise jamais la SAE pour suivre l'activité des champs MCO, SSR et HAD, mais mobilise le PMSI qui constitue la source de référence.

B- SAE-Diffusion : mêmes précautions d'usage que les bases administratives

Les données des bases administratives sont également restituées grâce à un site internet dénommé SAE-Diffusion ([Accueil - SAE Diffusion \(sante.gouv.fr\)](http://sae.diffusion.sante.gouv.fr)). Il permet de visualiser facilement les résultats de l'enquête bordereau par bordereau, en sélectionnant le contenu des résultats visualisés pour obtenir des résultats individuels (saisie de l'identifiant d'un établissement) ou agrégés selon des critères prédéfinis (région, département, catégorie, statut juridique, liste de numéros Finess, etc.).

Les données de SAE-Diffusion étant celles des bases administratives, les mêmes précautions d'usage s'appliquent :

- Considérer ces données comme provisoires, dans l'attente des bases statistiques : dès que celles-ci sont disponibles, les bases administratives peuvent être considérées comme « caduques » ;
- Ne pas utiliser ces données pour calculer des évolutions avec les années précédentes, mais les considérer plutôt comme une « photographie » sur l'année considérée.

C- Bases statistiques : fiabilisées pour les études, révisables

Contrairement aux bases administratives, qui doivent retranscrire exactement les réponses des établissements géographiques à la SAE telles que validées par eux suite à la phase de « validation », les bases statistiques permettent de réaliser, si nécessaire, des corrections de valeurs aberrantes, manquantes ou des corrections d'incohérences qui n'auraient pu être traitées pendant la phase de validation de la collecte.

À compter de la SAE 2018, la DREES apporte systématiquement une correction de la non-réponse partielle et totale des établissements encore en activité et qui répondaient les années précédentes. Les établissements qui n'ont pas validé leur questionnaire mais ont renseigné une partie des données (statut de l'enquête « en cours de saisie ») ainsi que ceux n'ayant pas saisi leur SAE (statut « non saisi ») sont ajoutés aux bases statistiques dès lors qu'ils sont identifiés comme étant encore actifs.

Les bases statistiques sont en outre élaborées plus tardivement que les bases administratives, ce qui peut laisser du temps à certains établissements pour compléter ou corriger leurs réponses après le scellement des bases administratives.

Ce travail d'amélioration des bases « en continu » se traduit par un écart parfois sensible entre les agrégats calculés à partir des bases administratives et ceux calculés à partir des bases statistiques. Pour exemple, le tableau ci-dessous illustre les écarts sur le nombre total de lits en état d'accueillir des patients au 31 décembre, suivant qu'est utilisée l'une ou l'autre des bases, pour les SAE 2018 à 2021.

Tableau 1 : nombre total de lits en état d'accueillir des patients au 31 décembre

	2018	2019	2020	2021
Bases administratives	395 693	392 193	386 807	382 587
Bases statistiques	395 449	391 788	386 903	382 804
Révision	-244	-405	96	217

Champ : lits de MCO, SSR, USLD et hospitalisation à temps plein de psychiatrie, France entière y compris le service de santé des armées.

Source : DREES, SAE 2018 à 2021, traitements DREES.

L'exemple ci-dessus prouve en outre que l'écart entre bases administratives et bases statistiques n'est pas stable dans le temps. Pour cette raison, les évolutions des agrégats ne sont pas les mêmes suivant qu'elles sont calculées avec les bases administratives ou avec les bases statistiques.

Ceci renforce la mise en garde de la DREES : les analyses en évolution ne doivent être conduites qu'avec les bases statistiques, qui intègrent le maximum d'informations et de retraitements.

Autre amélioration des bases « en continu », la DREES enregistre régulièrement des « remords » sur les bases statistiques antérieures à celle qui est en cours d'élaboration. En effet, durant chaque collecte de la SAE, lors de la phase de « validation » des données avec les établissements, il peut arriver que les établissements signalent une correction à apporter à leur réponse sur une ou plusieurs années antérieures, notamment lorsque des évolutions aberrantes sont repérées et sont imputables à une erreur déclarative sur la ou les années antérieures. De même, lors de la phase de retraitements statistiques conduisant à élaborer les bases statistiques, l'équipe SAE de la DREES peut également être amenée à ré-interroger la qualité des déclarations passées.

Dans les deux cas, ces besoins de corrections sont enregistrés dans un « fichier de remord » mobilisé par la DREES pour ses publications. De manière aperiodique, lorsque ses moyens le lui permettent, la DREES peut être amenée à réviser les bases statistiques du passé pour en améliorer la qualité en y

intégrant le contenu du « fichier de remords ». C'est ce qui a été fait en juillet 2022 avec la mise à disposition de versions corrigées des bases statistiques 2013 à 2020¹.

Enfin, le principe de « subsidiarité des sources » s'applique aussi : comme pour les bases « administratives », les données d'activité des bases « statistiques » qui sont pré-remplies avec le PMSI ne doivent pas être utilisées pour suivre l'activité des établissements de santé, il convient de consulter directement les bases PMSI (voir plus haut).

D- Des révisions à apporter aux bases statistiques 2020 et 2021

Durant la crise sanitaire, la collecte de la SAE a été compliquée, en raison de la pression à laquelle faisaient face les établissements répondants. Les collectes les plus concernées sont celles des SAE 2019 à 2021. La DREES a eu davantage à faire pour produire une base statistique aussi fiable que possible dans ce contexte, ce qui peut expliquer des révisions plus importantes en base statistique pour ces millésimes.

En outre, la DREES a poursuivi des travaux d'expertise complémentaires en continu. Ces travaux ont permis d'améliorer les données de la SAE 2019, dont la base statistique a été révisée à l'occasion de la publication méthodologique de juillet 2022 (voir plus haut).

Pour la qualité des données 2020 et 2021, les expertises se sont poursuivies en 2023. Des échanges avec plusieurs établissements ont permis d'établir que, sous la contrainte de moyens qu'a représenté la crise sanitaire, leurs réponses ont parfois été plus imprécises qu'habituellement, faute de temps disponible pour rechercher les bonnes informations. La collecte de la SAE 2022, réalisée dans un contexte un peu plus apaisé, a permis de relever davantage d'évolutions atypiques que lors d'une campagne classique.

Ces analyses ont conduit à alimenter le « fichier de remords » avec un nombre important de corrections portant sur les années 2020 et 2021. Pour exemple, le tableau ci-dessous illustre la somme des corrections à apporter au nombre total de lits en état d'accueillir des patients au 31 décembre, tel qu'il est actuellement retracé dans les bases statistiques des SAE 2018 à 2021.

Tableau 2 : besoins de corrections du nombre total de lits en état d'accueillir des patients au 31 décembre, par rapport aux bases statistiques de la SAE actuellement disponibles

	2018	2019	2020	2021
Total des corrections	-12	-7	-397	-1 803

Champ : lits de MCO, SSR, USLD et hospitalisation à temps plein de psychiatrie, France entière y compris le service de santé des armées.

Source : DREES, SAE 2018 à 2021, traitements DREES.

Cet exemple montre qu'une bonne partie des correctifs sur 2019 ont déjà été intégrés dans les bases corrigées mises à disposition en juillet 2022. En revanche, sur 2020 et plus encore sur 2021, les révisions sont importantes : une partie des réponses apportées en première intention s'est appuyée sur une définition de type « lits budgétés » ou « lits autorisés »². Ces concepts sont parfois plus facilement disponibles que le concept SAE, notamment dans les systèmes d'information des établissements de santé. Certains d'entre eux les ont mobilisés « dans l'urgence » pour répondre à la SAE, dans un contexte de forte tension sanitaire.

La prise en compte de ces corrections conduit bien évidemment à des évolutions du nombre de lits qui s'écartent, elles aussi, de celles qui peuvent être calculées avec les bases statistiques disponibles.

¹ Voir Delaporte et alli, DREES Méthodes n° 2, juillet 2022 : [La base ARES : agrégats régionaux sur les établissements de santé - Méthodologie de constructions de séries régionales de la SAE depuis 2013 | Direction de la recherche, des études, de l'évaluation et des statistiques \(solidarites-sante.gouv.fr\)](#)

² C'est-à-dire sans écarter les lits non disponibles du fait du manque de personnel, alors que la SAE le requiert.

E- Mise à disposition d'un « fichier de remords » des bases statistiques

La mise à disposition de bases statistiques intégrant les « remords » représente un travail coûteux. Actuellement, les équipes DREES dédiées à la SAE sont fortement occupées par la préparation de la SAE 2023, qui se cumule avec la refonte de la SAE (prévue pour la SAE 2024) qu'implique la réforme des autorisations de soins (entrée en vigueur à partir de juin 2023). La DREES ne va donc pas pouvoir procéder à cette mise à jour à court terme.

Pour offrir à ses utilisateurs une solution fiable pour tenir compte des corrections, la DREES va mettre à disposition son « fichier de remords », accompagné d'une courte documentation. Ceci permettra aux utilisateurs qui souhaitent reconstituer des « bases statistiques corrigées » de le faire dès la fin de l'année 2023.

Par ailleurs, la DREES va publier en décembre 2023 une étude portant sur les premiers résultats de la SAE 2022, documentant notamment les évolutions du nombre de lits en tenant compte du « fichier de remords ».

L'objectif de moyen terme reste, pour la DREES, d'intégrer aux bases statistiques les « remords » et de mettre à disposition une version corrigée des bases (comme en juillet 2022).

En savoir plus sur la SAE :

[La Statistique annuelle des établissements \(SAE\) | Direction de la recherche, des études, de l'évaluation et des statistiques \(solidarites-sante.gouv.fr\)](#)