

Construction d'une base de données longitudinales reposant sur l'appariement de registres administratifs

Philippe Wanner et Ilka Steiner¹

Résumé

Cette communication décrit la constitution d'une base de données élaborée en Suisse en vue de l'analyse de la migration et de l'intégration structurelle de la population d'origine étrangère. Cette base de données repose sur différents registres (registre des habitants, des assurances sociales, du chômage) et enquêtes et couvre 15 années (1998 à 2013). Des informations sur le statut migratoire et les caractéristiques socioéconomiques sont ainsi disponibles pour près de 4 millions d'étrangers ayant séjourné en Suisse entre 1998 et 2013. Cette base de données est le fruit d'une collaboration entre l'Office fédéral de la statistique et des chercheurs du NCCR On the Move.

Mots Clés : Appariement, Registres de population, Migration, Approches longitudinales.

1. Introduction

Au cours des dernières décennies, les migrations internationales ont joué un rôle essentiel dans la croissance démographique de nombreux pays industrialisés, et suscité de nombreux débats liés à l'intégration économique et sociale des migrants. La Suisse n'échappe pas à cette situation, puisque sa population dénombre près de 25% de personnes de nationalité étrangère et un tiers de personnes nées à l'étranger (Wanner, 2014). Principalement due à l'activité économique, cette migration a été, jusqu'au début des années 1970, marquée par le principe de rotation des travailleurs issus de pays voisins (principalement l'Italie), faiblement qualifiés, qui arrivent en Suisse pour un séjour temporaire et pour y exercer une activité professionnelle dans des secteurs tels que la construction, l'agriculture ou l'hôtellerie et la restauration. Puis, le séjour des migrants s'est allongé, les motifs à l'origine des flux migratoires sont devenus plus nombreux, les pays d'origine se sont diversifiés, et il en fut de même pour les caractéristiques professionnelles des migrants. De cette diversification de la migration résulta le besoin de mieux comprendre les comportements de vie et d'intégration de populations aux profils plus variés, autant d'un point de vue socioéconomique que culturel.

Or, le système statistique suisse a été pris par surprise par la transformation des flux migratoires et les données existantes, de nature transversale, peinent à saisir la complexité des flux migratoires et des processus d'intégration. Des questions aussi centrales que, par exemple, quelle est la part des migrants qui finissent par quitter la Suisse et de ceux qui y restent, ou quelle est la proportion des migrants qui se naturalisent, ne trouvent aucune réponse dans les statistiques traditionnelles. Quant à l'intégration professionnelle des migrants, elle ne peut être décrite qu'à une date donnée, mais son processus ne peut pas être étudié.

Dans ce contexte, un large projet portant sur les nouvelles formes de mobilité et de migration, le NCCR On the Move, financé par le Fonds national suisse pour la recherche scientifique, a vu le jour en 2014. Une partie de ce projet visait à mettre en place les outils statistiques, en l'occurrence une base de données longitudinales, visant à documenter la migration contemporaine et les caractéristiques de vie des migrants selon une approche longitudinale. Dans le cadre de la partie statistique de ce projet, un aspect prioritaire a été accordé à l'utilisation de données déjà existantes. Cette communication décrit les étapes marquant la constitution de cette base de données et commente les impacts de cette nouvelle statistique sur la connaissance de la migration contemporaine.

¹Philippe Wanner, NCCR On the Move et Institut de démographie et socioéconomie, Université de Genève, Pont d'Arve 40, 1211 Genève 4 (philippe.wanner@unige.ch). Ilka Steiner, Institut de démographie et socioéconomie, Université de Genève, Pont d'Arve 40, 1211 Genève 4.

2. Méthodes

2.1 Population sous étude

L'objectif est donc de constituer une base de données longitudinales, de nature sociodémographique, reposant sur des registres et statistiques déjà existants, en vue de décrire le séjour en Suisse de la population « migrante ». Cette population est définie selon le critère de la nationalité et non du lieu de naissance, pour deux raisons : d'une part, les pays germanophones ont de tout temps privilégié le critère du passeport, qui donne accès à certains droits (en particulier le droit de voter à l'échelle fédérale), à l'opposé des pays anglo-saxons qui privilégient généralement le lieu de naissance ; d'autre part, et cela résulte de la première raison, les statistiques se sont très longtemps focalisées sur le critère de la nationalité. Ainsi, de 1981 à 2010, un registre informatisé des étrangers (Registre central des étrangers) était disponible en Suisse, alors qu'aucun registre fédéral ne couvrait la population de nationalité suisse. Concernant la mesure de la population migrante au sens strict du terme (critère du lieu de naissance), aucune information n'était disponible, à l'exception des recensements décennaux: les personnes nées à l'étranger, mais ayant acquis la nationalité suisse, échappaient en effet à toute mesure.

Cependant, depuis fin décembre 2010 et l'instauration du registre harmonisé des habitants, la Suisse s'est dotée d'un enregistrement individuel de l'ensemble de sa population résidente, qu'elle soit Suisse ou étrangère, et quel que soit son statut migratoire. Ce registre des habitants fournit en effet la base de la statistique des personnes et des ménages, qui regroupe donc des informations pour chaque personne résidant de manière permanente ou non en Suisse.

La base de données longitudinale couvre la période 1998 à 2013, et pour cette raison, nous avons privilégié la nationalité comme indicateur (imparfait) du statut migratoire. Les personnes de nationalité étrangère résidant de manière permanente en Suisse sont incluses pour l'ensemble de la période. Les personnes de nationalité suisse, qui constituent un groupe de référence, sont incluses uniquement dès le 31 décembre 2010. Depuis la même date, il est également possible de distinguer les migrants des non-migrants (selon le critère du lieu de naissance).

Jusqu'en 2009, la base de données permet de suivre des personnes considérées d'une manière individuelle, sans possibilité d'identifier des ménages. Dès 2010, il est également possible de reconstituer les ménages (définis comme étant les personnes qui vivent dans le même logement) suite à l'entrée en vigueur du registre harmonisé des habitants, qui inclut un numéro de ménage. Cependant, les relations entre les divers membres du ménage ne sont pas connues : il n'est pas possible, par exemple, dans le cas où deux personnes vivent dans le même ménage, d'identifier s'il s'agit d'un couple ou de deux personnes sans lien conjugal.

2.2 Bases légales

Les travaux visant à constituer cette base de données l'ont été dans le cadre de la loi fédérale sur la statistique, qui rend possible les appariements de données à des fins de recherche. L'article 14a de cette loi prévoit en effet que « Pour exécuter ses tâches en matière de statistiques, l'office peut appairer des données à condition de les rendre anonymes. Si des données sensibles sont appariées ou si l'appariement de données permet d'établir des profils de la personnalité, les données appariées doivent être effacées une fois les travaux statistiques d'exploitation terminés ». Les données appariées peuvent être transmises aux chercheurs à condition qu'un contrat de protection soit conclu entre l'Office fédéral de la statistique (OFS) et le chercheur.

L'application par l'OFS de la loi sur la statistique est caractérisée par trois aspects : d'une part, les appariements de données se font dans les bureaux de l'office. Une place de travail a ainsi été mise à disposition en vue d'effectuer ces appariements in situ. D'autre part, les chercheurs n'ont pas accès à des informations non anonymes, ni à l'identifiant unique personnel (le numéro de sécurité sociale) qui permet certains appariements. Un pseudo-identifiant est produit par l'OFS à partir du numéro de sécurité sociale et mis à disposition des chercheurs. Enfin, chaque chercheur et groupe de recherche n'a accès qu'aux variables dont il a spécifiquement besoin pour sa recherche, avec un pseudo-identifiant spécifique à chaque projet de recherche.

2.3 Registres et statistiques inclus

Les registres inclus ont été sélectionnés en fonction de différents critères, parmi lesquels le caractère exhaustif des données (à une exception près). Un deuxième critère fait référence à l'information fournie par les registres, qui doit être en adéquation avec la problématique de recherche. A ce stade, on s'intéresse en particulier aux comportements migratoires, traduits par les changements de statut (arrivée, départ, changement de permis de séjour, naturalisation) et par les changements de lieux de domicile. En outre, différents travaux ont été initiés dans le domaine de l'intégration structurelle, mesurée en particulier par la position sur le marché du travail (incluant l'adéquation entre la profession exercée et la profession apprise). Cependant, il est envisagé d'inclure prochainement d'autres dimensions (par exemple en lien avec l'éducation ou les comportements démographiques) qui conduiraient à la prise en compte d'autres registres ou statistiques.

Le tableau 2.3-1 détaille les différents registres inclus. Trois d'entre eux (registre central des étrangers, registre des requérants d'asile et statistique des personnes et des ménages, qui repose essentiellement sur le registre harmonisé des habitants) fournissent l'information sur la population résidant en Suisse, et les quatre autres sur des caractéristiques économiques et professionnelles. Les registres des personnes sont structurés sous la forme de deux fichiers annuels, l'un comprenant l'effectif de la population à la fin de l'année, et l'autre les mouvements enregistrés dans la population au cours d'une année (par exemple une arrivée ou un décès). Une seule source n'est pas exhaustive (relevé structurel) et son intégration dans la base de données est liée au fait que cette enquête annuelle, qui remplace le recensement traditionnel exhaustif ayant eu lieu jusqu'en 2000, fournit des informations que l'on ne retrouve pas dans les autres registres, en particulier sur la profession apprise, la profession exercée, le nombre d'heures hebdomadaire d'activité, le type de ménage, la langue parlée et la langue principale, la religion, ou encore la pendularité et l'utilisation des transports publics ou privés.

Tableau 2.3-1
Registres inclus dans la base de données démographiques

	Caractéristiques	
	Population et période	Description
Registre central des étrangers	Personnes de nationalité étrangère, 1998 à 2010	Enregistrement exhaustif des étrangers titulaires d'un permis de séjour (ainsi que des travailleurs frontaliers) à la fin de l'année et des flux au cours d'une année (migrations, naissances, décès, changements de statut, etc.). Ce registre a été intégré en 2010 dans la statistique des personnes et des ménages.
Statistique des personnes et des ménages	Ensemble des personnes résidentes, 2010 à 2013	Enregistrement exhaustif des personnes résidentes en Suisse et des mouvements (migrations, changements de commune, naturalisation, etc.).
Registre des requérants d'asile	Requérants d'asile et admis provisoires, 1998 à 2010	Enregistrement exhaustif des personnes du domaine de l'asile à la fin de l'année et des flux au cours d'une année (migrations, naissances, décès, changements de statut, etc.). Ce registre a été intégré en 2010 dans la statistique des personnes et des ménages.
Relevé structurel	Echantillon de la population âgée de 15 ans et plus résidant en Suisse depuis plus de 12 mois, 2010 à 2013	Enquête portant sur les caractéristiques familiales, économiques, linguistiques et de mobilité d'un échantillon de plus de 200 000 personnes par année (Suisse et étrangères).
Comptes individuels	Personnes exerçant une activité rémunérée en Suisse, 1998 à 2013	Enregistrement exhaustifs des revenus professionnels soumis à cotisation de la prévoyance vieillesse. Ce registre exhaustif fournit le genre de cotisation (salariné, indépendant) et le montant du salaire.
Registre du chômage	Personnes inscrites au chômage, 1998 à 2013	Enregistrement exhaustif des personnes s'inscrivant pour une recherche d'emploi
Statistique de l'aide sociale	Bénéficiaires de prestations sociales sous conditions de ressources	Enregistrement exhaustif des bénéficiaires de l'aide sociale ²

² Actuellement, la statistique de l'aide sociale n'a été utilisée que pour un sous-échantillon de la population (requérants d'asile, réfugiés reconnus et personnes admises de manière provisoire en Suisse).

2.3 Travaux d'appariement

La mise en relation des différents registres a été effectuée par des appariements, lesquels reposent sur deux approches distinctes.

D'une part, la disponibilité d'un identifiant personnel dans les différents registres permet en effet de lier aisément plusieurs registres. Ce numéro d'identifiant de 13 chiffres (appelé AVS13) est totalement anonyme, et a été introduit à la fin de la première décennie du XIX^e siècle. Il remplace l'ancien numéro de sécurité sociale (AVS) qui était calculé selon un algorithme simple, à partir du nom de famille, du sexe de la personne et de la date de naissance, et ne garantissait donc pas l'anonymat. Le nouvel identifiant est disponible pour toute personne présente en Suisse en 2010 et après. Chaque étranger qui arrive en Suisse (immigration ou par naissance) se voit attribuer un numéro.

Les travaux d'appariement ont donc été accompagnés de la validation de cet identifiant, qui dans certains cas peut poser des problèmes. Ainsi, certaines personnes se sont vues attribuer de manière erronée deux numéros AVS13, et il a fallu tenir compte des mutations de ces numéros. Pour d'autres personnes, l'OFS ne disposait pas d'un numéro AVS13 au moment de la validation du registre harmonisé des habitants, et a attribué un numéro provisoire, qu'il a fallu remplacer ex post par le numéro effectif.

D'autre part, le registre des requérants d'asile (AUPER) nécessite, pour être apparié avec les autres registres, une approche différente. En effet, transitent dans ce registre les personnes en cours de procédure d'asile ou admises provisoirement en Suisse pour des raisons sécuritaires. Une fois un permis de séjour accordé (suite à la reconnaissance de leur statut de réfugié, mais parfois aussi suite à un mariage), ces personnes sont désactivées du registre AUPER pour réapparaître dans le registre central des étrangers. Or, il semble essentiel de disposer d'informations sur la procédure d'asile, en d'autres termes de relier les deux registres.

Cette liaison ne peut être effectuée, pour les personnes ayant obtenu un permis de séjour ou d'établissement avant 2010, que par un appariement reposant sur la disponibilité dans les deux registres, de variables non modifiables (par exemple date de naissance, sexe, nationalité à l'arrivée en Suisse) ou rarement modifiables (commune de domicile, état civil). Différentes tentatives ont été effectuées afin d'optimiser l'appariement, en appliquant des algorithmes existant dans la littérature (par ex. Fellegi et Sunter, 1969) ; finalement, compte tenu des spécificités des informations et de la qualité parfois médiocre des informations contenues dans le registre AUPER (la commune de domicile est souvent manquantes, les nationalités sont régulièrement corrigées etc.), un algorithme spécifique a été programmé. Celui-ci tient compte, en particulier, de deux caractéristiques: d'une part, les territoires nationaux se modifient, et un même individu peut voir sa nationalité se transformer avec le temps (par exemple, un Yougoslave devient de nationalité de Serbie-et-Monténégro, puis de nationalité serbe). On doit alors considérer des règles spécifiques d'appariement; d'autre part, certains enregistrements sont manquants (par exemple, une personne ayant reçu le statut de réfugié en 2005 peut apparaître tardivement – en 2008 – dans le registre central des étrangers): cette situation nécessite de faire des tentatives d'appariement en tenant compte d'un décalage temporel. Les procédures d'appariements ont été effectuées avec le logiciel SAS et sont documentées ailleurs (Wanner et al., 2016).

3. Résultats

3.1 Trajectoires

Les résultats obtenus se présentent sous la forme de différents fichiers annuels (plus d'une centaine) comprenant un pseudo-identifiant unique à l'ensemble des registres. Les fichiers peuvent représenter des états de la population (par exemple, 15 fichiers annuels du registre des étrangers, pour les 31 décembre des années 1997 à 2010) ou des mouvements (13 fichiers de mouvements du registre des étrangers, pour les années 1998 à 2010).

La reconstitution des trajectoires de vie s'effectue alors simplement en mettant ensemble ces différents fichiers, ainsi que les différentes variables qui intéressent le chercheur. Certaines trajectoires peuvent porter sur un épisode migratoire de courte durée (par exemple, une arrivée et un départ dans un court intervalle de temps), d'autres pour l'ensemble de la période étudiée.

Près de 10,64 millions d'individus ont été suivis sur une période comprise entre une année et 15 ans. Le nombre est plus élevé que la taille de la population suisse (8,2 millions résidents permanents en 2013) car la migration est caractérisée par un remplacement régulier de la population migrante, certains migrants partent tandis que d'autres migrants reviennent.

Au total, 3 431 527 personnes de nationalité étrangère ont été incluses au moins une fois dans le fichier "état" du registre des étrangers: 672 432 d'entre elles (soit 19,6%) ont été suivies durant l'ensemble de la période 1998-2013, soit enregistrées 17 fois à la fin de l'année entre le 31 décembre 1997 et le 31 décembre 2013 ; en revanche, 415 700 personnes n'ont été enregistrées qu'à une fin d'année (12,1%). En outre, quelque 7,2 millions d'individus de nationalité suisse ou arrivées en Suisse après 2010 figurent dans le registre harmonisé des habitants: quelque 84,6% d'entre elles ont été enregistrées sur l'ensemble de la période de 2010 à 2013. Le nombre de trajectoires ainsi reconstituées autorise donc une analyse fine des données.

3.2 Validation

Bien que considérées comme de plutôt bonne qualité, les sources utilisées souffrent de problèmes de qualité inhérents aux registres, et il est essentiel de valider les informations avant de les utiliser. On a été particulièrement attentifs aux différentes sources d'erreurs mentionnées par Roos et Nicol (1999), à savoir des duplications ou enregistrements manquants (par exemple disparition d'individus), des contradictions dans les informations (par exemple un individu dont la date de naissance est plus élevée que la date de l'immigration), des mises à jour tardives de dimensions pouvant se modifier dans le temps (par exemple l'état civil), et un manque de cohérence dans les différentes sources d'information (par exemple un individu de sexe masculin dans une source, et de sexe féminin dans une autre). Des cas ont été identifiés pour ces différentes sources d'erreurs et corrigés en tenant compte de critères précis.

La disponibilité d'informations portant à la fois sur l'état de la population (à la fin de chaque année) et les mouvements (naissances, décès, immigrations, émigrations, naturalisations, changement de statut, etc.) a également permis de valider les trajectoires reconstituées, et de corriger des cas de disparitions ou d'apparitions non documentées.

Les erreurs ayant été identifiées ont été traitées en tenant compte de règles de corrections précises et systématiques et documentées (voir Steiner et Wanner, 2015).

3.3 Analyses longitudinales initiées

Différentes utilisations de cette base de données démographiques ont déjà été effectuées et parfois soumises à publication dans des revues spécialisées. Ces analyses adoptent une approche longitudinale, puisque la base de données autorise un suivi des trajectoires migratoires.

Un premier axe de recherche a porté sur les facteurs intervenant sur l'émigration de la population migrante en Suisse, émigration qui peut être synonyme d'un retour dans le pays d'origine ou d'un départ dans un pays tiers (les deux cas pouvant être distingués grâce à une information disponible dans le registre harmonisé des habitants). Les résultats obtenus ont montré que des facteurs liés à une faible intégration structurelle (par exemple le fait de bénéficier d'un faible revenu du travail ou de ne pas exercer l'activité professionnelle correspondant aux compétences acquises lors de la formation) augmentent le risque d'une émigration de retour, alors que des facteurs traduisant une réussite professionnelle (par exemple un revenu plus élevé que celui anticipé en fonction de la formation acquise) accroissent la probabilité de partir dans un état tiers.

Un deuxième axe de recherches a porté sur les rythmes d'intégration de la population du domaine de l'asile (requérante d'asile ou réfugiée) selon une optique longitudinale et une durée d'observation allant jusqu'à 15 années (Wanner et al., 2016). Les appariements effectués dans ce cadre ont permis de mobiliser des variables mesurant le chômage, le recours à l'aide sociale, ou le niveau de revenus. Les analyses effectuées ont d'une part documenté des niveaux d'intégration extrêmement variables dans les communautés concernées par l'asile, et d'autre part montré la durée élevée de présence, nécessaire à l'amélioration toute relative du statut socioprofessionnel de la population du domaine de l'asile.

En conclusion, les appariements effectués et la préparation de cette base de données démographiques conduisent à une meilleure compréhension de différents phénomènes migratoires.

Bibliographie

Fellegi, I. P. Sunter, A.P, (1969). « A Theory for Record Linkage. » *Journal of the American Statistical Association*, 64(328): p. 1183-1210.

Roos, L., Nicol, P., Cageorge S. (1987). « Using administrative data for longitudinal research: comparisons with primary data collection », *Journal of Chronic Diseases*, 40(1), p. 41-49.

Steiner, I., Wanner, P. (2015), Towards a new data set for the analysis of migration and integration in Switzerland. Working Paper no 1. Neuchâtel: NCCR On the Move,

Wanner, P. (2014), *Une Suisse à dix millions d'habitants. Enjeux et débats*, Lausanne: PPUR.

Wanner, P., Bertrand, A.L., Steiner, I. (2016), *Intégration structurelle et déqualification de la population réfugiée en Suisse*, rapport non publié, Genève: Université de Genève.