

# Les bases statistiques SAE 2022 – Documentation

## Sommaire :

A- Accès aux données et contenu des dossiers .....	1
B- Spécificité de l'exploitation de la base SAS .....	3
C- Rappels et précautions d'usage pour l'exploitation des bases .....	3
D- Contenu des corrections .....	4
E- Ruptures de séries .....	6
F- Tables fournies .....	6
a- Données d'activité et de capacités .....	6
b- Données sur le personnel médical rémunéré .....	8
c- Données sur le personnel non médical rémunéré .....	9
d- Nouvelles bases mises à disposition .....	9

La **Statistique annuelle des établissements de santé (SAE)** est une enquête administrative, exhaustive et obligatoire auprès des établissements de santé installés en France (métropole et DOM), y compris les structures qui ne réalisent qu'un seul type d'hospitalisation ou qui ont une autorisation pour une seule activité de soins. Les données sont collectées par la DREES (Direction de la recherche, des études, de l'évaluation et des statistiques), direction du ministère des solidarités et de la santé.

Elle produit deux types de données :

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

## A- Accès aux données et contenu des dossiers

Pour consulter les bases statistiques SAE 2022, il faut dé-zipper le dossier contenant les tables. Pour cela, il faut cliquer sur *SAE 2022 – Bases statistiques.zip*, et enregistrer le fichier dans un répertoire. Il faut ensuite double-cliquer sur ce dossier, et cliquer sur « Extraire ».

Une fois dé-zippé, le répertoire « SAE 2022 Bases statistiques » contient 4 sous-répertoires :

- un sous-répertoire « Bases statistiques » ;
- un sous-répertoire « Fichier de transport – bases SAS » ;
- un sous-répertoire « Documentation ».

**Le sous-répertoire « Bases statistiques »** contient deux dossiers :

- un dossier « Base SAS » avec les différentes tables redressées au format SAS ;
- un dossier « Base CSV » avec les différentes tables redressées au format CSV.

Chacun de ces dossiers contient l'intégralité des bases (dans deux formats différents). Lorsqu'elles sont suffixées par la lettre « a », en fin de nom de fichier, cela indique que la base statistique est identique à la base administrative, contrairement au cas où elles sont suffixées par la lettre « r », qui indique que la base statistique est redressée.

**Le sous-répertoire « CAPACT\_PM\_PNM »**, qui contient les tables de séries longues, est supprimé et sera prochainement mis à jour et regroupé dans un seul dossier.

**Le sous-répertoire « Fichier de transport – bases SAS »** contient les *fichiers de transport* pour les bases statistiques de la SAE 2022. Les tables et formats SAS sont en effet créés en 32bits à la DREES. Dans le cas où votre système d'exploitation (et SAS) est en 64bits, cela empêche l'utilisation des formats. Les *fichiers de transport* devraient vous permettre de pallier ce problème et d'utiliser les formats. Vous trouverez la procédure d'utilisation à cette adresse :

<http://support.sas.com/kb/44/047.html>.

**Le sous-répertoire « Documentation »** contient :

- les formats SAS ;
- l'aide au remplissage de l'enquête (SAE\_AR\_2022.pdf). Ce document est fourni aux établissements enquêtés au moment de la collecte. Il contient des informations sur les concepts ainsi que des consignes pour la réponse à l'enquête ;
- les nomenclatures de l'enquête (Nomenclatures\_2022.xls), incluant les nomenclatures spécifiques à la diffusion. Ce document est fourni aux établissements enquêtés au moment de la collecte ;
- le fichier SAE2022\_Dictionnaire\_variables\_par\_bordereaux.xlsx : il établit, pour chaque bordereau de la SAE, une correspondance entre le visuel du questionnaire et les variables des fichiers :
  - l'onglet **Recap TABLES** recense l'ensemble des tables diffusées avec leur bordereau correspondant ;
  - l'onglet **Légende** est une aide à la lecture ;
  - les onglets **ID1** à **Q24** indiquent, case par case, le nom de chaque variable pour chacun des bordereaux de l'enquête ;
  - l'onglet **SYGEN** reprend, case par case, le nom de chacune des variables de la table. La table SYGEN correspond à la fiche de synthèse « Activité générale et activités autorisées » (synthèse des bordereaux). Cet onglet sert aussi de correspondance pour la base HPR (Hôpitaux de proximité) ;
  - l'onglet **GHT** donne la liste des variables présentes dans la base agrégée au niveau groupement hospitalier de territoire (GHT).
- le fichier Liste Variables.xlsx : il reprend les informations du fichier précédent (SAE2022\_Dictionnaire\_variables\_par\_bordereaux.xlsx) mais sous forme de tableau. Il contient, pour chaque table SAS, la liste des variables ainsi que leur libellé (et éventuellement la modalité associée), et la case du questionnaire à laquelle elle correspond. Pour certaines variables, il n'y a pas de correspondance entre celle-ci et une case du questionnaire : il s'agit de variables créées a posteriori. Elles sont repérées par l'indication « (Var SAS) » dans la colonne Bordereau ;
- le fichier Liste GHT\_2022.xlsx : pour chaque GHT, il liste les entités juridiques qui le compose ;
- **Une note méthodologique sur les spécificités des bases de la SAE 2022, ainsi que leur bon usage**
- ce document.

Vous pourrez trouver d'autres informations concernant la SAE sur :

- le site internet de la DREES : <https://drees.solidarites-sante.gouv.fr/sources-outils-et-enquetes/00-la-statistique-annuelle-des-etablissements-sae>
- la FAQ de la SAE : [Accueil | Aides et FAQ enquête SAE \(sante.gouv.fr\)](#)

Pour toute question, vous pouvez nous écrire à l'adresse : [DREES-INFOS@sante.gouv.fr](mailto:DREES-INFOS@sante.gouv.fr)

## B- Spécificité de l'exploitation de la base SAS

Lors de l'utilisation des bases SAS, vous pouvez choisir d'afficher les tables avec ou sans formats. Pour utiliser les formats de la SAE, il faut copier ces lignes de code au début de votre programme SAS :

```
libname fSAE "chemin du dossier où se trouve formats.sas7bcat";  
options fmtsearch= (fSAE);
```

Si lors d'une instruction vous ne voulez pas utiliser les formats, il faut ajouter à votre procédure l'instruction suivante : `format _all_;`

Par exemple, pour créer une version de table sans formats :

```
data table_sans_formats;  
set table_avec_formats;  
format _all_;  
run;
```

Si vous ne voulez pas du tout utiliser les formats, il faut ajouter cette ligne de code : `options nofmterr;`

## C- Rappels et précautions d'usage pour l'exploitation des bases

### a. Le champ de l'enquête et les tables d'identification

Le champ de la SAE est constitué des établissements de santé enregistrés dans le répertoire Finess et qui ont exercé en 2022 une activité de soins autorisée par leur ARS.

Le recueil de la SAE s'effectue sur un site de collecte en ligne. Cette enquête permet de récupérer des informations sur les capacités, l'activité, les plateaux techniques et le personnel des établissements de santé. Le taux de réponse de la SAE 2022 (pourcentage d'établissements enquêtés ayant transmis et validé leurs données) est de 99,9 %.

Dans le questionnaire et sur le site de collecte, deux bordereaux contiennent des données d'identification sur les établissements de santé : les bordereaux ID1 et ID2. Ces informations sont en partie importées du répertoire Finess.

Dans les bases CSV et SAS fournies, deux tables, réorganisées par rapport aux bordereaux du questionnaire, contiennent les données d'identification des établissements :

- La table **ID\_2022** contient les données d'identification des entités directement interrogées par la SAE : données de l'établissement issues des bordereaux ID1 et ID2 (raison sociale, statut juridique, catégorie d'établissement, etc.) et données propres à la collecte de l'enquête (vague d'interrogation, statut de validation).
- La table **IDB\_2022** contient les données d'identification des entités non interrogées directement, à savoir les entités juridiques (EJ) non interrogées (EJ publiques de groupe 02, EJ des centres privés de dialyse de groupe 06 et les CLCC ou EJ de radiothérapie de groupe 08) et les établissements géographiques non interrogés directement (USLD et psychiatrie).

Les établissements réalisant uniquement une activité de santé mentale font l'objet d'une interrogation groupée : un unique établissement par département est désigné pour être interrogé et pour répondre pour l'ensemble des structures de psychiatrie de la même EJ situées dans son département. Dans le cas particulier des établissements privés ne faisant que de la psychiatrie et tous situés dans le même département, c'est l'EJ de ces établissements qui est interrogée et qui répond pour l'ensemble des établissements de psychiatrie qui lui sont rattachés. Dans certains cas, cette EJ n'est pas située dans le même département ou dans la même région que les établissements qui lui sont rattachés.

Les données collectées sont localisées au lieu d'implantation de l'établissement répondant, y compris lorsque la réponse est collectée au niveau de l'EJ.

Depuis 2016, deux nouvelles variables, REG\_DIFF et DEP\_DIFF, sont disponibles dans la table ID diffusée sur Data.Drees, permettant de rendre compte du lieu réel de l'activité dans les cas particuliers où établissements et EJ ne sont pas localisés au même endroit.

Dans le cas particulier des établissements privés ne faisant que de la psychiatrie et tous situés dans le même département, les variables REG et DEP sont relatives à la localisation de l'EJ répondante et les variables REG\_DIFF et DEP\_DIFF se rapportent aux établissements géographiques dépendant de cette EJ.

#### b. Spécificités d'exploitation de certaines tables

Les tables de la SAE sont créées de telle sorte qu'une ligne correspond aux données d'un établissement dans son ensemble.

Cependant, dans le cas de bordereaux contenant des tableaux de grande taille, afin de faciliter la lecture des tables et d'éviter d'avoir trop de variables, une variable de la table SAS/CSV correspondra à plusieurs cases du questionnaire : on crée alors une variable « modalité » qui distingue chaque case concernée. Ainsi, une ligne ne correspondra pas aux données d'un établissement dans son ensemble mais à un croisement « établissement » x « sous-modalité ».

Exemples :

- dans la table concernant la psychiatrie (bordereau PSY), on distingue la psychiatrie générale, la psychiatrie infanto-juvénile et la psychiatrie pénitentiaire ;
- dans les tables de personnel rémunéré par les établissements, la distinction porte sur le code « métier » (bordereaux Q20 et Q23) ou sur la discipline d'équipement (bordereaux Q21 et Q24) ;
- dans la table sur les urgences (bordereau URGENCES), on distingue les urgences générales et pédiatriques.

Les différentes modalités sont rappelées dans l'onglet « Recap TABLES » du dictionnaire des variables.

Pour ces tables, il existe donc différentes manières d'agrégier les données : somme sur tous les établissements par sous-modalité, somme de toutes les sous-modalités pour chaque établissement, etc.

**Néanmoins, il faut tout de même faire attention à certaines tables dans lesquelles il existe une sous-modalités « Total » ou des sous-totaux (par exemple dans les tables PSY ou SSR). Dans ce cas, il faut choisir entre deux options : utiliser l'agrégat déjà calculé, ou recalculer le total en excluant les totaux et sous-totaux existants.**

Exemple : dans la table concernant la psychiatrie (bordereau PSY), la variable DIS (discipline de psychiatrie) peut prendre quatre modalités : GEN (psychiatrie générale), INF (psychiatrie infanto-juvénile), PEN (psychiatrie en milieu pénitentiaire) et TOT (total psychiatrie).

La modalité TOT est directement calculée sur le site de collecte et correspond à la somme GEN+INF+PEN. Pour éviter les doubles-comptes dans les agrégats, il faut donc choisir entre utiliser uniquement la modalité TOT ou utiliser uniquement les trois autres modalités, GEN/INF/PEN.

## D- Contenu des corrections

Les corrections de la SAE 2022 sont de deux ordres : harmonisation et redressement des données.

**L'harmonisation des données** concerne la mise en cohérence des données entre bordereaux (cohérence entre les bordereaux Q20/Q21 et Q23/Q24 particulièrement, dont les ETP salariés et les effectifs de libéraux doivent être égaux), ainsi que l'harmonisation des autorisations avec le référentiel

ARHGOS (notamment pour les autorisations des maternités, des urgences, en soin de suite et réadaptation, etc.).

Attention, pour la SAE 2022, il peut exister des écarts entre les bordereaux Q23 et Q24 sur les ETP salariés et les effectifs libéraux. En effet, les directeurs de soins infirmiers et non infirmiers ont été regroupés dans une seule catégorie dans le bordereau Q23 et non dans Q24, ce qui empêche la mise en cohérence des données Q23 et Q24 pour ces catégories.

**Le redressement des données** : 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 des données collectées, 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, nous apportons également une correction de la non réponse partielle et totale. 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 en « non saisi ») sont ajoutés aux bases statistiques.

- Pour les établissements « en cours de saisie », les données complétées par l'établissement sont récupérées et intégrées aux bases statistiques, et les quelques données manquantes (notamment les capacités d'accueil et les données de personnel) sont ajoutées en utilisant les données N-1 :
- Pour les établissements n'ayant pas saisi leur SAE ou l'ayant validée mais n'ayant pas rempli un bordereau qu'ils auraient dû remplir (parce qu'ils ont l'autorisation par exemple), nous récupérons les données PMSI pour l'activité et nous imputons les données manquantes en utilisant, le plus souvent, les données de l'année précédente (pour les bordereaux et variables les plus importants). Par la suite, la dénomination « non réponse totale par bordereau » correspond à la correction de cette non réponse totale ou partielle.

Après intégration de ces établissements, le taux de couverture de l'enquête est de 99,9 % (contre 99,9 % de taux de réponse en base administrative).

## E- Ruptures de séries

La SAE a subi une refonte très importante en 2013, sur laquelle vous trouverez des informations sur le site de la DREES ou dans la documentation relative aux bases administratives. **En raison de cette refonte, les évolutions 2012-2013 sont difficiles à analyser.**

Pour faciliter l'exploitation des données, nous avons mis à disposition des tables qui regroupent les années 2000 à 2022. Cela ne signifie pas que l'évolution annuelle est interprétable : sur un grand nombre de variables, les concepts ont changé. Ainsi, les données d'activité importées du PMSI ne se basent pas exactement sur les mêmes notions que l'activité qui était déclarée dans la SAE précédemment : les évolutions 2012-2013 restent difficiles à analyser.

**Attention** : depuis 2011, les données incluent Mayotte.

**Les bordereaux Q20 à Q24 ont été dépriorisés pour la SAE 2020 suite aux difficultés des établissements à répondre aux différentes questions. Ils ne sont donc pas présents dans les bases statistiques de la SAE 2020. De ce fait, les séries longues PM\_01\_22 et PNM\_01\_22 ne comportent pas l'année 2020 : elles passent directement de 2019 à 2021.**

## F- Tables fournies

### a- Données d'activité et de capacités

#### Pour l'année 2022

Depuis la SAE 2019, les bordereaux PHARMA et BIO sont interrogés alternativement. Pour la SAE 2022, les bordereaux PHARMA et DOULEUR ont été collectés. Pour la SAE 2023, les bordereaux BIO et PALIA seront collectés.

Les tables auxquelles des corrections ont été apportées, par rapport aux bases administratives, sont listées ci-dessous en précisant les modifications effectuées :

- La table **ID\_2022R** : elle est au même format que la table ID\_2022 des bases administratives, mais elle contient des corrections sur la variable ETAT\_SAISIE ainsi que sur le champ.
- La table **PN\_2022R** : elle est au même format que la table PN\_2022 des bases administratives, mais elle contient des corrections sur l'activité (prise en charge des personnes en situation de précarité, prise en charge sanitaire des personnes détenues, addictologie).
- La table **FILTRE\_2022R** : elle est au même format que la table FILTRE\_2022 des bases administratives, mais elle contient des corrections sur les activités et équipements disponibles. Elle est également corrigée de la non-réponse totale au bordereau.
- La table **MCO\_2022R** : elle est au même format que la table MCO\_2022 des bases administratives, mais elle contient des corrections sur les capacités, le nombre de journées exploitables et sur l'activité. Elle est également corrigée de la non-réponse totale au bordereau.
- La table **MCOGER\_2022R** : elle est au même format que la table MCOGER\_2022 des bases administratives, mais elle contient des corrections sur les capacités (lits et places) et sur l'activité. Elle est également corrigée de la non-réponse totale au bordereau.
- La table **PSY\_2022R** : elle est au même format que la table PSY\_2022 des bases administratives, mais elle contient des corrections sur les capacités et l'activité (prise en charge à temps complet, partiel et ambulatoire). Elle est également corrigée de la non-réponse totale au bordereau.
- La table **PSY2\_2022R** : elle est au même format que la table PSY2\_2022 des bases administratives, mais contient des corrections sur les déclarations de services médico-psychologiques régionaux et les ETP associés dans le personnel concourant à l'activité.

- La table **SSR\_2022R** : elle est au même format que la table SSR\_2022 des bases administratives, mais elle contient des corrections sur les autorisations ainsi que sur le nombre de lits, places, journées et séjours. Elle est également corrigée de la non-réponse totale au bordereau.
  - La table **USLD\_2022R** : elle est au même format que la table USLD\_2022 des bases administratives, mais elle contient des corrections sur l'activité. Elle est également corrigée de la non-réponse totale au bordereau.
  - La table **BLOCS\_2022R** : elle est au même format que la table BLOCS\_2022 des bases administratives, mais elle contient des corrections sur le nombre de sites, de salles, de salles de surveillance et de postes de réveil. Elle est également corrigée de la non-réponse totale au bordereau.
  - La table **IMAGES\_2022R** : elle est au même format que la table IMAGES\_2022 des bases administratives, mais elle contient des corrections sur le nombre d'équipement et l'activité, notamment du fait de la correction de la non-réponse totale au bordereau.
  - La table **PERINAT\_2022R** : elle est au même format que la table PERINAT\_2022 des bases administratives, mais elle contient notamment des corrections sur les types de maternités et les capacités. Elle est également corrigée de la non-réponse totale au bordereau.
  - La table **PERINAT\_P\_2022R** : elle est au même format que la table PERINAT\_P\_2022 des bases administratives, mais elle contient notamment des corrections sur le personnel concourant à l'activité
- 
- La table **CHIRCAR\_2022R** : elle est au même format que la table CHIRCAR\_2022 des bases administratives, mais elle contient des corrections sur les autorisations et le champ.
  - La table **DIALYSE\_2022R** : elle est au même format que la table DIALYSE\_2022 des bases administratives, mais elle contient des corrections sur les capacités et le nombre de séances notamment. Elle est également corrigée de la non-réponse totale au bordereau.
  - La table **DIALYSE\_P\_2022R** : elle est au même format que la table DIALYSE\_P\_2022 des bases administratives, mais elle contient des corrections sur le personnel concourant à l'activité
- 
- La table **CANCERO\_2022R** : elle est au même format que la table CANCERO\_2022 des bases administratives, mais elle contient des corrections sur l'activité et le champ. Elle est également corrigée de la non-réponse totale au bordereau.
  - La table **URGENCES\_2022R** : elle est au même format que la table URGENCES\_2022 des bases administratives, mais elle contient des corrections sur les autorisations, notamment suite à la correction de la non-réponse totale au bordereau.
  - La table **URGENCES2\_2022R** : elle est au même format que la table URGENCES2\_2022 des bases administratives, mais elle contient des corrections sur le nombre de passages aux urgences générales notamment. Elle est également corrigée de la non-réponse totale au bordereau.
  - La table **SMURSAMU\_2022R** : elle est au même format que la table SMURSAMU\_2022 des bases administratives, mais elle contient des corrections sur les autorisations SMUR et SAMU, sur l'activité de SMUR et de SAMU et sur les carences ambulancières. Elle est également corrigée de la non-réponse totale au bordereau.
  - La table **REA\_2022R** : elle est au même format que la table REA\_2022 des bases administratives, mais elle contient des corrections sur les capacités et l'activité notamment. Elle est également corrigée de la non-réponse totale au bordereau.

b- Données sur le personnel médical rémunéré

La table **Q20\_2022R** (personnel médical rémunéré par spécialité exercée) : elle est au même format que la table Q20\_2022 des bases administratives, mais elle contient des corrections sur les ETP et les effectifs rémunérés, notamment suite à la correction de la non-réponse totale au bordereau et à la mise en cohérence avec le bordereau Q21.

La table **Q21\_2022R** (personnel médical rémunéré par groupe de disciplines d'équipement) : elle est au même format que la table Q21\_2022 des bases administratives, mais elle contient des corrections

sur les ETP et les effectifs rémunérés, notamment suite à la correction de la non-réponse totale au bordereau et à la mise en cohérence avec le bordereau Q20.

La table **Q22\_2022R** (internes et faisant fonction d'internes) : elle est au même format que la table Q22\_2022 des bases administratives, mais elle contient des corrections sur les effectifs.

Pour rappel, suite à la dépriorisation des bordereaux de personnel Q20 à Q24 en cours de collecte de la SAE 2020, ces données – non vérifiées par la DREES – ne sont pas diffusées dans les bases statistiques 2020.



c- Données sur le personnel non médical rémunéré

La table **Q23\_2022R** (personnel non médical rémunéré) : elle est au même format que la table Q23\_2022 des bases administratives, mais elle contient des corrections sur les effectifs et ETP rémunérés, notamment suite à la correction de la non-réponse totale au bordereau.

La table **Q24\_2022R** (personnel non médical rémunéré par disciplines d'équipement) : elle est au même format que la table Q24\_2022 des bases administratives, mais elle contient des corrections sur les ETP de personnel non médical rémunéré, notamment suite la correction de la non-réponse totale au bordereau et à la mise en cohérence avec le bordereau Q23.

Attention, des écarts pour les ETP des salariés et les effectifs et libéraux entre les bordereaux Q23 et Q24 subsistent pour la catégorie des directeurs de soins (Q23) / directeurs de soins infirmiers et non infirmiers (Q24).

Pour rappel, suite à la dépriorisation des bordereaux de personnel Q20 à Q24 en cours de collecte de la SAE 2022, ces données – non vérifiées par la DREES – ne sont pas diffusées dans les bases statistiques 2020.

d- Nouvelles bases mises à disposition

Depuis la SAE 2016, les tables SYGEN\_AAAAR et GHT\_AAAAR sont mises à disposition :

- La table **SYGEN\_2022R** reprend les informations contenues dans la partie « Activité générale et activités autorisées » du bordereau SYGEN. Elle est au même format que la table SYGEN\_2021 des bases administratives, mais elle est mise à jour des redressements apportés aux données des différents bordereaux du questionnaire.

Pour les bordereaux interrogés en alternance une année sur deux (bordereaux BIO et PHARMA), leurs variables restent présentes dans la table SYGEN (pour faciliter les alimentations automatiques des systèmes d'information), mais seules celles du bordereau interrogé pour l'année considérée sont alimentées, tandis que celles du bordereau non interrogé restent à vide. Pour la SAE 2022, seules les données de PHARMA sont donc alimentées avec les données collectées.

- La table **GHT\_2022R** contient des données agrégées sur le contour des groupements hospitaliers de territoires (GHT) au 31 décembre tels que le nombre de lits, de places, de journées et de séjours et des informations sur les équipements, les plateaux techniques et le personnel des établissements de santé. Elle est au même format que la table GHT\_2022 des bases administratives, mais elle est mise à jour des redressements apportés aux données des différents bordereaux du questionnaire.

Pour les bordereaux interrogés en alternance (bordereaux BIO et PHARMA), là encore les variables des deux bordereaux sont présentes dans la table GHT, mais seules celles du bordereau interrogé pour l'année considérée sont alimentées, tandis que celles du bordereau non interrogé restent à vide.

Depuis la SAE 2018, la table HPR\_AAAAR est mise à disposition. La table **HPR\_2022R** contient les informations issues de la table SYGEN\_2022R sur le contour des hôpitaux de proximité uniquement, selon la liste initialement définie par l'arrêté de mai 2019 et mise à jour grâce au référentiel Finess.