

**Conseil économique et social**

Distr. générale
19 août 2009
Français
Original: anglais

Commission économique pour l'Europe**Conférence des statisticiens européens****Groupe d'experts des recensements de la population et des habitations****Douzième réunion**

Genève, 28-30 octobre 2009

Point 4 de l'ordre du jour provisoire

Qualité des recensements et contrôle de la divulgation**Évaluation de la qualité du recensement intégré
de 2008 – Israël****Note du Bureau central de statistique d'Israël****I. Introduction**

1. Le recensement et ses résultats ont été évalués au regard de plusieurs aspects de l'évaluation de la qualité, à savoir couverture, exactitude, cohérence, complétude, fiabilité et validité^{1, 2}. Dans la publication de la Commission économique des Nations Unies pour l'Europe intitulée «Recommandations de la Conférence des statisticiens européens pour les recensements de la population et des habitations de 2010» sont énumérés certains des aspects de la qualité que doivent présenter les trois phases du recensement (conception, mise en œuvre et évaluation)³ et qui sont l'exactitude, l'actualité, l'accessibilité, l'interprétabilité et la cohérence. Israël a procédé à un recensement intégré en 2008 non assorti d'une phase postcensitaire puisqu'il était supposé que ce recensement et ses deux échantillons fourniraient les informations nécessaires à l'évaluation, qui a été facilitée par le fait qu'il s'agissait d'un recensement intégré fondé à la fois sur des sources administratives et sur des opérations de terrain. La collecte de données a été entièrement informatisée à l'aide d'ordinateurs portables qui ont permis de réunir et de communiquer chaque jour des renseignements à une base de données centrale. Il a ainsi été possible de suivre les

¹ Willem de Vries (2002), Dimensions of Statistical Quality, Inter-agency meeting on coordination of statistical activities, SA/2002/6/Add.1.

² Leo, Pipnio, Yang W. Lee and Ricahrd Y., Wang (2002), Data Quality Assessment, Communication of the ACM April 2002/vol. 45, no, 4ve.

³ Recommandations de la Conférence des statisticiens européens pour les recensements de la population et des habitations de 2010, Commission économique des Nations Unies pour l'Europe, 2006.

opérations sur le terrain au moyen d'outils spéciaux à la pointe de la technologie. L'évaluation sera achevée à la fin de 2009, mais la plupart des travaux y afférents sont terminés.

II. Description succincte des méthodes et procédures appliquées dans le recensement intégré

2. Le recensement intégré s'appuie pour beaucoup sur des sources administratives, principalement le registre de la population, et d'autres sources de même type, qui permettent d'actualiser les informations concernant les personnes «inactives» et d'établir des «familles administratives». La qualité des renseignements dépend des besoins, des procédures, des règlements et des lois en vertu desquels les données administratives sont réunies et tenues à jour. Ces dernières ne sont pas forcément conformes aux définitions et besoins du recensement. Deux enquêtes par sondage ont été conçues de manière à remédier aux lacunes des sources administratives, leurs principaux objectifs étant les suivants:

a) Estimer la qualité des principales sources administratives et ajuster en conséquence les estimations démographiques visant les «résidents habituels» et leur répartition géographique;

b) Collecter des données sur les caractéristiques socioéconomiques et démographiques de la population et des ménages dans les «zones statistiques» et les petites localités. Le registre de la population est l'élément fondamental du recensement intégré⁴. On a eu recours à deux bases d'échantillonnage indépendantes. La première, qui avait pour but d'estimer le sous-dénombrement inhérent au registre, était un échantillon établi à partir d'une base de données géographique créée spécialement pour le recensement, qui divisait le pays en 40 000 cellules couvrant la totalité du pays et représentant chacune une moyenne de 50 ménages (la fourchette étant de 30 à 70 ménages). Le deuxième échantillon, qui visait à estimer le surdénombrement, était établi à partir du registre lui-même. La population des institutions a été entièrement dénombrée selon la procédure de recensement classique.

III. Opérations sur le terrain

3. Un questionnaire informatisé a été soumis à l'échantillon géographique à l'aide d'un ordinateur portable (système d'entretien individuel assisté par ordinateur) tandis que l'échantillon établi d'après le registre de la population a été interrogé par téléphone (système d'entretien téléphonique assisté par ordinateur). Pour le premier type d'enquête, le pays a été divisé en 12 régions. Le travail sur le terrain a duré trois mois, le premier étant consacré à la formation et à l'initiation, ainsi qu'au marquage des bâtiments et logements échantillonnés. La phase effective de collecte de données a duré sept semaines. La «date du recensement» était le 27 décembre 2008, la campagne de collecte ayant débuté le 28 décembre. Les enquêteurs ont été recrutés pour une période de trois mois (de décembre 2008 à février 2009) et ont perçu des rémunérations différentes selon la qualité des données recueillies et la difficulté de la tâche. La plus faible rémunération correspondait au salaire minimum national.

⁴ http://www.cbs.gov.il/www/mifkad/census2008_e.pdf.

IV. Registre amélioré de la population

4. Le registre de la population contient des données sur les personnes qui ont été, à un moment donné, des nationaux ou des «résidents permanents» depuis qu'il a été institué en novembre 1948, peu après la création de l'État d'Israël. Les personnes qui, d'une manière définitive, ne résident plus en permanence dans le pays se voient attribuer le statut juridique d'«inactif» (qui s'applique principalement aux personnes décédées ou qui ont officiellement informé le Ministère de l'intérieur qu'ils vivaient dans un autre pays). Les personnes très âgées (plus de 110 ans) et dont les données n'ont pas été actualisées dans le registre depuis longtemps sont marquées comme étant «inactives», c'est-à-dire qu'ils ne figurent probablement pas au nombre de la population. Les personnes qui sont suspectées d'être absentes du pays depuis longtemps et de façon continue et dont les données n'ont pas été actualisées dans le registre pendant la période écoulée depuis leur dernier départ constaté sont marquées comme étant «potentiellement inactives», leur «adresse résidentielle» étant supprimée du registre. Les indications susmentionnées sont attribuées administrativement par le Ministre de l'intérieur. Tous les types de résidents «inactifs» sont supprimés des «listes électorales» aux fins des élections parlementaires et municipales. Les sources administratives d'appui incorporées dans le registre amélioré étaient les suivantes:

a) Fichiers d'assurances nationaux (y compris l'assurance maladie nationale) qui indiquent si telle ou telle personne a droit à l'assurance maladie et aux prestations sociales nationales comme les allocations pour enfants, les allocations pour handicapés, etc.;

b) Liste des compteurs électriques;

c) Registre d'immatriculation des véhicules.

5. Les fichiers d'appui ont servi à actualiser les adresses et à créer des «familles administratives». La qualité du registre amélioré et surtout celle des familles administratives est encore en cours d'évaluation.

V. Qualité de l'infrastructure géographique

6. Trois processus précensitaires ont permis d'évaluer la couverture géographique et de procéder à des actualisations:

a) De meilleurs outils d'orientation ont été mis au point pour les localités ne comptant pas d'adresses enregistrées (en général des villages arabes sans nom de rue), douze mois avant la phase de dénombrement, et des points de repère ont été ajoutés à la couche informatisée des «bâtiments et rues»;

b) Une autre évaluation de la couverture des bâtiments a été réalisée un mois avant le dénombrement. Des points de repère ont été ajoutés à la main sur les cartes imprimées par des coordinateurs géographiques recrutés spécialement pour faciliter l'actualisation et l'orientation. Dans 50 localités, 3 694 points de repère ont été ajoutés à la couche informatisée et environ 900, ajoutés à la main sur les cartes papier, sont actuellement incorporés dans la base de données géographique numérique;

c) Il a été nécessaire d'actualiser la couche des bâtiments pour les nouveaux bâtiments situés dans les cellules échantillonnées. L'échantillon a été établi à partir d'une base de données géographique numérique, qui a été gelée quinze mois avant le dénombrement sur le terrain. Les données informatisées se rapportaient à la base de données de l'année précédente. De nouveaux bâtiments qui ne figuraient pas dans la base de données au moment de l'échantillonnage ont été ajoutés pendant la phase précensitaire lorsque les enquêteurs ont établi la liste de tous les logements situés dans les bâtiments

échantillonnés et ajouté les bâtiments qui étaient absents des cartes papier – 9 032 bâtiments (5 %) ont été ajoutés aux cellules échantillonnées (au total 181 001 bâtiments de la liste initiale). Le nombre de nouveaux bâtiments correspondait à la fourchette prévue pour la période (plus de deux ans) écoulée entre la date des photographies aériennes et le dénombrement effectif. Cet écart a constitué une source de préoccupation majeure et un autre travail des «coordinateurs géographiques» a consisté à actualiser à la main les cartes papier établies pour toutes les tâches effectuées par le personnel de terrain.

VI. Qualité de la collecte directe de données (enquête par échantillonnage géographique visant à estimer le sous-dénombrement)

7. Ci-après sont présentés les cinq aspects de l'évaluation de la qualité du recensement intégré, ainsi que les divers outils utilisés.

a) Couverture du dénombrement: en moyenne, 5 % de logements en moins ont été localisés sur le terrain par rapport au nombre de ménages prévu *a priori* dans les zones échantillonnées, la fourchette étant comprise entre +6 et -21 % de logements dans différentes localités;

Résultats du dénombrement

Ménages interrogés	83 %
Questionnaire complet	78 %
Questionnaire partiel	5 %
Logements sans réponse (fermés, non occupés ou liés à une activité économique)	9 %
Refus de répondre	4 %

Chaque logement a été visité au moins cinq fois avant que la situation définitive ne soit déterminée.

b) Le supérieur de l'enquêteur a procédé à un nouveau dénombrement par téléphone pour évaluer la fiabilité des données (brèves questions de vérification) et a recontacté environ 5 % des ménages de l'échantillon pour vérifier qu'ils ont été personnellement interrogés. Des fautes ont été rarement décelées;

c) Tests de cohérence en ligne: les questionnaires informatisés (à l'aide du logiciel BLAISE) comprenaient des tests de cohérence intégrés. Deux types de paramètres de cohérence étaient utilisés, le premier étant un «avertissement», qui permettait de poursuivre l'entretien après avoir répondu à l'avertissement (par exemple, «Conjoint de même sexe que la personne interrogée, prière de vérifier» ou «Vérifier l'âge de l'enfant»). Le deuxième type de paramètre était une «erreur», qui empêchait de poursuivre l'entretien (par exemple «Année d'immigration antérieure à l'année de naissance, prière de vérifier les données saisies»). Le nombre d'«erreurs» et d'«avertissements» prouvés dans les questionnaires informatisés n'a pas encore été estimé;

d) L'évaluation en ligne de la validité et de la fiabilité: un système informatisé a été développé pour la «vérification en ligne» des données recueillies. Il s'est avéré indispensable pendant le recensement pilote final (répétition de 2006) lorsque que certains enquêteurs ont été suspectés d'avoir trouvé des moyens de réduire leur charge de travail

tout en étant pleinement rémunérés. Conjuguée aux tests de cohérence intégrés dans les questionnaires, la transmission quotidienne de données informatisées permettait pratiquement de vérifier en ligne les données transmises et le travail des enquêteurs. Cette procédure a permis de déceler les fautes et d'en rendre compte immédiatement aux supérieurs sur le terrain et au centre de recensement. Un ensemble d'interrogations prédéfinies a été exécuté chaque matin à propos des transmissions de données de la nuit précédente. Les interrogations n'ont pas été testées avant le recensement puisque les besoins et possibilités ont été définis après le test pilote final. De nouvelles interrogations ont été ajoutées tout au long du processus.

Exemples du système d'interrogation en ligne

<i>Type d'interrogation</i>	<i>Paramètre</i>	<i>Type de vérification</i>	<i>Définition opérationnelle</i>
Diminution intentionnelle du nombre de membres du ménage	Nombre de membres du ménage	Comparaison de la taille du ménage à la taille moyenne des ménages voisins	Plus d'un écart-type par rapport à la taille moyenne des ménages
	Actualisation fréquente concernant les membres du ménage	Suppression fréquente de personnes de la liste des membres du ménage	2 % à 5 % : suspect Plus de 5 % : improbable
Informations non obtenues de la personne interrogée	Numéros d'identification doubles	Plus d'un enregistrement ayant le même numéro d'identification dans tous les enregistrements précédents pour le même enquêteur	5 % à 10 % : suspect Plus de 10 % : improbable
	Numéro d'identification légitime erroné	Numéro d'identification ne figurant pas dans le registre de la population ou appartenant à une personne inactive	Moins de 5 % : suspect Plus de 5 % : improbable
	Dates de naissance réparties de manière irrégulière	Répartition de chaque élément de la date de naissance	30 % à 40 % : suspect Plus de 40 % : improbable
Questions sautées	Fréquence des réponses «Je ne fais pas d'études» ou «Je n'ai jamais fait d'études»	Nombres de personnes ayant répondu ainsi	10 % à 20 % : suspect Plus de 20 % : improbable
	Fréquence des réponses «Je ne sais pas» ou «Refus de répondre»	Nombre de questionnaires comportant plus de quatre réponses de ce genre	10 % à 20 % : suspect Plus de 20 % : improbable
Absence de variables essentielles	Absence du prénom	Pourcentage de personnes pour lesquelles la variable est absente	1 % à 5 % : suspect Plus de 5 % : improbable
	Absence du nom de famille	Pourcentage de personnes pour lesquelles la variable est absente	1 % à 5 % : suspect Plus de 5 % : improbable
	Absence de l'âge/de la date de naissance	Pourcentage de personnes pour lesquelles la variable est absente	1 % à 5 % : suspect Plus de 5 % : improbable

Chaque jour, les responsables de régions et le centre de recensement ont reçu par voie électronique une liste indiquant les informations suivantes: nom de l'enquêteur, type de faute commise, degré de chaque faute, nombre de questionnaires et nombres de relevés. Le système n'était pas parfait mais adéquat et très souple et, ainsi qu'il a été mentionné, des interrogations ont été ajoutées ou modifiées tout au long du processus;

e) Appariement entre les données recueillies et celles provenant du registre amélioré de la population et validation des résultats – qualité des paramètres d'identification obtenus. Les résultats de l'appariement sont présentés dans le graphique ci-après. L'appariement entre les données collectées sur le terrain et celles du registre amélioré est primordial pour la méthode du recensement intégré. Il n'est à première vue envisageable que pour les résidents israéliens et a été réalisé avec succès pour 98,3 % de résidents vivant dans des ménages et pour 95,8 % de résidents institutionnels, ce qui démontre que la qualité des informations d'identification obtenues dans le cadre du recensement était adéquate.

VII. Enquête concernant le remplissage de données (enquête par sondage/entretien téléphonique visant à estimer le surdénombrement) – Éléments d'évaluation de la qualité

8. Diverses procédures ont permis d'évaluer la qualité en rapport avec l'enquête concernant le remplissage de données:

a) Premièrement, le personnel du centre de recensement a écouté un échantillon d'entretiens et d'enregistrements vocaux;

b) Deuxièmement, un échantillon d'enquêteurs téléphoniques a fait l'objet d'un suivi continu de la part de professionnels du centre chargés d'évaluer la validité et la fiabilité des données. Les résultats de ce processus ne sont pas encore définitifs;

c) Troisièmement, il a été procédé à une estimation de la proportion de personnes et de ménages omis du dénombrement sur le terrain, ainsi que du nombre de personnes qui ont pu être jointes à leur adresse enregistrée lors de l'entretien téléphonique mais qui n'ont pas été dénombrées sur le terrain. On obtient ainsi une estimation des personnes omises du dénombrement sur le terrain. Dans l'enquête concernant le remplissage de données (entretien téléphonique), 42 337 personnes ont signalé que leur adresse «à la date du recensement» était identique à celle sous laquelle elles étaient inscrites dans le registre de la population et environ 1 % a été incontestablement omis lors de l'opération sur le terrain. Il s'agit des personnes dont le ménage a été interrogé mais dont le nom a été omis du questionnaire ou de personnes dont le ménage tout entier a été omis.

VIII. Procédures automatisées

9. Le recensement intégré fait appel à des procédures automatisées, seuls les éléments résiduels étant traités manuellement. L'édition et l'imputation sont entièrement automatiques, l'automatisation des procédures ayant pour but d'assurer l'actualité et la cohérence des renseignements.

a) Lorsque des réponses ont été imputées aux catégories «autres» et que des saisies ont été effectuées selon la réponse verbale de la personne interrogée, une procédure automatisée spéciale fondée sur des chaînes de texte vérifie qu'il n'y avait pas de catégorie imputable et 15 % des catégories «autres» ont été imputés à une catégorie déterminée;

b) Les branches d'activité économique et les principales professions ont fait l'objet d'un codage automatique, ainsi que 75 % des enregistrements, dont 5 % ont été de nouveau codés manuellement, le but étant d'estimer la qualité de l'automatisation. Les 25 % résiduels ont été codés manuellement grâce à un système assisté par ordinateur (ce qui a nécessité l'emploi de 25 personnes chargées du codage pendant huit mois) et 10 % des relevés codés manuellement ont été de nouveau codés de manière aléatoire par différentes personnes. Les codes incompatibles ont fait l'objet d'un nouveau codage par des spécialistes. Les résultats de l'évaluation de la qualité devraient être obtenus en novembre 2009;

c) Les informations socioéconomiques et démographiques font actuellement l'objet d'une édition et d'une imputation automatisées reposant sur les systèmes NIM et CANCEIS. Un document décrivant les règles et procédures sera présenté à la réunion de décembre 2009 de la CEE.

IX. Conclusion

10. Lorsqu'un appariement semble possible, les résultats du processus (qui n'est pas encore entièrement achevé puisqu'un traitement manuel est encore en cours) sont de 98,3 % pour les ménages et de 95,8 % pour les institutions. Les résultats de l'opération sur le terrain sont plus élevés que les 95 % prévus. Les résultats de l'appariement ayant trait aux sous-groupes de population «difficiles à contacter» (fondés sur les résultats du précédent recensement de 1995) sont également de l'ordre de 90 %. Cela est sans doute dû à une vaste campagne de promotion spécialement conçue pour chaque sous-groupe. Le recensement intégré basé sur deux grands échantillons (environ 20 % de la population) et un dénombrement complet des institutions ont permis d'améliorer les opérations et de concentrer les efforts sur les ménages échantillonnés et de réduire la charge imposée à la population.

11. Les résultats finals qui devraient être diffusés un an après le recensement sont obtenus grâce à un recours intensif à des procédures automatisées à toutes les étapes du processus. Le double système d'estimation permet d'évaluer la qualité du recensement sans faire appel à une «enquête postcensitaire».

Graphique
Appariement des résultats de l'enquête par sondage visant à estimer le sous-dénombrement

