



Swiss National Cohort (SNC) et statistique démographique longitudinale (DVS)

Intégration de la « Swiss National Cohort » (SNC) dans la statistique démographique longitudinale (DVS)

Neuchâtel, janvier 2023

Éditeur:	Office fédéral de la statistique (OFS)
Renseignements:	info.pop@bfs.admin.ch, tél. 058 467 25 25
Rédaction:	Nathalie Monnet, POP
Contenu:	Nathalie Monnet, POP
Domaine:	01 Population
Langue du texte original:	Français
Traduction:	Services linguistiques de l'OFS

Concept de mise en page:	Section DIAM
Download :	www.statistique.ch
Copyright:	OFS, Neuchâtel 2023 Reproduction autorisée seulement avec l'accord préalable des auteurs

Table des matières

1	Etat actuel	3
2	Intégration SNC-DVS	3
3	Réserves concernant l'intégration SNC-DVS	4
4	Conclusion	5

1 Etat actuel

La « Swiss National Cohort » (SNC) est une base de données qui a été produite par l'Institut d'épidémiologie, de biostatistique et de prévention de l'Université de Zurich et de l'institut de médecine sociale et préventive de l'Université de Berne (ISPM - organe d'exécution du projet). Le projet SNC permet de lier les individus présents dans la statistique de la population et des ménages en 2010-2011 aux données des recensements 1990 et 2000 grâce à un *appariement probabiliste*¹.

Depuis janvier 2022, l'OFS est en charge des données SNC, la base de données comprend :

- L'extrait SNC qui contient tous les individus présents dans les recensements 1990 (VZ 1990), 2000 (VZ 2000) et la statistique de la population et des ménages (STATPOP) de 2010 à 2019. Une partie des variables provient de plusieurs autres statistiques: la statistique du mouvement naturel de la population (BEVNAT) pour la date de décès et les causes de décès, le relevé structurel (RS) et la statistique des bâtiments et des logements (StatBL), l'enquête suisse sur la santé (ESS) 1992 ainsi qu'un indice socio-économique créé par l'ISPM.
- Un fichier de clés qui permet d'utiliser le lien probabiliste entre le numéro AVS à 13 positions (NAVS) des individus présents dans STATPOP et les identifiants des recensements.

La statistique démographique longitudinale (DVS) contient l'ensemble des mouvements à partir de 2011 des personnes présentes dans au moins un fichier STATPOP ainsi qu'une table des attributs avec les variables qui ne changent pas dans le temps (date de naissance, lieu de naissance, lien sur les parents/enfants). De plus, des effectifs au 31 décembre ont été créés sur la base de la table des attributs et celles de mouvements pour aider les utilisateurs dans leurs travaux d'analyse.

D'un côté, la SNC contient des états de la population en 1990, 2000, et de 2010 à 2019. De l'autre côté, la DVS couvre la partie à partir de 2010, sauf pour les causes de décès ainsi que les variables du relevé structurel². Intégrer les informations des recensements 1990-2000 dans la statistique démographique longitudinale permettrait d'allonger la biographie des individus (par exemple, en ajoutant l'information sur leur commune de résidence en 1990 et 2000) ainsi qu'ajouter des individus qui ne sont plus présents à partir de 2010. Les étapes nécessaires sont détaillées dans la deuxième section.

2 Intégration SNC-DVS

Pour l'intégration des informations des recensements 1990 et 2000 dans la DVS, les étapes suivantes seraient nécessaires :

1. S'assurer de la qualité de l'appariement probabiliste de la SNC, par exemple, en (1) mettant à jour les NAVS dans les données 1990 et 2000 ; (2) regardant combien de personnes sont revenues après 2012 en Suisse (si absentes en 2010-2011) et qui auraient pu être présentes en 1990 et 2000, car le lien n'a pas été établi dans le projet SNC. D'autres techniques pourraient également être utilisées afin de tester les liens.
2. Reprendre la date de naissance et le lieu de naissance dans la table des attributs (DVS) pour les personnes qui sont dans les recensements 1990 et 2000.
3. Harmoniser la définition du lieu de domicile : les recensements sont basés sur la population au domicile économique, alors que la DVS sur la population résidente au domicile principal.
4. Créer les mouvements entre 1990 et 2010 : seules les naissances, les décès en Suisse et les acquisitions de nationalité suisse pourraient être créés mais sans aucune information sur les autres variables au moment de l'évènement (commune de domicile, nationalité, état civil, etc.). Les migrations, les déménagements, les changements de statut, les changements de type de population, les changements d'état civil, etc. ne seraient pas faisables.

¹ Les recensements ne disposent pas de numéros AVS mais d'un identifiant unique pour chaque individu, sans que celui-ci soit le même pour chaque recensement.

² Ces informations peuvent être obtenues par une demande d'appariement entre la DVS et les divers fichiers qui ont permis la création du projet SNC (causes de décès, statistique des bâtiments et logements, relevé structurel, etc.) : [Appariement de données pour le compte de tiers | Office fédéral de la statistique \(admin.ch\)](https://www.bfs.admin.ch/bfs/fr/home/accueil/actualites/actualites/2022/01/appariement-de-donnees-pour-le-compte-de-tiers.html).

3 Réserves concernant l'intégration SNC-DVS

Malgré l'apport biographique pour les années 1990 et 2000, il y a plusieurs arguments en défaveur de l'intégration des données SNC dans la statistique démographique longitudinale :

1. Peu de demandes : en 2022, uniquement quatre demandes concernent les données SNC + VZ + STATPOP (demandes d'appariement). Rares sont les demandes qui nécessitent des données avant 2010 et nous pensons que ce qui est à disposition actuellement est suffisant. En effet, les données SNC sont de bonne qualité en elles-mêmes, c'est-à-dire qu'elles sont suffisantes pour les demandes reçues.
2. Différences conceptuelles : les recensements comprennent des définitions différentes du type de population (population économique). La statistique démographique longitudinale, quant à elle, est basée sur la population résidente au domicile principal. De plus, la SNC ne comprend que deux états (1990 et 2000) alors que la DVS est une statistique d'événements (tous les mouvements à partir de 2011). Un important travail de cohérence serait nécessaire, sans y apporter beaucoup d'informations supplémentaires. En effet, les années 1991-1999 et 2001-2009 ne sont pas disponibles (les liens avec BEVNAT par exemple n'ont pas été établis).
3. Incertitudes sur la qualité : l'appariement probabiliste effectué par l'ISPM afin de retrouver le NAVS des individus présents dans les recensements 1990 et 2000 se base sur les individus présents dans STATPOP entre 2010-2011. Un potentiel problème provient des individus présents dans un des recensements et dans STATPOP sauf entre 2010-2011. Par exemple, si une personne était en Suisse en 2000 et 2014 mais pas en 2010-2011, aucun lien n'a été établi pour cette personne et elle apparaît comme deux personnes distinctes dans la SNC.

D'autres problèmes de qualité sont liés à la difficulté d'apparier des individus avec un écart de 10 ans, en particulier pour les personnes entre 15-25 ans en 1990 ou en 2000. En effet, la probabilité de changement d'état civil, de commune de domicile, etc. est plus élevée pour cette catégorie que d'autres classes d'âge. Les données montrent un pic de liens manquants entre les identifiants des recensements et des NAVS pour les personnes nées dans les années 1980. Ces problèmes de qualité ne permettent pas une concordance optimale avec le travail fait pour la statistique démographique longitudinale, qui est basée sur des liens déterministes en utilisant le numéro AVS.

4 Conclusion

En conclusion, les arguments avancés sont en défaveur de l'intégration de la SNC dans la DVS. Actuellement, la DVS ne couvre pas encore toutes les thématiques couvertes par les données STATPOP. Le travail de création d'un produit complet est dans un premier temps la priorité pour la statistique démographique longitudinale. De plus, toutes les informations de l'intégration de la SNC dans DVS sont déjà disponibles en l'état actuel, comme le montre le schéma ci-dessous, en utilisant le fichier des clés. Ne pas intégrer la SNC dans la DVS n'empêche pas la mise en lien entre les deux fichiers. En effet, les deux fichiers ne sont pas basés sur les mêmes types de liens (SNC selon un appariement probabiliste, DVS selon un lien déterministe basé sur le numéro AVS). Selon les réserves émises ci-dessus, il n'est pas judicieux de prendre des ressources pour un travail qui ne semble pas apporter une plus-value substantielle à la situation actuelle.

