

**Commission économique pour l'Europe**

Conférence des statisticiens européens

**Groupe d'experts des recensements
de la population et des habitations**

Vingt-quatrième réunion

Genève, 21-23 septembre 2022

Point 5 de l'ordre du jour provisoire

**Transitions dans les méthodes appliquées pour les recensements :
plans, expérience acquise et innovations****Évolution du recensement permanent de la population
en Italie : enseignements à retenir du premier cycle
et de la conception du recensement permanent
au-delà de 2021****Note de l'Institut national italien de statistique****Résumé*

Le recensement permanent de la population et des habitations (ci-après « le recensement permanent ») a été conçu conformément au programme de modernisation de l'Institut national italien de statistique (Istat), qui place le système intégré de registres statistiques au cœur de la production statistique. Dans ce système, les enquêtes sur le terrain ont pour rôle de faciliter la tenue des registres, c'est-à-dire, au sens large, d'en évaluer la qualité, ainsi que d'ajouter les informations manquantes, incomplètes ou de qualité insuffisante. Cela permet de disposer chaque année de statistiques détaillées en matière de recensement.

L'élément central du recensement permanent est le registre de base de la population, dont les principales sources sont les registres locaux de la population des municipalités italiennes. Deux enquêtes par sondage (enquête aréolaire et enquête par listes) sont menées chaque année pour évaluer et corriger les erreurs de couverture du registre de base de la population et collecter les données nécessaires à la production des résultats du recensement.

L'utilisation des données administratives a encore été accélérée en raison de la pandémie et de l'annulation consécutive des enquêtes sur le terrain concernant la vague de recensement de 2020. Afin de prédire les effectifs de population au niveau municipal par

* Établie par Antonella Bernardini, Angela Chieppa, Nicoletta Cibella, Gerardo Gallo, Fabrizio Solari et Donatella Zindato.

Note : Les appellations employées dans le présent document ne reflètent aucune prise de position du Secrétariat de l'Organisation des Nations Unies quant au statut juridique de pays, territoires, villes ou zones quelconques, ou de leurs autorités, ni quant au tracé de leurs frontières ou limites.



âge, sexe et citoyenneté, un processus intégrant les données disponibles provenant des vagues passées et les « signes de vie » administratifs a été mis en place pour établir les critères déterministes appliqués aux enregistrements individuels figurant dans le registre de base de la population. En raison de cette poussée obligatoire vers l'utilisation d'un plus grand nombre de données administratives, il a fallu repenser le cadre statistique nécessaire à l'évaluation de la qualité des processus d'estimation du recensement permanent. Plus généralement, il a fallu aussi repenser la conception du recensement permanent en se servant des données tirées d'enquête pour évaluer la qualité d'une estimation de la population entièrement fondée sur des registres.

I. Introduction

1. Le recensement permanent de la population et des habitations (ci-après « le recensement permanent ») a été conçu conformément au programme de modernisation de l'Institut national italien de statistique (Istat), qui place le système intégré de registres statistiques au cœur de la production statistique. Dans ce système, les enquêtes sur le terrain ont pour rôle de faciliter la tenue des registres, c'est-à-dire, au sens large, d'en évaluer la qualité, ainsi que d'ajouter les informations manquantes, incomplètes ou de qualité insuffisante. Cela permet de disposer chaque année de statistiques détaillées en matière de recensement.
2. Au cours des deux premières vagues (2018-2019), en raison des problèmes de qualité rencontrés dans le cadre des activités sur le terrain, la conception théorique a été partiellement modifiée pour inclure l'utilisation de l'enquête par listes (non prévue initialement pour estimer les erreurs de couverture) afin d'estimer les erreurs que représente la surcouverture et pour inclure l'utilisation des « signes de vie » administratifs afin de corriger la sous-couverture de l'enquête par listes.
3. En outre, en raison de la pandémie de COVID-19 et du retrait consécutif des enquêtes par sondage incluses dans la conception du recensement permanent, l'Istat a opté en 2020 pour l'utilisation exclusive des signes de vie en vue d'estimer les erreurs de couverture du registre de la population.
4. En raison de cette poussée obligatoire vers l'utilisation d'un plus grand nombre de données administratives, il a fallu repenser le cadre statistique nécessaire à l'évaluation de la qualité des processus d'estimation du recensement permanent. Plus généralement, il a fallu aussi repenser la conception du recensement permanent en se servant des données tirées d'enquête pour évaluer la qualité d'une estimation de la population entièrement fondée sur des registres.

II. Première conception du recensement permanent de la population

5. L'élément central du recensement permanent est le registre de base de la population, dont les principales sources sont les registres locaux de la population des municipalités italiennes. Avec le registre statistique de base des adresses et les registres thématiques relatifs à l'éducation et à l'emploi, le registre de la population constitue la base de la production de données démographiques. Deux enquêtes par sondage (enquête aréolaire et enquête par listes) sont menées chaque année pour évaluer et corriger les erreurs de couverture du registre de base de la population et pour collecter des données pour les variables qui ne sont pas disponibles (ou ne le sont que partiellement) dans les registres.
6. Le modèle de capture-recapture a été adopté pour les estimations directes des erreurs de couverture du registre de base de la population : les données présentes dans le registre de la population constituent la « première capture » tandis que les données recueillies sur le terrain représentent la « deuxième capture ». À la différence d'une enquête postcensitaire type qui est destinée à mesurer la sous-couverture, dans la conception du recensement permanent, la deuxième « capture » vise à mesurer et à corriger à la fois la sous-couverture et la surcouverture du registre de base de la population.
7. Plus précisément, l'enquête aréolaire a été utilisée pour estimer le nombre d'individus résidant habituellement dans la municipalité qui ne sont pas inclus dans le registre de base de la population (sous-couverture du registre) tandis que l'enquête par listes a servi à estimer le nombre d'individus inclus dans le registre qui ne résident plus habituellement dans la municipalité (surcouverture du registre). De plus, les signes de vie provenant de la base de données intégrée des résidents habituels et classés selon la durée, le type et la fiabilité de la source ont été intégrés dans le processus d'estimation afin de corriger la sous-couverture de l'enquête. Les individus figurant dans le registre de base de la population qui n'avaient pas été dénombrés ont donc été considérés comme des résidents habituels s'ils étaient associés à de forts signes de vie (c'est-à-dire au moins huit mois).

8. Une estimation indirecte à l'aide de modèles liés à de petites zones a ensuite été utilisée à la fois afin d'améliorer la qualité des estimations directes pour les municipalités échantillonnées et afin de calculer les estimations pour les municipalités non échantillonnées. À l'issue de ce processus, on a finalement dénombré la population en appliquant des coefficients de correction pour les erreurs en matière de sous-couverture et de surcouverture aux individus inscrits dans le registre de base de la population (voir Gallo et Zindato, 2018).

9. Outre les problèmes posés par les activités sur le terrain déjà mentionnés, d'autres difficultés liées à la conception du premier cycle concernaient la taille de l'échantillon, ce qui ne permettait pas d'estimer les coefficients de pondération des corrections au niveau de désagrégation souhaité, ainsi que la sous-couverture de l'enquête aréolaire en raison de la qualité insuffisante de la base d'échantillonnage des adresses.

III. Utilisation de données administratives pour le dénombrement de la population de résidents habituels en 2020

10. L'utilisation des données administratives dans le cadre du recensement permanent a encore été accélérée en raison de la pandémie et de l'annulation consécutive des enquêtes sur le terrain concernant la vague de recensement de 2020. Afin de dénombrer la population en 2020, l'Istat a opté pour l'utilisation exclusive des signes de vie en vue d'estimer les erreurs de couverture du registre de la population.

11. Pour ce faire, il a été recouru aux critères de classification appliqués aux registres statistiques. Plus précisément, un processus intégrant les données disponibles provenant des vagues précédentes et les signes de vie a été mis en place dans le but d'établir des critères déterministes appliqués aux enregistrements individuels dans les registres statistiques et ainsi d'estimer la population résidant habituellement en Italie. Ces critères ont été établis sur la base de connaissances spécialisées et d'un processus structuré et itératif, dans lequel une partie des variables est définie continuellement au fur et à mesure du traitement des signaux administratifs et de l'analyse des données (Chieppa *et al.*, 2018).

12. Pour l'établissement de la base de données intégrée des résidents habituels, les sources pertinentes pour le recensement de la population de résidents habituels sont choisies et ordonnées hiérarchiquement au sein du système intégré de microdonnées qui, conformément aux dispositions de la loi relative à la confidentialité, collecte des informations provenant de sources administratives pour appuyer les processus de production statistique, tant dans le domaine social que dans le domaine économique. Le processus d'intégration de la base de données intégrée des résidents habituels comprend le traitement de données provenant de plus de quarante archives administratives, chacune contenant des informations de base sur les signes de vie (événements) des individus et couvrant plusieurs années. Parmi les principales sources figurent les registres thématiques des professions et de l'éducation, le registre des déclarations fiscales et les archives de la sécurité sociale, ainsi que le registre des biens immobiliers. Pour chaque événement administratif, des informations sur la localisation de l'événement (au moyen des codes de la province et de la municipalité) sont également enregistrées.

13. La base de données intégrée des résidents habituels utilisée pour le dénombrement de la population de 2020 intègre des microdonnées, obtenues du 1^{er} janvier 2019 au 31 décembre 2020. L'observation longitudinale des signaux directs sur deux ans permet de capturer des profils bien précis concernant la présence des individus sur le territoire. Dans certains cas, ces profils concernant la présence en Italie identifient clairement les résidents habituels, tandis que dans d'autres cas, les signes de vie sont de faible intensité ou identifient des travailleurs saisonniers (dans les deux cas, les profils ne peuvent être associés à la définition d'un résident habituel). Les signes se rapportant au travail et aux études, ainsi que les baux d'habitation ou les prestations sociales du système de sécurité nationale, peuvent être classés comme des signes de vie directs pour ce qui est de la résidence habituelle en Italie ; ces enregistrements offrent des détails considérables, à savoir la durée de l'activité, sa localisation (municipalité et adresse) et certains attributs particuliers (contrat de travail, établissement scolaire fréquenté/cours suivis, etc.) pertinents pour évaluer la force du signe de vie (Istat, 2021). Des exemples de signes de vie indirects sont plutôt les registres des

impôts sur le revenu (déclarations fiscales, déclarations d'impôt sur le revenu, etc.) ainsi que la possession d'un véhicule automobile selon le registre public des véhicules automobiles ou la possession d'une propriété selon les archives cadastrales, tous ces éléments fournissant des signes indirects de la résidence habituelle en Italie.

14. Après le traitement des signes de vie directs et la détermination de la municipalité où les études ou l'activité professionnelle ont eu lieu, l'étape suivante du processus consiste à intégrer les individus présentant des signes de vie directs au registre de base de la population. Plus précisément, cette étape permet de recenser pour chaque municipalité, d'une part, tous les individus ayant leur résidence habituelle en Italie (c'est-à-dire ayant un signe de vie direct) qui ne sont pas enregistrés dans le registre de base de la population et, d'autre part, les individus inscrits dans ce registre mais sans signe de vie direct.

15. L'étape suivante consiste à comparer le signe de vie indirect (provenant du registre des impôts) relatif aux « membres de la famille à charge » et aux propriétaires d'un véhicule automobile ou d'une unité immobilière aux individus enregistrés dans le registre de base de la population qui se sont avérés ne pas présenter de signe de vie lors des étapes précédentes.

16. Enfin, la dernière étape consiste à identifier les individus n'ayant ni signe de vie direct ni signe de vie indirect, c'est-à-dire à déterminer la surcouverture du registre de base de la population. Pour ce sous-groupe de population, cependant, une vérification supplémentaire est effectuée afin d'identifier les « conjoints » des personnes de référence qui présentent un signe de vie direct. Ces conjoints de personnes ayant un signe de vie direct, qui autrement se retrouveraient dans l'ensemble des personnes classées comme faisant l'objet d'une surcouverture dans le registre de base de la population parce qu'elles n'ont pas de signe de vie direct ou indirect, sont au contraire considérés comme des résidents habituels.

17. On trouvera dans le tableau 1 une comparaison entre les individus présentant un signe de vie selon la base de données intégrée des résidents habituels et les individus enregistrés comme résidents habituels dans le registre de base de la population. La base de données susmentionnée recense près de 62 millions d'individus présentant un signe de vie administratif, mais parmi eux, seuls 59,2 millions peuvent être considérés comme résidant habituellement en Italie. Par ailleurs, la population correctement inscrite dans le registre de base de la population s'élève à 58,7 millions. La sous-couverture nationale du registre de base de la population représente près de 325 000 individus, tandis que la surcouverture des registres de population représente un peu plus d'un million.

Tableau 1

Population dénombrée et population totale au 31 décembre 2020 suite à l'intégration entre le registre de base de la population et la base de données intégrée des résidents habituels

<i>Description des résultats</i>	<i>Type de registre ou d'archives</i>	<i>Population dénombrée totale</i>	<i>Dénombrement du recensement de la population</i>
Population correctement placée dans le registre de base de la population	Registre de base de la population et base de données intégrée des résidents habituels	58 713 660	Oui
Sous-couverture au niveau national	Uniquement base de données intégrée des résidents habituels	324 932	Oui
Unités incertaines	Uniquement registre de base de la population	197 621	Oui
Sous-couverture au niveau national	Uniquement registre de base de la population	1 005 908	Non
Unités incertaines	Uniquement base de données intégrée des résidents habituels	288 211	Non
Population non incluse dans le dénombrement	Uniquement base de données intégrée des résidents habituels avec des signes de vie inutiles	1 410 497	Non

<i>Description des résultats</i>	<i>Type de registre ou d'archives</i>	<i>Population dénombrée totale</i>	<i>Dénombrement du recensement de la population</i>
Population dénombrée lors du recensement de 2020	Base de données intégrée des résidents habituels + registre de base de la population	59 236 231	Oui
Population totale	Base de données intégrée des résidents habituels + registre de base de la population	61 961 252	

Source : Istat, 2022.

18. En outre, comme le montre le tableau, le registre de base de la population compte près de 200 000 individus dont le statut de résident habituel ne peut être clairement établi sur la base des signes de vie (signes de vie faibles ou inexistants) et pour lesquels un choix prudent a été opéré, c'est-à-dire qu'ils ont été inclus dans le dénombrement final de la population. En revanche, les personnes dont les signes de vie sont incertains (plus de 1,4 million de personnes dont les signes administratifs sont faibles ou mal localisés) et qui n'étaient pas des résidents habituels selon le registre de base de la population n'ont pas été prises en compte dans le dénombrement censitaire. Enfin, il faut mentionner que l'erreur consistant en une mauvaise localisation dans le registre de la population (individus enregistrés dans une municipalité mais qui résident habituellement dans une autre) n'a pas été évaluée jusqu'à présent.

19. Comme en 2018 et 2019, l'objectif du dénombrement de 2020 était de produire des estimations de la taille de la population en corrigeant les erreurs de couverture du registre de base de la population. On a atteint cet objectif en identifiant, d'une part, les individus inscrits dans le registre de base de la population comme résidents habituels mais qui ne figurent pas dans les autres sources administratives (surcouverture du registre de base de la population) et, d'autre part, les individus figurant dans les données administratives comme résidents habituels mais non inscrits dans le registre de base de la population (sous-couverture de celui-ci). Ce processus pourrait également être décrit comme l'identification des personnes faisant l'objet d'une surcouverture dans la base de données du registre de base de la population intégrée à la base de données intégrée des résidents habituels compte tenu des signes de vie.

20. Il s'agissait en effet d'une innovation importante, qui assurait une correspondance entre le dénombrement censitaire et les enregistrements individuels, ce processus étant distinct de celui de 2018 et de 2019, dans lequel on avait corrigé les erreurs de couverture en appliquant des coefficients de pondération aux enregistrements du registre de base de la population (Istat, 2020 ; 2021).

21. L'année 2020 représente le point de départ d'une utilisation plus intensive des données administratives aux fins de l'estimation de la taille de la population. L'Istat s'efforce d'améliorer l'utilisation des signes de vie dans le nouveau cycle (postérieur à 2021) du recensement permanent de la population et des habitations. Un tournant sera marqué avec l'acquisition de nouvelles sources (par exemple, les archives des services publics concernant notamment les données sur la consommation d'énergie obtenues grâce aux compteurs intelligents). Ces sources fourniront très probablement des éléments d'évaluation objectifs se rapportant au lieu réel de résidence habituelle (Albert et Rajagopal, 2013).

22. Dans le deuxième cycle du recensement permanent qui débutera en 2022, les enquêtes conserveront un rôle crucial pour ce qui est de collecter des données sur les variables non remplaçables (ou seulement partiellement remplaçables) par des données administratives (enquête par listes) et pour ce qui est de fournir des mesures de la qualité concernant l'estimation de la taille de la population entièrement fondée sur des registres. Une formalisation du processus d'estimation de la taille de la population entièrement fondé sur des données administratives est à l'étude, y compris le choix des mesures de la qualité et la définition de l'enquête (enquête de contrôle) appuyant le nouveau processus d'estimation.

VI. Recensement permanent de la population au-delà de 2021

23. Pour la vague 2021 du recensement permanent, l'objectif principal est de produire des dénombrements de la population entièrement fondés sur des signes de vie obtenus à partir des données administratives. De nouvelles techniques sont à l'étude. Elles consistent à utiliser des données tirées d'enquête pour améliorer les profils des signes de vie (modèles des classes latentes et autres méthodes des sciences des données) liés à des groupes de population difficilement identifiables par les seuls critères déterministes. En même temps, on utilisera des données tirées d'enquête pour produire des estimations de l'erreur caractérisant ces dénombrements de la population.

24. Pour le deuxième cycle du recensement permanent, qui débutera en 2022, l'architecture proposée s'appuie sur la définition d'un registre étendu de la population résultant d'un processus d'intégration entre le registre de base de la population et les archives administratives contenant des signes de vie. En Italie, le registre étendu de la population entrera en vigueur en tant que processus d'intégration entre le registre de base de la population et la base de données intégrée des résidents habituels.

25. Il est supposé que le registre étendu de la population ne peut pâtir que d'une surcouverture (c'est-à-dire qu'il n'est pas touché par une sous-couverture). Les profils des signes de vie sont ensuite définis pour recenser les sous-populations dont les individus sont supposés avoir un comportement similaire en matière de surcouverture, et chaque individu du registre étendu de la population est affecté à un profil particulier de signe de vie h , $h = 1, \dots, H$. Les profils des signes de vie peuvent être définis au moyen soit des informations fournies par des experts, soit des preuves résultant de modèles statistiques, soit d'une combinaison des deux.

26. À partir de 2023, un autre objectif sera de produire une liste des personnes faisant l'objet d'une surcouverture dans le registre étendu de la population. À cette fin, une fonction d'indicateur fondée sur le profil des signes de vie est définie (Bernardini, Cibella et Solari, 2022), selon laquelle tous les individus de chaque profil sont classés de manière soit à être inclus dans le dénombrement de la population, soit à en être exclus. Ce choix donne lieu à des estimations fondées sur des registres qui sont biaisées et peut être considéré comme une approche de la dichotomisation du dénombrement proposée par Zhang (2019), qui prédit plutôt des probabilités de profils de surcouverture dans les valeurs $[0,1]$.

27. De plus, outre l'erreur que représente une surcouverture, un registre étendu de la population peut pâtir d'une mauvaise localisation s'agissant des zones locales, c'est-à-dire que des individus peuvent être correctement inclus dans le registre susmentionné mais affectés à la mauvaise zone de localisation. Pour évaluer l'erreur que constitue une mauvaise localisation, il convient d'utiliser une version améliorée du registre étendu de la population, dans laquelle les enregistrements se réfèrent à des paires « individu-adresse » au lieu des seuls individus (voir Zhang, 2021c et Bernardini, Cibella et Solari, 2022). Cela éviterait d'ajouter des composantes relatives aux sélections biaisées dans les estimations des enquêtes et, par conséquent, dans les mesures de la qualité associées aux estimations de la taille de la population fondées sur des registres (voir Bernardini, Cibella et Solari, 2022).

28. On pourrait évaluer les mesures de la qualité concernant l'estimation de la taille de la population en comparant les estimations fondées sur des registres et les estimations des enquêtes (pour plus de détails, voir Zhang, 2022) ou en suivant l'approche de la couverture de l'évaluation proposée par Zhang (2021b), selon laquelle la couverture de l'intervalle de confiance est choisie comme mesure de précision. Suivant cette approche, connue sous le nom d'approche de l'enquête de contrôle, il faut recourir à des enquêtes pour évaluer les mesures de la qualité au lieu de produire des estimations. En ce qui concerne le recensement permanent, une enquête de contrôle est prévue à partir de 2023.

29. Selon l'approche de l'enquête de contrôle, les inefficacités de l'enquête telles que celles rencontrées en 2018 et 2019 n'ont d'effet que sur l'estimation des mesures de la qualité et non sur l'estimation de la taille de la population.

A. Définition des profils de signes de vie et d'une fonction d'indicateur fondée sur les profils de signes de vie

30. Ainsi qu'il est mentionné plus haut, on définit les profils de signes de vie du registre étendu de la population pour déterminer les sous-populations dont les individus sont censés avoir un comportement similaire en matière de surcouverture. À cette fin, un point crucial est le choix de la méthode visant à définir les profils en question. Les règles de classification peuvent être établies à l'aide de l'approche de l'apprentissage automatique (Michalski, 1983) ou de méthodes fondées sur le score de propension (voir par exemple Stuart, 2010). Une autre approche possible consiste à explorer les informations contenues dans les archives administratives selon un processus de découverte de connaissances dans les bases de données (voir Chieppa *et al.*, 2019).

31. Des modèles statistiques peuvent être utilisés pour prédire les taux de surcouverture concernant chaque profil θ_h . À cette fin, les meilleures options semblent être les modèles log-linéaires ajustés et les modèles de classes latentes.

32. Au lieu de définir la fonction d'indicateur fondée sur les signes de vie dans chaque profil comme étant une valeur constante égale à 0 ou 1, des valeurs aléatoires 0 ou 1 peuvent être générées à partir de certaines probabilités prédites spécifiées θ_h . Cela permettrait d'obtenir des estimations non biaisées fondées sur des registres (modèles).

33. Pour obtenir une liste stable d'individus faisant l'objet d'une surcouverture dans le temps, on peut attribuer des numéros aléatoires permanents à chaque individu au lieu de générer des numéros aléatoires indépendants chaque année.

B. Principales caractéristiques de l'enquête de contrôle

34. Ainsi qu'il a été mentionné précédemment, une enquête de contrôle sera menée en 2023 pour fournir des mesures de la qualité concernant l'estimation de la taille de la population fondée sur des registres.

35. À cette fin, le modèle d'échantillonnage le plus adéquat consiste à échantillonner des individus à partir du registre étendu de la population ou des paires d'individus et d'adresses à partir de la version améliorée de ce registre. L'approche de l'échantillonnage permet de réduire les coûts par rapport aux enquêtes aéroliques.

36. En outre, la taille globale de l'échantillon peut être plus petite que celle d'une enquête fondée sur une base d'échantillonnage traditionnelle visant à estimer la taille de la population.

37. Pour certains ensembles d'individus, il peut y avoir des informations incomplètes ou peu fiables concernant l'adresse de leur résidence habituelle potentielle. Dans ce cas, d'autres techniques d'échantillonnage doivent être utilisées. Les solutions possibles sont l'échantillonnage aérolique (si les individus sont censés être concentrés dans des parties particulières du territoire) ou l'échantillonnage indirect (on utilise par exemple l'adresse de leur lieu de travail ou d'étude (signes de vie) pour essayer de collecter des informations utiles sur leur véritable adresse).

38. Contrairement aux enquêtes traditionnelles visant à estimer un ou plusieurs paramètres démographiques, la définition de l'attribution des échantillons dans une enquête de contrôle n'est pas simple et intuitive. Deux propositions fondées sur l'attribution de puissance définie par Bankier (1988) sont présentées dans Bernardini, Cibella et Solari (2022). Selon le concept de base, les profils de signes de vie pour lesquels le taux de surcouverture n'est pas proche de 0 ou de 1 devraient être divisés en K_h profils h_k , $k = 1, \dots, K_h$, pour lesquels le taux de surcouverture des valeurs correspondantes θ_{h_k} est plus proche de 0 ou de 1 que de θ_h . Pour atteindre cet objectif, des informations auxiliaires supplémentaires sont nécessaires et une solution possible pour collecter les informations requises est le suréchantillonnage des profils critiques.

39. Enfin, une enquête aérolique à petite échelle pourrait être réalisée afin d'évaluer une éventuelle sous-couverture du registre étendu de la population. Une autre façon de mesurer la sous-couverture consiste à introduire dans l'enquête de contrôle une composante

« vérification inverse des enregistrements » (c'est-à-dire une vérification croisée des enregistrements relatifs aux non-répondants). Bien que l'intégration des différentes enquêtes puisse poser des problèmes en matière d'organisation et de travail de terrain, d'un point de vue purement statistique, elles peuvent être reliées à l'aide d'un échantillonnage fondé sur des graphiques (voir Zhang, 2021a).

V. Conclusions

40. Pour remplacer le recensement décennal, l'Istat a lancé en 2018 le recensement permanent de la population et des habitations, qui s'appuie sur l'intégration de données administratives à des informations collectées à partir de deux enquêtes par sondage (enquête aréolaire et enquête par listes) menées annuellement dans des municipalités autoreprésentatives, ainsi que tous les quatre ans, selon un schéma de rotation, dans les municipalités non autoreprésentatives (Falorsi, 2017). Selon cette conception, les données issues des registres sont complétées par des données obtenues sur le terrain.

41. L'utilisation des données administratives a été encore accélérée en raison de la pandémie et de l'annulation consécutive des enquêtes sur le terrain concernant la vague de recensement de 2020.

42. En raison de cette poussée obligatoire vers une plus large utilisation de données administratives, il a fallu repenser à la fois le cadre statistique nécessaire à l'évaluation de la qualité et les processus d'estimation globaux du recensement permanent, en se servant des données tirées d'enquête pour mesurer la qualité d'une estimation de la population entièrement fondée sur des registres.

43. À cette fin, le traitement des données du recensement de 2021, actuellement en cours, sera d'une grande importance, étant donné que des données tirées des enquêtes et des données administratives seront disponibles. Les comparaisons entre les différents modèles d'estimation, l'intégration des données administratives et des données tirées des enquêtes ainsi que l'évaluation de la qualité du travail sur le terrain sont autant de domaines de recherche importants permettant d'améliorer la conception des futurs cycles de recensement permanent de la population et des habitations.

Bibliographie

- Albert, A. & Rajagopal, R. (2013), Smart Meter Drive Segmentation: What Your Consumption Says About You. *IEEE Transactions on power systems*: 4019–4030.
- Bankier, M. D. (1988), Power allocations: determining sample sizes for subnational areas, *The American Statistician*, 42:3, 174–177.
- Bernardini A., Cibella N. & Solari F. (2022). A Statistical Framework for Register Based Population Size Estimation, Technical Report, Istat Advisory Committee on Statistical Methods 2022 Springtime Meeting, Roma, 19–20 May 2022.
- Chieppa A., Gallo G., Tomeo V., Borrelli F. & Di Domenico S. (2019). Knowledge discovery for inferring the usually resident population from administrative registers, *Mathematical Population Studies Mathematical Demography*, 26:2, 92–116. <https://doi.org/10.1080/08898480.2017.1418114>.
- Falorsi, S. (2017), The Italian experience on the Population and Housing Census: the Master Sample, UNECE Meeting, October 4–6 (2017), https://unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.41/2017/Meeting-Geneva-Oct/Day2_1130_Italy_falorsi_presentation.ppt__1_.pdf.
- Gallo, G. & Zindato, D. (2018). Annex H. Italy case study, in UNECE, Guidelines on the Use of Registers and Administrative Data for Population and Housing Censuses, Geneva, <https://unece.org/guidelines-use-registers-and-administrative-data-population-and-housing-censuses-0>.

- Gallo, G. & Zindato, D. (2021). Italy: The combined use of survey and register data for the Italian Permanent Population Census count in UNECE, Guidelines for Assessing the Quality of Administrative Sources for Use in Censuses (endorsed by the 69th plenary session of the Conference of European Statisticians), <https://unece.org/statistics/publications/CensusAdminQuality>.
- Istat (2020). Nota tecnica sulla produzione dei dati del Censimento Permanente: la stima della popolazione residente per sesso, età cittadinanza, grado di istruzione e condizione professionale per gli anni 2018 e 2019, <https://www.istat.it/it/files/2020/12/NOTA-TECNICA-CENSIPOP.pdf>.
- Istat (2020). Nota tecnica sulla produzione dei dati del Censimento Permanente: la popolazione residente per genere, età, cittadinanza e grado di istruzione al 31.12.2020, https://www.istat.it/it/files/2021/12/NOTA-TECNICA-CENSIMENTO-POPOLAZIONE_2020.pdf.
- Zhang, L.-C. (2019). On provision of UK neighbourhood population statistics beyond 2021, Report for ONS, <https://arxiv.org/pdf/2111.03100.pdf>.
- Zhang, L.-C. (2021a). Graph Sampling, Chapman and Hall/CRC.
- Zhang, L.-C. (2021b). Proxy expenditure weights for Consumer Price Index: audit sampling inference for big-data statistics. *Journal of the Royal Statistical Society, Series A*, 184:2, 571–588.
- Zhang, L.-C. (2021c). Discussion at ISI session ‘Population statistics using administrative data instead of census’, Virtual 63rd ISI World Statistics Congress, 11–16 July 2021.
- Zhang, L.-C. (2022b). Complementarities of survey and population registers, in N. Balakrishnan, T. Colton, B. Everitt, W. Piegorsch, F. Ruggeri and J. L. Teugels (eds.), Wiley StatsRef: Statistics Reference.
-