

Rapport de méthodes

Estimation de la couverture du recensement de la population de l'an 2000

0

Statistische Grundlagen und Übersichten
Bases statistiques et vues d'ensemble
Basi statistiche e presentazioni generali

Procédure d'enquête
et plan d'échantillonnage
de l'enquête de couverture



Office fédéral de la statistique
Bundesamt für Statistik
Ufficio federale di statistica
Uffizi federal da statistica

Neuchâtel, 2002

Die vom Bundesamt für Statistik (BFS) herausgegebene Reihe «Statistik der Schweiz» gliedert sich in folgende Fachbereiche:

La série «Statistique de la Suisse» publiée par l'Office fédéral de la statistique (OFS) couvre les domaines suivants:

-
- 0 Statistische Grundlagen und Übersichten
 - 1 Bevölkerung
 - 2 Raum und Umwelt
 - 3 Arbeit und Erwerb
 - 4 Volkswirtschaft
 - 5 Preise
 - 6 Industrie und Dienstleistungen
 - 7 Land- und Forstwirtschaft
 - 8 Energie
 - 9 Bau- und Wohnungswesen
 - 10 Tourismus
 - 11 Verkehr und Nachrichtenwesen
 - 12 Geldpolitik, Finanzmärkte und -akteure
 - 13 Soziale Sicherheit
 - 14 Gesundheit
 - 15 Bildung und Wissenschaft
 - 16 Kultur, Medien, Zeitverwendung
 - 17 Politik
 - 18 Öffentliche Finanzen
 - 19 Rechtspflege
 - 20 Gesellschaft in Bewegung (Querschnittsthemen)

- 0 Bases statistiques et vues d'ensemble
 - 1 Population
 - 2 Espace et environnement
 - 3 Travail et vie active
 - 4 Economie nationale
 - 5 Prix
 - 6 Industrie et services
 - 7 Agriculture et sylviculture
 - 8 Energie
 - 9 Construction et logement
 - 10 Tourisme
 - 11 Transports et communications
 - 12 Politique monétaire, marchés et acteurs financiers
 - 13 Protection sociale
 - 14 Santé
 - 15 Education et science
 - 16 Culture, médias, emploi du temps
 - 17 Politique
 - 18 Finances publiques
 - 19 Droit et justice
 - 20 Société en mutation (thèmes interdisciplinaires)
-

Estimation de la couverture du recensement de la population de l'an 2000

Procédure d'enquête et plan d'échantillonnage
de l'enquête de couverture

Auteurs

Anne Renaud et Philippe Eichenberger

Office fédéral de la statistique

Editeur

Office fédéral de la statistique

Préambule

Le projet «estimation de la couverture du recensement» fait partie des projets visant à évaluer la qualité du recensement de la population de l'an 2000. Il a été initié au service de méthodes statistiques (METH) de l'OFS puis transmis à la section de la structure de la population et des ménages (BHS) de l'OFS pour la réalisation de l'enquête. Les travaux décrits dans ce rapport ont été effectués par Anne Renaud et Philippe Eichenberger de METH.

Ce projet a été possible grâce à la participation de nombreuses personnes internes et externes à l'OFS. Nous désirons remercier Rachel Fritschi de la section BHS pour l'organisation de l'enquête, Romain Douard, Jitka Fligr et André Zinggeler de la section de l'utilisation du territoire (RAUM) de l'OFS pour les données du REAB et les cartes géographiques, Emerete Gauro et Floriano Bertozzi de l'Ufficio di statistica del Cantone Ticino (Ustat) pour les données du Tessin, Flavia Molinari et Giovanni Longu du service linguistique (LING) de l'OFS pour l'aide dans le tri des codes des bâtiments tessinois, et Markus Eichenberger de l'office fédéral de l'informatique et de la télécommunication (BIT) pour les programmes SAS de traitement des listes de ménages provisoires.

Résumé

Les défauts de couverture d'un recensement de la population en Suisse sont estimés pour la première fois dans le cadre du recensement de l'an 2000 (RFP2000).

L'enquête post-censitaire nécessaire pour les estimations selon le système dual est nommée enquête de couverture (EC2000). Elle est basée sur un plan d'échantillonnage à plusieurs niveaux avec deux bases d'échantillonnage indépendantes au premier niveau (communes au Tessin et numéros postaux dans le reste de la Suisse). La méthode d'échantillonnage permet de contrôler les tailles des échantillons aux différents niveaux, tout en limitant la variabilité des poids d'échantillonnage et en tenant compte des difficultés pratiques de l'enquête sur le terrain.

Durant l'enquête, on crée la liste des ménages dans un échantillon d'environ 18'000 bâtiments. Le sous-échantillon final de 15'800 bâtiments comporte 27'000 ménages pour lesquels on recueille les informations de base du recensement. Cet échantillon de ménages permet d'atteindre le nombre désiré de personnes dans les groupes pour lesquels on désire une estimation de la couverture.

Mots-clé

rapport de méthodes; recensement de la population; VZ2000; RFP2000; enquête de couverture; EC2000; méthodes; estimations; couverture; sous-dénombrement; plan d'échantillonnage; plusieurs niveaux; stratification; allocation; post-enumeration; coverage survey.

Complément d'information:	Anne Renaud, tél. 032 713 62 65 Anne.Renaud@bfs.admin.ch
Réalisation:	Service de méthodes statistiques, OFS
Diffusion:	Office fédéral de la statistique CH-2010 Neuchâtel Tél. 032 713 60 60 / Fax 032 713 60 61 Order@bfs.admin.ch
Internet:	http://www.statistik.admin.ch
Numéro de commande:	338-0009
Prix:	gratuit
Série:	Statistique de la Suisse
Domaine:	0 Bases statistiques et vues d'ensemble
Langue du texte original:	Français
Graphisme/Layout:	OFS
Copyright:	OFS, Neuchâtel 2002 La reproduction est autorisée, sauf à des fins commerciales, si la source est mentionnée.
ISBN:	3-303-00242-8

Table des matières

1	Introduction	5
2	Procédure de l'enquête de couverture EC2000	6
3	Procédure d'échantillonnage	10
4	Plan d'échantillonnage NORD	12
4.1	Données	12
4.2	Méthode d'échantillonnage	14
4.3	Plan au niveau des numéros postaux	20
4.4	Plan au niveau des tournées	26
4.5	Plan au niveau des bâtiments	28
4.6	Sous-échantillonnage des bâtiments	30
5	Plan d'échantillonnage TICINO	34
5.1	Données	34
5.2	Méthode d'échantillonnage	35
5.3	Plan au niveau des communes	36
5.4	Plan au niveau des bâtiments	38
5.5	Sous-échantillonnage de bâtiments	44
6	Échantillon final de l'EC2000	46
7	Conclusion	47
A	Recensement de la population de l'an 2000	49
B	Enquête pilote de l'EC2000	50
C	Listes provisoires de ménages	51
D	Échantillon des numéros postaux (plan NORD)	53
E	Programmes SAS et S-PLUS	55

1 Introduction

Dans tout recensement, certaines personnes ne reçoivent pas de questionnaires et d'autres sont comptées deux fois. L'effet global de ces défauts de couverture produit habituellement un sous-dénombrement net de la population. Le sous-dénombrement est souvent limité en moyenne nationale, mais il prend des valeurs plus importantes pour certains groupes de personnes (ex. jeunes hommes, très jeunes enfants ou personnes âgées). On parle de sous-dénombrement global et de sous-dénombrement différencié. Quelques exemples sont donnés dans les Tables 1 et 2.

TAB. 1 – Exemples de taux de sous-dénombrement globaux. Références : Brown et al. (1999), McLennan (1996), ABS (1997), Statistique Canada (1999), Hogan (1993) et Davis (2001).

Pays	Taux et année
Royaume-Uni	<1% (avant 1991) ; 2.2% (1991)
Australie	1.9% (1986) ; 1.8% (1991) ; 1.6% (1996)
Nouvelle-Zélande	1.2% (1996)
Canada	2.5% (1996)
États-Unis	1.6% (1990), 1.2% (2001)

TAB. 2 – Exemples de taux de sous-dénombrement différenciés. Références : Brown et al. (1999), McLennan (1996) et Hogan (1993).

Pays	Taux et année
Royaume-Uni	>20% pour jeunes hommes dans les grandes villes (1991)
Australie	1.1% - 3.1% selon le "State/Territory" (1996)
Australie	2.0% pour les hommes et 1.1% pour les femmes (1996)
Australie	4.3% pour les hommes entre 20 et 24 ans (1996)
États-Unis	<1% pour les "Whites" ; 4.5 - 5% pour les "Blacks" et "Hispanics" (1990)

La méthode d'estimation des défauts de couverture est habituellement basée sur le système dual, voir par exemple Wolter (1986) pour la méthode, Fienberg (1992) pour une bibliographie et Hogan (1992) pour une application aux États-Unis. L'enquête post-censitaire requise pour l'estimation duale est nommée post-enumeration survey ou coverage survey dans les pays anglophones. Cette enquête est organisée peu après le recensement et consiste à énumérer - indépendamment du recensement - les personnes dans un échantillon d'unités géographiques dans le pays. La comparaison des résultats de l'enquête avec ceux du recensement permet alors de faire des estimations du sous-dénombrement global et différencié selon différents groupes d'intérêt.

Les défauts de couverture d'un recensement de la population sont estimés pour la première fois en Suisse dans le cadre du recensement de l'an 2000 (RFP2000). L'enquête

post-censitaire est nommée enquête de couverture (EC2000) et consiste à énumérer les ménages et personnes dans un échantillon de bâtiments en Suisse, indépendamment et quelques mois après le jour du recensement (5 décembre 2000). Les ressources permettent de traiter un échantillon de 27'000 ménages répartis dans environ 300 communes ou numéros postaux. On désire avoir environ 1'000 ménages au Tessin et 26'000 ménages dans le reste de la Suisse.

Le projet a comme but l'estimation des défauts de couverture classiques des recensements (ex. différences entre les hommes et les femmes, différences entre villes et campagnes). De plus, il vise à obtenir des résultats différenciés selon la variante de relevé et donc à détecter d'éventuelles problèmes spécifiques au RFP2000. (Voir Annexe A pour des informations sur le recensement et les variantes de relevé.)

Ce rapport décrit la procédure d'enquête et le plan d'échantillonnage de l'EC2000. Une version résumée et en anglais est disponible dans Renaud (2002).

2 Procédure de l'enquête de couverture EC2000

Les procédures d'enquêtes choisies pour les enquêtes post-censitaires à l'étranger sont étroitement liées aux structures administratives du pays, aux données existantes, aux buts et aux ressources disponibles. Le point commun est l'utilisation de données indépendantes du recensement afin de satisfaire aux hypothèses de la méthode duale. Les données de ce type n'étant en général pas disponibles, elles doivent être récoltées spécialement pour les besoins de l'enquête.

Pour avoir des estimations des taux de couverture de bonne qualité, l'enquête doit avoir lieu le plus tôt possible après le jour du recensement pour éviter de grands changements, et pas trop tôt après le jour du recensement pour avoir un maximum de questionnaires du recensement en retour et éviter des interactions.

La procédure d'enquête de l'EC2000 se résume dans les cinq phases suivantes :

1. sélection d'un échantillon de bâtiments ;
2. élaboration de la liste des ménages habitant les bâtiments de l'échantillon ;
3. sous-échantillonnage des bâtiments afin de réduire la taille et atteindre les 27'000 ménages requis ;
4. contacts et interviews avec les ménages habitant les bâtiments de l'échantillon ;
5. traitement des données, estimations, analyses et documentation.

Une enquête pilote a permis de tester la procédure d'enquête et les formulaires. Elle est brièvement décrite dans l'Annexe B.

Notons que la procédure d'enquête de l'EC2000 fait l'hypothèse que toutes les personnes résidant en Suisse habitent dans des bâtiments et que nous possédons une liste exhaustive des bâtiments habités. Pour maximiser la couverture de la population cible (population résidente), on utilisera donc des listes avec les bâtiments codés comme habités et les bâtiments codés comme inhabités (ex. industrie).

Les interviews de l'EC2000 ont eu lieu entre la mi-avril et la fin mai 2001, c'est-à-dire après le deuxième rappel du recensement, lorsque le taux de retour était env. de 95%. Le contrôle des listes de ménages par les facteurs était échelonné entre la mi-janvier et la fin février 2001. Elles ont donc été créées de 1.5 à 2.5 mois après le recensement, période durant laquelle on observe peu de déménagements.

Phase 1 : échantillon de bâtiments

L'échantillon des bâtiments de l'EC2000 est construit en deux parties. La première partie recouvre le territoire Suisse sans le Tessin (échantillon NORD) et la deuxième partie recouvre le Tessin (échantillon TICINO). L'échantillon des bâtiments de l'EC2000 est sélectionné à l'OFS. La procédure d'échantillonnage est décrite dans les sections 3 à 6.

Phase 2 : liste des ménages habitant les bâtiments de l'échantillon

La phase 2 a pour but de créer la liste exhaustive des ménages résidants dans les bâtiments de l'échantillon. Elle est très importante car la qualité des estimations de la couverture repose en bonne partie sur la qualité de cette liste. Elle demande aussi un grand travail sur le terrain car aucune liste de ménages n'existe actuellement en Suisse.

Une liste provisoire de ménages est tout d'abord créée sur la base de l'annuaire de téléphone électronique (ETV). On recherche dans ETV les entrées correspondant aux adresses des bâtiments de l'échantillon à l'aide du logiciel ETV Batchman de l'entreprise NOVA ESPRIT. Notons que les listes provisoires ne peuvent être créées que pour les bâtiments ayant une adresse postale complète. Les listes obtenues avec ETV Batchman sont nettoyées afin de retirer les entrées multiples et les entreprises (programme SAS développé par Markus Eichenberger, BIT). Les numéros de téléphone trouvés dans ETV sont conservés en vue des interviews. (Voir l'Annexe C pour les détails de la démarche.)

Des formulaires bâtiments sont ensuite imprimés avec les listes provisoires des ménages trouvés dans ETV. Ces formulaires sont envoyées aux offices de Poste concernés et les employés postaux les contrôlent et les actualisent. Les facteurs indiquent aussi le type de bâtiment (exclusivement, partiellement ou pas à usage d'habitation) et l'occupation des logements (temporaire ou à l'année). Les facteurs reçoivent des cartes géographiques pour identifier les bâtiments sans adresses postales complètes (cartes créées en collaboration avec la section RAUM de l'OFS).

Notons qu'une étape préliminaire est nécessaire au Tessin car l'échantillon des bâtiments comporte plusieurs bâtiments difficiles à identifier sur le terrain (sans adresse et sans coordonnées). Les communes sont mises à contribution pour identifier les bâtiments sur des cartes à partir des données du cadastre et du nom du propriétaire. Les bâtiments a priori vides (ex. poulailler) sont retirés de la liste avant l'envoi aux offices de Poste.

La saisie des corrections des facteurs permet d'obtenir une liste finale des ménages dans les bâtiments de l'échantillon. Les numéros de téléphone sont recherchés pour les ménages ajoutés par les facteurs.

Les listes provisoires, l'impression des formulaires de bâtiments et la saisie des corrections

sont réalisées à l'OFS. Le contact avec les offices de Poste est sous la responsabilité de l'entreprise Data Care AG de Kriens, filiale de la Poste.

Phase 3 : sous-échantillonnage des bâtiments

Le nombre de ménages habitant un bâtiment donné n'est pas connu. Il est donc difficile de faire une estimation précise du nombre de ménages dans les bâtiments de l'échantillon de la phase 1. La procédure consiste donc à prendre un nombre suffisant de bâtiments dans la phase 1 et de sous-échantillonner ensuite afin d'obtenir les 27'000 ménages désirés. Le sous-échantillonnage est effectué à l'OFS. La méthode est décrite dans les sections 4.6 (plan NORD) et 5.5 (plan TICINO).

Phase 4 : contact avec les ménages et interviews

Le contact avec les ménages se fait par téléphone (CATI) si un numéro de téléphone est connu et en face-à-face (CAPI¹) dans le cas contraire. Des cartes permettent aux interviewers CAPI d'identifier les ménages habitant les bâtiments sans adresse postale complète. Seule la prise de contact varie entre les interviews CATI et CAPI. Les questions sont identiques, assistées par ordinateur et possibles en 10 langues afin de limiter la non-réponse due à des problèmes de langue. Les réponses aux questions sur le ménage et les personnes qui le composent sont relevées auprès d'un unique membre du ménage ayant au moins 15 ans (proxi).

L'interview se déroule comme suit : (1) questions communes au ménage (contrôle de l'adresse, relevé de l'étage et du nombre de pièces du logement) ; (2) liste des membres du ménage (noms et prénoms) ; et (3) pour chaque membre du ménage : réponse à un questionnaire individuel (sexe, date de naissance, état civil, nationalité, langue, situation dans le ménage, vie active, éventuel deuxième domicile et déménagement entre le jour du recensement et le jour de l'interview). Les questions sont principalement reprises du questionnaire du recensement et reformulées pour le contact par téléphone. Une version courte du questionnaire (nombres globaux) est disponible pour les cas de ménages collectifs (ex. homes).

Les questions permettent l'appariement entre les données du recensement et de l'enquête, la recherche de personnes à d'autres adresses (déménagements, deux domiciles) et la détermination des caractéristiques des personnes pour une estimation du sous-dénombrement global et différencié.

Les interviews sont réalisées par l'institut de sondage IHA-GfM.

Phase 5 : traitement des données, estimations, analyse et documentation

La population cible pour les estimations est la population résidente en Suisse, selon le domicile économique, vivant dans un ménage privé et dans l'état au jour du recensement,

¹CATI : Computer Assisted Telephone Interview. CAPI : Computer Assisted Personal Interview.

voir les définitions dans l'Annexe A. Les personnes résidentes dans des ménages collectifs sont exclues des estimations.

On définit trois échantillons au niveau des personnes :

- P-sample : échantillon des personnes de l'EC2000 dans les bâtiments de l'EC2000 ;
- E-sample : échantillon des personnes recensées dans les bâtiments de l'EC2000 ;
- E-sample étendu : échantillon des personnes recensées dans les bâtiments de l'EC2000 et dans les bâtiments alentours (domaines à définir).

On dit qu'une énumération (record) du recensement est correcte si (1) il s'agit d'une vraie personne et non pas d'un enregistrement fantaisiste (Mickey Mouse ou Pussi-le-chat), (2) la personne est née avant le jour du recensement ou décédée après le jour du recensement, (3) l'enregistrement contient suffisamment d'information pour identifier la personne (au moins un nom et un prénom), (4) l'enregistrement n'est pas un double, (5) la personne est incluse dans le recensement là où elle devrait l'être (géographiquement et au domicile économique). Une énumération ne satisfaisant pas les critères d'une énumération correcte est dite erronée.

L'estimateur DSE du total selon la méthode duale a la forme suivante (méthode du US Census Bureau, voir par ex. Griffin, 2000) :

$$DSE = C \frac{CE}{N_e} \left(\frac{M}{N_p} \right)^{-1} \quad (1)$$

avec C le total recensé, N_e le total pondéré des enregistrements dans le E-sample étendu, CE le total pondéré des enregistrements corrects dans le E-sample étendu, N_p le total pondéré des enregistrements dans le P-sample et M le total pondéré des enregistrements du P-sample appariés avec le recensement. Les éléments de l'équation (1) sont illustrés dans la Figure 1.

Le P-sample permet d'estimer la sous-couverture ; via l'appariement avec le recensement. Le E-sample étendu permet d'estimer la sur-couverture ; via l'identification des enregistrements erronés. Les domaines de recherche pour l'appariement et la détection des enregistrements erronés sont identiques afin d'éviter des biais dans les estimations ("balancing error", voir par exemple Navarro, 2000).

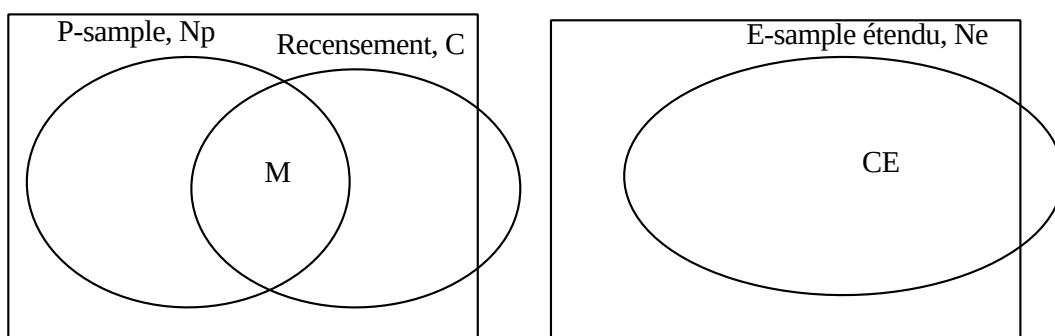


FIG. 1 – Illustration des termes C , N_p et M (ovales et intersection, à gauche), et N_e et CE (ovale et intersection, à droite). Les cadres représentent la population cible.

L'estimateur dual est basé sur deux hypothèses principales : l'hypothèse d'indépendance entre le recensement et l'enquête post-censitaire et l'hypothèse que toutes les personnes ont la même probabilité d'être énumérées. L'hypothèse d'indépendance est respectée le mieux possible par l'utilisation de jeux de données indépendants du recensement et le choix de la procédure d'enquête de l'EC2000. Des contaminations entre EC2000 et recensement sont cependant possible, par exemple via une corrélation entre la non-réponse à l'EC2000 et un raz-le-bol face au recensement, ses rappels et les demandes de renseignements pour les questionnaires partiellement remplis. L'hypothèse d'homogénéité des probabilités d'énumération est abordée par le biais d'une application de la méthode duale dans un ensemble de post-strates contenant des personnes ayant un comportement similaire face au recensement (ex. hommes entre 20 et 30 ans habitant dans des villes de plus de 10'000 habitants). Les résultats sont ensuite combinés pour obtenir des estimations pour les différents groupes d'intérêt.

Le traitement des données est basé sur la procédure utilisée aux USA en 1990 et en 2000 (voir par exemple Childers, 2000). Les étapes importantes recouvrent la préparation des données de l'EC2000 (P-sample), la recherche des données dans le recensement (E-sample restreint et étendu), l'appariement entre le P-sample et le recensement avec un traitement différencié pour les personnes ayant déménagé ou pas entre le jour du recensement et l'EC2000, la détermination des enregistrements erronés dans le E-sample étendu et la codification des différents résultats.

La méthode d'estimation commence par la définition des échantillons finaux puis traite l'imputation des status manquants (type de domicile, enregistrement correct ou match non résolu), la pondération (y compris le modèle de non-réponse), la définition des post-strates, les estimations duales dans les post-strates, les estimations synthétiques dans les différents groupes d'intérêt et l'estimation de la variance des résultats.

Différentes évaluations, notamment celle de la base de sondage des bâtiments, permettront de donner des informations complémentaires sur les défauts de couvertures estimés, leur précision et leur qualité. Les résultats seront ensuite analysés, documentés et préparés pour la diffusion.

3 Procédure d'échantillonnage

Deux plans d'échantillonnage indépendants sont développés pour l'EC2000. Le plan pour le territoire de la Suisse sans le Tessin (NORD) et le plan pour le Tessin (TICINO). Le partage géographique est dû au fait que le Tessin organise le recensement de manière indépendante et que les sources de données au niveau des bâtiments n'ont pas la même structure. Les résultats sont ensuite combinés pour avoir un unique jeu de données pour toute la Suisse.

Les deux plans d'échantillonnages sont des plans à plusieurs niveaux (voir Table 3) avec stratification au premier niveau (numéros postaux et communes). On fait un relevé exhaustif des ménages et des personnes dans les bâtiments afin de simplifier la procédure d'appariement avec les données du recensement. La différence principale entre les deux plans se trouve dans l'utilisation ou non du découpage postal de la Suisse. Les adresses

des bâtiments du Tessin ne sont en effet que très partielles dans le jeu de données à disposition.

TAB. 3 – Niveaux des plans d'échantillonnages NORD et TICINO.

Plan NORD	Plan TICINO
(1) numéros postaux (NPA)	(1) communes
(2) circonscriptions de distribution de la Poste (tournées)	
(3) bâtiments	(2) bâtiments
(4) ménages (exhaustif)	(3) ménages (exhaustif)
(5) personnes (exhaustif)	(4) personnes (exhaustif)

Le plan d'échantillonnage à plusieurs niveaux permet de réduire les coûts d'enquête car il regroupe du travail sur le terrain et limite le nombre d'offices de Poste à contacter. Il implique cependant une perte de précision par l'effet de grappe qu'il induit. La stratification au premier niveau permet non seulement de contrebalancer une partie de la perte de précision due à l'effet de grappe mais aussi de contrôler la taille de l'échantillon dans les différents groupes d'intérêt (ex. variantes de relevé, tailles de communes). Elle permet également de maîtriser les coûts en favorisant certaines unités primaires au détriment d'autres unités primaires dans lesquels on prévoit des difficultés dans la réalisation de l'enquête sur le terrain.

L'échantillon NORD est choisi sur la base des données du Registre des adresses de bâtiments (REAB) de l'OFS. Le REAB est utilisé pour le recensement des bâtiments et logements mais pas pour le recensement des personnes. L'exigence d'indépendance des sources de l'EC2000 et du recensement est donc conservée au niveau des ménages et personnes.

L'échantillon TICINO est choisi sur la base des données du registre Tessinois des bâtiments. Ce registre est construit sur la base d'une statistique fiscale exhaustive ("Dichiarazione per la stima ufficiale della sostanza immobiliare", Stima). La Stima n'a pas servi pour le recensement des personnes au Tessin. L'indépendance des sources entre recensement et enquête est donc également respectée.

La différence entre la population cible (population résidante) et la population d'échantillonnage (population résidant dans les bâtiments de la liste) est difficile à estimer. De part le ratissage large au niveau des bâtiments, on estime qu'elle est petite. Une évaluation du cadre d'échantillonnage principal (liste de bâtiments) est prévue dans le cadre du projet. Elle permettra de déterminer les problèmes de couverture de la population cible due à la qualité des listes de bâtiments à disposition.

La relation entre la taille de l'échantillon et la précision des résultats obtenus avec le système dual est difficile à évaluer. Les défauts de couverture n'ont pas été estimés dans les derniers recensements et le RFP2000 intègre de nombreuses innovations dans la méthode de relevé. On prend donc les résultats obtenus à l'étranger comme base pour nos estimations, notamment ceux obtenus en Australie en 1996 (McLennan, 1996). Ce pays avait en effet organisé une enquête de couverture (post-enumeration survey) avec un plan d'échantillonnage à plusieurs niveaux (districts, blocks, logements) qui se rapproche de la

structure du plan choisi en Suisse.

Hypothèse pour les estimations de précision : les données pour 10'000 personnes (net) permettent d'obtenir un écart-type de 0.3% pour un sous-dénombrement entre 1.2 et 1.8%. La non-réponse étant estimée à 15% des personnes interrogées, on estime que les données de 11'800 personnes (brut) permettent d'obtenir des résultats avec une précision acceptable dans un groupe d'intérêt.

Le plan d'échantillonnage est construit de façon à obtenir des résultats acceptables (au sens de l'hypothèse de précision retenue) dans les grands groupes démographiques (hommes-femmes, grandes classes d'âge, etc.), les classes de taille des communes (moins et plus de 10'000 habitants) et les variantes de relevé du recensement CLASSIC et TRANSIT.

Les 27'000 ménages (brut) de l'EC2000 sont répartis en deux régions : 1'000 ménages dans 20 communes au Tessin (plan TICINO) et 26'000 ménages dans 285 NPA du reste de la Suisse (plan NORD).

4 Plan d'échantillonnage NORD

Le plan d'échantillonnage NORD (Suisse sans le Tessin) est construit de façon à atteindre les tailles d'échantillon minimales dans les différents groupes d'intérêt. Il vise également à obtenir des probabilités d'inclusion contrôlées pour tous les bâtiments de chacune des variantes de relevé. L'introduction de probabilités inégales permet de tenir compte de la difficulté du relevé sur le terrain. On évitera cependant une trop grande variabilité des poids car les valeurs extrêmes, toujours difficiles à traiter, se sont révélées extrêmement dommageables dans le cadre des estimations de la couverture des recensements aux États-Unis.

4.1 Données

Le plan d'échantillonnage utilise diverses sources de données. Ces données permettent de définir les bases de sondage aux différents niveaux, les informations auxiliaires pour la stratification au premier niveau et les estimations de la taille des échantillons de personnes et de ménages.

4.1.1 Répertoire d'adresses de bâtiments (REAB)

Le répertoire d'adresses des bâtiments (REAB ; section RAUM de l'OFS) combine les données du recensement de 1990 et la base de données "Géocodage". Il est complété par les données actuelles des gérances immobilières et par des informations venant des communes. Ce répertoire est utilisé comme base pour le recensement 2000 des bâtiments et logements (préimpression de bordereaux, état au début novembre 2000). Le REAB comporte environ 1.6 million de bâtiments, définis comme des entités indépendantes ou séparées par un mur mitoyen.

Dans le plan d'échantillonnage, on utilise les variables identifiantes des bâtiments (code

de référence, adresse, coordonnées géographiques XY) et diverses autres variables telles la variable OBJET (Table 4). Pour les besoins de la procédure d'enquête, on prend également en compte les données sur les personnes de contact (nom et adresse). Notons que les coordonnées hectométriques disponibles dans certaines régions sont ignorées car elles ne sont pas suffisamment précises pour identifier des bâtiments (ex. BL et ZH).

TAB. 4 – Variable OBJET dans le REAB.

OBJET	Description
1	avec utilisation à des fins d'habitation
2	sans fin d'utilisation à des fins d'habitation
4	habitation mobile/provisoire
7	bâtiment fictif (1 par commune, "Kirchspitze")
8	bâtiment inconnu
9	bâtiment démoli

Le REAB forme la source principale des données du plan NORD. Tous les bâtiments du REAB, excepté les bâtiments fictifs (OBJET=7) sont retenus. On ne prend donc pas uniquement les bâtiments recensés (OBJET=1 et 4). Ce choix élargi permet de maximiser la couverture de la population cible.

Chaque bâtiment du REAB possède un numéro postal (NPA). L'ensemble des NPA présents dans le REAB forme la base de sondage au premier niveau d'échantillonnage.

4.1.2 Données sur les communes

La base de donnée ADMIN DB du recensement 2000 (section BHS de l'OFS) contient la liste des communes politiques de la Suisse avec des caractéristiques telles que la langue administrative, le canton et la variante de relevé du recensement (état au 15 novembre 2000).

Les variantes de relevé sont regroupées dans les 4 catégories principales (CLASSIC, SEMI-CLASSIC, TRANSIT et FUTURE) sans tenir compte des spécificités liées à l'organisation des envois des questionnaires (mail management).

Les données de la statistique de l'état annuel de la population (ESPOP ; section BEV de l'OFS) sont utilisées pour les estimations de la population résidante dans les communes. Les données au 1er janvier 1991 et fin 1999 sont prises en compte.

4.1.3 Estimation de la population dans les numéros postaux

Les données du géocodage (section RAUM de l'OFS) font le lien entre les données au niveau des numéros postaux et les données au niveau des communes. Ces données fournissent, par hectare, le nombre de ménages et de personnes à l'état du recensement de 1990. La définition géographique des NPA (actualisée dans l'état du 1er janvier 1999) permet alors (1) de déterminer la commune majoritaire (en terme de population) sur le territoire de chaque numéro postal, et (2) d'estimer la part de la population de la commune majoritaire se trouvant dans le numéro postal.

Les hypothèses suivantes sont retenues pour les estimations du nombre de ménages et de personnes à la fin 1999 dans chaque numéro postal : (1) la population du numéro postal a augmenté depuis 1990 selon le même taux que sa commune majoritaire, et (2) le nombre moyen de personnes par ménage est égal à celui de 1990. Si les données de population d'une commune ne sont pas connues pour 1990 (ex. nouvelles communes), le taux d'augmentation global de la Suisse (6%) remplace le taux communal.

Notons qu'une partie de la population du recensement de 1990 n'a pas pu être assignée aux NPA de 1999 (env. 2%). Les données de population par NPA sont donc en moyenne un peu sous-estimées.

4.1.4 Circonscriptions de distribution

Les circonscriptions de distribution de la Poste sont des regroupements de rues et lieux-dits utilisés pour l'acheminement des lettres et colis. Elles correspondent à des tournées de facteurs et sont définies dans le fichier des circonscriptions (Botenfile de la Poste). Un programme de la Poste permet d'assigner le numéro postal de distribution (parfois différent de celui de l'adresse postale) et le numéro de circonscription aux adresses données en entrée. Il a été appliqué pour le tri du courrier lors de l'envoi des questionnaires du recensement. Les circonscriptions ne peuvent être assignées que si (1) des circonscriptions sont définies dans le numéro postal, (2) l'adresse postale est complète (rue, numéro de maison, numéro postal) et (3) l'adresse postale existe dans le Botenfile. L'entreprise DCL a appliqué le programme sur la liste des bâtiments du REAB se trouvant dans les NPA tirés au premier niveau d'échantillonnage. Cette étape permet la création de la base de sondage au deuxième niveau.

4.1.5 Listes de ménages et de personnes

Les listes de ménages et de personnes dans les bâtiments de l'échantillon sont créées dans le cadre de l'EC2000 (voir la procédure d'enquête, section 2). Les listes de ménages sont élaborées en deux temps (1) listes provisoires sur la base de l'annuaire électronique ETV et (2) adaptation des listes par les facteurs. Les listes des personnes dans les ménages de l'échantillon sont créées au début des interviews avec les ménages.

4.2 Méthode d'échantillonnage

La méthode d'échantillonnage du plan NORD utilise des tirages proportionnels à la taille sans remplacement (PPS sampling) au premier et au deuxième niveau (NPA et tournées), puis aléatoire simple sans remplacement (SRS sampling) au troisième niveau (bâtiments). La mesure de taille dans les tirages PPS est une fonction du nombre de bâtiments dans l'unité d'échantillonnage.

Pour une distribution équitable du travail entre les offices de Poste et les facteurs, on tire un nombre fixe *TCS* (Target Cluster Size) de bâtiments dans chaque NPA. Ces *TCS* bâtiments sont ensuite répartis entre les tournées sélectionnées dans le NPA (par ex.

$TCS = 60$ bâtiments répartis en 3 tournées de 20). Tous les bâtiments sont pris dans les NPA avec moins que TCS bâtiments et la répartition en tournées n'est réalisées que si l'information sur les tournées est disponible dans le NPA.

4.2.1 Niveau des numéros postaux

L'ensemble $U_{(NPA)} = \{1, \dots, i, \dots, M\}$ des NPA de la région NORD est partitionné en H strates de tailles M_h , $h = 1, \dots, H$. Le nombre de bâtiments dans le NPA i de la strate h est noté B_{ih} .

Au premier niveau d'échantillonnage, on tire m_h NPA dans chaque strate h . L'échantillon des NPA est noté $s_{(NPA)}$. Il est de taille $m = \sum m_h$.

La probabilité de sélection π_{ih} du NPA i de la strate h est donnée par :

$$\pi_{ih} = P(i \in s_{(NPA)}) = m_h \frac{B_{ih}^+}{\sum_{i \in h} B_{ih}^+} \quad (2)$$

où $B_{ih}^+ = \max(B_{ih}, TCS)$ est la mesure de taille définie pour le tirage PPS.

La mesure de taille B_{ih}^+ du NPA i de la strate h est égale au nombre de bâtiments si $B_{ih} \geq TCS$. Elle est égale à TCS si $B_{ih} \leq TCS$. Les avantages de l'utilisation de cette mesure de taille sont décrits dans la section 4.2.5.

4.2.2 Niveau des tournées

Soit $U_{(tour,ih)} = \{1, \dots, j, \dots, C_{ih}\}$ l'ensemble des tournées dans le NPA $i = 1, \dots, M_h$ de la strate $h = 1, \dots, H$. Le nombre de bâtiments dans la tournée j du NPA i de la strate h est noté B_{jih} .

Au deuxième niveau d'échantillonnage, on tire c_{ih} tournées dans le NPA i de la strate h . L'échantillon des tournées est noté $s_{(tour)}$. Il est de taille $c = \sum \sum c_{ih}$.

La probabilité d'inclusion $\pi_{j|ih}$ de la tournée j du NPA i de la strate h , sachant que le NPA i est dans l'échantillon $s_{(NPA)}$, est donnée par :

$$\pi_{j|ih} = P(j \in s_{(tour)} | i \in s_{(NPA)}) = c_{ih} \frac{B_{jih}^+}{\sum_{j \in (i,h)} B_{jih}^+} \quad (3)$$

où $B_{jih}^+ = \max(B_{jih}, TCS/c_{ih})$ est la mesure de taille de la tournée j du NPA i de la strate h .

La mesure de taille B_{jih}^+ au niveau des tournées est du type de celle au niveau des NPA B_{ih}^+ . L'avantage de ce choix est démontré plus bas.

4.2.3 Niveau des bâtiments

Soit $U_{(bat,jih)} = \{1, \dots, k, \dots, B_{jih}\}$ l'ensemble des bâtiments de la tournée $j = 1, \dots, C_{ih}$ du NPA $i = 1, \dots, M_h$ de la strate $h = 1, \dots, H$.

Au troisième niveau de l'échantillonnage, on tire $b_{jih} = \min(B_{jih}, TCS/c_{ih})$ bâtiments dans la tournée j du NPA i de la strate h . L'échantillon des bâtiments est noté $s_{(bat)}$. Il est de taille $b = \sum \sum \sum b_{jih}$.

La probabilité d'inclusion $\pi_{k|jih}$ du bâtiment k de la tournée j dans le NPA i de la strate h , sachant que la tournée j fait partie de l'échantillon $s_{(tour)}$, est donnée par :

$$\pi_{k|jih} = P(k \in s_{(bat)} | j \in s_{(tour)}) = \frac{b_{jih}}{B_{jih}} = \frac{TCS/c_{ih}}{B_{jih}^+}. \quad (4)$$

Complément : L'égalité entre les deux derniers termes de l'équation découle de la définition de la mesure de taille B_{jih}^+ des tournées, en appliquant les deux propriétés suivantes (1) Soient a, b , et $c \in \mathbb{R}$ et $c > 0$ alors $\min(a, b)/c = \min(a/c, b/c)$, et (2) Soient a et $b \in \mathbb{R}$ et $ab > 0$ (a et b strictement positifs ou strictement négatifs), alors $\min(a, b) = 1/\max(1/a, 1/b)$. En effet :

$$\begin{aligned} \frac{b_{jih}}{B_{jih}} &= \frac{\min(B_{jih}, TCS/c_{ih})}{B_{jih}} = \min(1, \frac{TCS/c_{ih}}{B_{jih}}) \\ &= \frac{1}{\max(1, \frac{B_{jih}}{TCS/c_{ih}})} = \frac{TCS/c_{ih}}{\max(TCS/c_{ih}, B_{jih})} = \frac{TCS/c_{ih}}{B_{jih}^+} \end{aligned}$$

4.2.4 Probabilité d'inclusion globale des bâtiments

La probabilité d'inclusion π_k du bâtiment k de la tournée j dans le NPA i de la strate h dans l'échantillon des bâtiments $s_{(bat)}$ est donnée par :

$$\begin{aligned} \pi_k &= P(k \in s_{(bat)}) = \pi_{ih} \pi_{j|ih} \pi_{k|jih} \\ &= m_h \frac{B_{ih}^+}{\sum_{i \in h} B_{ih}^+} c_{ih} \frac{B_{jih}^+}{\sum_{j \in (i, h)} B_{jih}^+} \frac{TCS/c_{ih}}{B_{jih}^+} \\ &= m_h \frac{TCS}{\sum_{i \in h} B_{ih}^+} \rho_{ih} \end{aligned} \quad (5)$$

avec

$$\rho_{ih} = \frac{B_{ih}^+}{\sum_{j \in (i, h)} B_{jih}^+} \leq 1. \quad (6)$$

Les π_k dépendent de caractéristiques au niveau de la strate h , de la valeur de TCS et du rapport ρ_{ih} entre la mesure de taille du NPA et la somme des mesures de taille des tournées. Ils ne dépendent pas de la tournée j .

Le quotient ρ_{ih} est égal à 1 si le NPA i a (1) une unique tournée ($C_{hi} = 1$) ou (2) plusieurs tournées ($C_{hi} > 1$) de tailles suffisantes ($B_{jih} \geq TCS/c_{ih}, \forall j \in \{1, \dots, C_{hi}\}$). Nous verrons qu'il reste proche de 1 dans l'application à l'EC2000. Ceci nous permet d'affirmer que les π_k dépendent très peu du NPA i .

Complément : Le quotient $\rho_{ih} > 0$ est inférieur ou égal à 1. En effet, tout NPA i avec C_{ih} tournées peut être classifié dans un des types suivants :

1. le NPA est de taille B_{ih} et toutes les tournées sont de taille $B_{jih} \geq TCS/c_{ih}$: on a (1) $B_{ih}^+ = B_{ih}$ car $B_{ih} = \sum_j B_{jih} \geq c_{ih}TCS/c_{ih} \geq TCS$ et (2) $\sum_j B_{jih}^+ = \sum_j B_{jih} = B_{ih}$, ce qui donne $\rho_{ih} = B_{ih}/B_{ih} = 1$.
2. le NPA est de taille $B_{ih} > TCS$ et au moins une tournée j est de taille $B_{jih} < TCS/c_{ih}$: on a (1) $B_{ih}^+ = B_{ih}$ et (2) $\sum_j B_{jih}^+ > \sum_j B_{jih} = B_{ih}$, ce qui donne $\rho_{ih} = B_{ih}/\sum_j B_{jih}^+ < 1$.
3. le NPA est de taille $B_{ih} \leq TCS$ et au moins une tournée j est de taille $B_{jih} < TCS/c_{ih}$: on a (1) $B_{ih}^+ = TCS$ et (2) $\sum_j B_{jih}^+ > c_{ih}TCS/c_{ih} > TCS$, ce qui donne $\rho_{ih} = TCS/\sum_j B_{jih}^+ < 1$.

Les probabilités d'inclusion des personnes et des ménages sont égales à celles du bâtiment dans lequel ils habitent. En effet, tous les membres de tous les ménages des bâtiments de l'échantillon des bâtiments font partie de l'échantillon de l'EC2000.

4.2.5 Un peu plus sur les mesures de taille

L'introduction des mesures de taille $B_{ih}^+ = \max(B_{ih}, TCS)$ et $B_{jih}^+ = \max(B_{jih}, TCS/c_{ih})$ a pour conséquences (1) que les très petites unités ont une probabilité relative plus grande de faire partie de l'échantillon (les mesures de taille sont toujours égales ou supérieures aux tailles réelles), et (2) que la probabilités d'inclusion des bâtiments d'une même strate sont égales - au facteur ρ_{ih} près - quelque soient la taille du NPA et de la tournée.

Si l'on prend des mesures de taille égales au nombre de bâtiments (tailles classiques), la probabilité d'inclusion des bâtiments π'_k prend la forme :

$$\begin{aligned}\pi'_k &= m_h \frac{B_{ih}}{\sum_{i \in h} B_{ih}} c_{ih} \frac{B_{jih}}{\sum_{j \in (i,h)} B_{jih}} \frac{b_{jih}}{B_{jih}} \\ &= m_h \frac{TCS}{\sum_{i \in h} B_{ih}} \rho'_{jih}\end{aligned}\quad (7)$$

avec

$$\rho'_{jih} = \min\left(\frac{c_{ih}B_{jih}}{TCS}, 1\right) \leq 1 \quad (8)$$

Les facteurs ρ_{ih} et ρ'_{jih} sont tous les deux inférieurs ou égaux à 1. Ils sont égaux à 1 lorsque le NPA a une unique tournée ou si toutes les tournées du NPA ont au moins TCS/c_{ih} bâtiments. Cependant, le facteur ρ'_{jih} , à l'inverse de ρ_{ih} , dépend des caractéristiques de la tournée j . De plus, il prend des valeurs plus éloignées de 1 que ρ_{ih} .

Illustration : Prenons un NPA i avec $B_{ih} = 410$ bâtiments répartis dans quatre tournées de tailles $B_{jih} = (150, 150, 100, 10)$. On tire $c_{ih} = 3$ tournées et la valeur de TCS est fixée à 60. La mesure de taille du NPA est $B_{ih}^+ = \max(B_{ih}, 60) = 410$ et les mesures de taille des tournées sont $B_{jih}^+ = (150, 150, 100, 20)$. On a donc $\rho_{ih} = 410/420 = 0.976$ et $\rho'_{jih} = (1, 1, 1, 0.5)$ selon la tournée j . Les probabilités d'inclusion des bâtiments du NPA i prennent des valeurs du simple au double dans le cas de l'utilisation de la taille classique. Elles sont égales et proches de 1 dans le cas de l'utilisation de la mesure de taille choisie dans la méthode d'échantillonnage.

Il est intéressant de noter qu'un tirage PPS avec les mesures de tailles choisies pour l'EC2000 peut être vu comme un tirage hybride PPS-SRS. Une augmentation de la valeur de TCS correspond en effet à un tirage de plus en plus d'unités avec probabilité égale.

Dans le cas général, si la mesure de taille est égale à la taille classique B_{ih} (ou $B_{ih}^+ = \max(B_{ih}, TCS)$ avec $TCS \leq \min(B_{ih})$), nous avons un tirage purement PPS. Si la mesure de taille est donnée par $B_{ih}^+ = \max(B_{ih}, TCS)$ avec $TCS \geq \max(B_{ih})$, nous avons un tirage SRS. Toutes les mesures de taille intermédiaires, avec $\min(B_{ih}) < TCS < \max(B_{ih})$, correspondent donc à des tirages hybrides PPS-SRS.

4.2.6 Echantillonnage quasi-self-weighting et méthode- φ

La probabilité de sélection π_k dépend principalement du plan d'échantillonnage au niveau des NPA, voir équation (5). Une attention spéciale est donc mise sur la stratification et l'allocation des NPA, tout en prenant en compte les différences de difficultés attendues pour le travail sur le terrain.

La méthode d'allocation introduite ci-dessous (méthode- φ) a pour but de choisir les probabilités d'inclusion des bâtiments avec un éloignement contrôlé d'un échantillonnage à probabilités égales (self-weighting). La méthode- φ peut être vue comme une allocation optimale selon une certaine fonction de coût constante dans des groupes de strates.

Base de l'allocation

Nous n'avons a priori aucune information sur la valeur de l'estimateur dual et sa variabilité dans la population suisse. La base de l'allocation des m NPA dans les strates $h = 1, \dots, H$ consiste donc en une allocation proportionnelle :

$$m_h = m \frac{B_h^+}{\sum_h B_h^+} = m \frac{B_h}{B^+} \quad (9)$$

où la taille $B_h^+ = \sum_{i \in h} B_{ih}^+$ de la strate h est la somme des mesures de taille des NPA qu'elle contient, et $B^+ = \sum_h B_h^+$ est la somme des tailles des strates.

Avec (5) et (9), la probabilité d'inclusion π_k du bâtiment k de la tournée j du NPA i de la strate h , tout comme celle de ses ménages et personnes, devient :

$$\pi_k = \frac{m}{B^+} TCS \left[\frac{B_{ih}^+}{\sum_{j \in (i,h)} B_{jih}^+} \right] = \frac{m}{B^+} TCS \rho_{ih} \quad (10)$$

La probabilité π_k est donc - au facteur ρ_{ih} près - indépendante des caractéristiques de la strate, du NPA et des tournées. Cette propriété de quasi-self-weighting ("quasi" car le facteur ρ_{ih} amène une petite variabilité) est intéressante pour la suite du plan d'échantillonnage car elle évite une grande variabilité des poids d'échantillonnage; variabilité dommageable pour les estimations.

Notons aussi que l'utilisation de la mesure de taille au niveau des strates (au lieu de la taille classique $B_h = \sum B_{ih}$) permet d'éviter l'apparition dans les π_k d'un facteur $\rho_h'' = B_h / \sum_{i \in h} B_{ih}^+$ dépendant de la strate.

Allocation selon la méthode- φ

On reprend l'idée de base de l'allocation proportionnelle en combinaison avec une démarche plus flexible. Cette démarche permet d'étendre l'idée de plan d'échantillonnage quasi-self weighting à des groupes de strates sur-, respectivement sous-représentés, tout en respectant la taille totale de l'échantillon des NPA.

Supposons que les strates $h = 1, \dots, H$ puissent être réparties en trois groupes : (1) strates avantageuses $G_1 = \{1, \dots, H_1\}$, (2) strates intermédiaires $G_2 = \{H_1 + 1, \dots, H_2\}$, et (3) strates difficiles $G_3 = \{H_2 + 1, \dots, H\}$. Les tailles des groupes de strates sont données par $B_{G_r} = \sum_{h \in G_r} B_h$, $r = 1, 2, 3$. La taille totale est égale à $B = \sum_r B_{G_r} = \sum_h B_h$. Les mesures de tailles correspondantes sont données par $B_{G_r}^+ = \sum_{G_r} B_h^+$ et $B^+ = \sum_r B_{G_r}^+$.

L'allocation se fait en deux étapes.

La première étape alloue les m NPA dans les groupes de strates G_1 , G_2 et G_3 comme suit :

$$\begin{aligned} m_{G_1} &= m(B_{G_1}^+ + \varphi)/B^+ \\ m_{G_2} &= mB_{G_2}^+/B^+ \\ m_{G_3} &= m(B_{G_3}^+ - \varphi)/B^+ \end{aligned} \quad (11)$$

Le paramètre φ peut prendre des valeurs dans l'intervalle $[0, B_{G_3}^+]$. Il permet un transfert de $m\varphi/B^+$ NPA du groupe difficile G_3 au profit du groupe avantageux G_1 . Ceci correspond à un éloignement de l'allocation proportionnel pour tenir compte de la difficulté de relevé dans le groupe de strates G_3 .

La deuxième étape consiste à allouer les NPA à l'intérieur des groupes de strates. Cette allocation est proportionnelle à la mesure de taille des strates. On a donc, pour la strate h dans le groupe de strate G_r :

$$m_h = m_{G_r} \frac{B_h^+}{B_{G_r}^+} \quad (12)$$

avec les éléments m_{G_r} , B_h^+ et $B_{G_r}^+$ définis plus haut.

Probabilités d'inclusion des bâtiments selon la méthode- φ

Avec l'allocation des NPA selon la méthode- φ , la probabilité d'inclusion π_k du bâtiment k du NPA i de la strate h est donnée par, voir équations (5), (11) et (12) :

$$\pi_k = \begin{cases} m \frac{TCS}{B^+} \rho_{ih} \left[1 + \frac{\varphi}{B_{G_1}^+} \right] & \text{if } k \in G_1 \\ m \frac{TCS}{B^+} \rho_{ih} & \text{if } k \in G_2 \\ m \frac{TCS}{B^+} \rho_{ih} \left[1 - \frac{\varphi}{B_{G_3}^+} \right] & \text{if } k \in G_3 \end{cases} \quad (13)$$

avec $0 < \rho_{ih} = B_{ih}^+ / \sum_{j \in (i,h)} B_{jh}^+ \leq 1$.

La probabilité d'inclusion des bâtiments, ménages et personnes ne dépend donc que de m , de TCS , du facteur ρ_{ih} et du groupe dans lequel se trouve la strate. Le paramètre φ permet de contrôler l'effet du groupe. Le seul effet non contrôlé est donc le facteur ρ_{ih} mais ce dernier est en pratique très proche de 1 (voir plus bas).

Notons que le paramètre φ s'exprime en fonction du paramètre plus intuitif $\gamma \geq 1$ défini comme le quotient entre la probabilité d'inclusion maximale (groupe G_1) et la probabilité d'inclusion minimale (groupe G_3). Si ρ_{ih} est fixé à 1, le paramètre φ est donné par :

$$\varphi = (\gamma - 1) \frac{B_{G_1}^+ B_{G_3}^+}{(B_{G_3}^+ + \gamma B_{G_1}^+)} \quad (14)$$

Si l'on désire limiter le rapport entre les probabilités maximales et minimales à $\gamma = 2$ par exemple, la valeur de φ doit inférieure ou égale à $(B_{G_1}^+ B_{G_3}^+) / (B_{G_3}^+ + 2B_{G_1}^+)$.

4.3 Plan au niveau des numéros postaux

Le plan d'échantillonnage au niveau des numéros postaux est à la base de la mise en oeuvre de la méthode d'échantillonnage. La définition de la stratification et des groupes de strates tout comme le choix des paramètres TCS et φ déterminent en effet presque entièrement le plan d'échantillonnage aux niveaux inférieurs (circonscriptions de distribution et bâtiments).

4.3.1 Base de sondage

La base de sondage est formée de la liste des $M = 3699$ NPA à 6 chiffres présents dans le REAB, après exclusion des NPA du Liechtenstein et du Tessin.

Les informations auxiliaires disponibles pour chaque NPA sont le nombre de bâtiments de différents types (variable OBJET, voir Table 4, p. 13), le nombre estimé de personnes et de ménages, les informations sur la commune majoritaire dans le NPA (variante de relevé, langue, population de la commune), et le nombre de bâtiments avec des problèmes d'adresse; voir Table 5. Les données de base sont décrites dans la section 4.1.

TAB. 5 – Définition du code d'adresse "ADR".	
ADR	Description
0	adresse complète ("nom de rue" et ("numéro de maison" ou "désignation de bâtiment"))
1	adresse pas complète mais coordonnées métriques
2	adresse pas complète et pas de coordonnées métriques

Remarquons que d'autres informations auxiliaires sont disponibles (ex. type de distribution de la Poste). Elles ne sont pas utilisées dans le plan d'échantillonnage mais sont conservées pour les besoins de la procédure d'enquête.

4.3.2 Stratification

Les NPA sont distribués dans 21 strates selon la variante de relevé du recensement, la taille de la commune et la difficulté de relevé dans l'EC2000.

Les variantes de relevé du recensement sont les variantes CLASSIC, SEMI-CLASSIC, TRANSIT et FUTURE.

Les communes majoritaires (voir section 4.1) sont classifiées en trois classes de taille : petite (moins de 500 habitants), moyenne (entre 500 et 10'000 habitants), et grande (à partir de 10'000 habitants). Notons que les variantes CLASSIC et SEMI-CLASSIC sont présentes exclusivement dans les petites et moyennes communes.

La définition de la difficulté de relevé prend en compte le nombre de bâtiments du NPA, la proportion de problèmes d'adresses et la proportion de bâtiments potentiellement vides. La classification est donnée dans la Table 6 et la définition des difficultés dans la Table 7.

TAB. 6 – Définition des types de NPA.

Taille NPA	petit	$< TCS$ bât.
	grand	$\geq TCS$ bât.
Problème adresse	mini	$<10\%$ bât. ADR=2 et $<20\%$ bât. ADR=1
	petit	$<10\%$ bât. ADR=2 et $\geq 20\%$ bât. ADR=1
	grand	$\geq 10\%$ bât. ADR2
Bâtiments vides	peu	$<10\%$ bât. OBJET>1
	beaucoup	$\geq 10\%$ bât. OBJET>1

TAB. 7 – Définition des difficultés de relevé (groupes de strates).

	Taille NPA	Prob. adresse	Bâtiments vides	Groupe
Avantageux	grand	mini	peu	G_1
Intermédiaire	grand	petit	peu	G_2
Intermédiaire	grand	mini	beaucoup	G_2
Difficile 1	grand	grand	-	G_3
Difficile 1	grand	petit	beaucoup	G_3
Difficile 2	petit	-	-	G_3

Les NPA sont classifiés selon leur taille (petit et grand). La limite est fixée à $TCS = 60$. Les petits NPA correspondent à des entreprises uniques, à des quartiers ou encore à des très petites communes. Le relevé dans les NPA de petite taille est considéré comme difficile car ces NPA ne sont pas "rentables" (contact avec un office de Poste pour peu de personnes dans l'échantillon) et impliquent un relevé fortement groupé géographiquement (tous les bâtiments participent à l'EC2000, fort effet de grappe). Dans le cas de petites communes, la participation à l'enquête correspond de plus à réaliser un deuxième recensement de la commune avec les problèmes d'acceptance que cela risque de produire.

Les problèmes d'adresse sont définis selon les proportions de bâtiments avec code d'adresse ADR=1 et ADR=2. Dès que la proportion de code d'adresse ADR=2 dépasse 10%, le relevé dans le NPA est considéré comme difficile (identification difficile du bâtiment par le facteur puis l'interviewer). Les problèmes d'adresse ADR=1 sont moins graves car des cartes permettent d'identifier les bâtiments, mais non négligeables lorsqu'ils atteignent une proportion importante de bâtiments.

La proportion de bâtiments avec OBJET>1 (Table 4, p. 13) est aussi prise en compte dans la définition des difficultés de relevé. Un NPA avec beaucoup de bâtiments OBJET>1 a

potentiellement beaucoup de bâtiments sans habitants. La participation à l'enquête n'est donc pas très rentable et le relevé peut être difficile dans le cas de zones industrielles.

TAB. 8 – Définition des strates 1 à 21 (variante x difficulté x taille de commune).

		CLASSIC	SEMI-CLASSIC	TRANSIT	FUTURE
Avant.	Petite	1	8	15	-
	Moyenne	2	9	15	21
	Grande	-	-	16	21
Inter	Petite	3	10	17	-
	Moyenne	4	11	17	21
	Grande	-	-	18	21
Diff.1	Petite	5	12	19	21
	Moyenne	6	13	19	21
	Grande	-	-	19	-
Diff.2		7	14	20	-

La stratification choisie donne la possibilité de contrôler les tailles d'échantillon au niveau des variantes de relevé et des tailles de communes tout en prenant en compte les difficultés attendues dans le relevé, voir Table 8. Le nombre de NPA par strate varie entre 26 (strate 21) et 1126 (strate 15). On note que les NPA dans les villes sont principalement dans le groupe de strates avantageuses. Cela est bénéfique aux estimations vu que ces NPA ont potentiellement plus de problèmes de couverture que les NPA dans les régions rurales.

4.3.3 Allocation

Le nombre de NPA à distribuer dans les 21 strates est fixé à $m = 285$. Ces 285 NPA sont répartis entre les 4 variantes de relevé puis entre les strates des différentes variantes. L'allocation est choisie sur la base de simulations du nombre espéré de bâtiments, de ménages et de personnes dans l'échantillon. Le paramètre TCS est constant pour toutes les variantes. La méthode- φ est appliquée indépendamment dans chaque variante, ce qui conduit à une valeur de φ par variante.

Choix de TCS

Le nombre de bâtiments tirés dans chaque NPA est fixé à $TCS = 60$. Elle est choisie de façon à respecter les deux contraintes suivantes : (1) total de 26'000 ménages dans les 285 NPA et (2) 11'800 personnes dans la variante CLASSIC, en tenant compte des valeurs reportées dans la Table 9. Le but lié à la variante CLASSIC est très contraignant car seul 3.8% de la population est recensée selon cette méthode et le nombre d'habitants par bâtiment est petit en comparaison avec les autres variantes².

²Dans la variante CLASSIC, 11'800 personnes sont réparties dans 4400 ménages et 4684 bâtiments. Si les 26'000-4'400=21'600 ménages restant sont répartis dans les bâtiments avec une moyenne de 1.8 ménages/bâtiment, nous arrivons à 12'000 bâtiments hors variante CLASSIC. Le total des 4684+12'000=16'684 bâtiments est alors distribué dans les 285 NPA avec une moyenne de 58.5 bâtiments par NPA.

TAB. 9 – Nombre M de NPA et B de bâtiments par variante. Estimations P et H du nombre de personnes et de ménages et diverses densités H/B , P/B et P/H .

Variante	M	B	P	H	H/B	P/B	P/H
CLASSIC	737	102765	258155	96446	0.94	2.51	2.68
SEMI-CLASSIC	266	62545	170935	64202	1.03	2.73	2.66
TRANSIT	2670	1426452	6273591	2648630	1.86	4.40	2.37
FUTURE	26	15433	58410	24066	1.56	3.78	2.43
Total	3699	1607195	6761091	2833344	1.76	4.21	2.39

Choix de φ et de la taille de l'échantillon dans chaque variante

La méthode- φ est appliquée dans les variantes CLASSIC, SEMI-CLASSIC et TRANSIT, avec arrondis des valeurs à l'entier le plus proche et choix de minimum 2 NPA par strate. Le facteur φ est fixé à 0 dans la variante FUTURE (une strate, pas de groupes de difficultés).

Dans chaque variante donnée, on fait varier le nombre m_{var} de NPA et le paramètre φ et on estime le nombre de bâtiments, de ménages et de personnes dans l'échantillon, ainsi que le rapport γ entre le poids maximal et le poids minimal au niveau des bâtiments. Le nombre de bâtiments est estimé sur la base d'un tirage de $\min(B_{ih}, TCS)$ bâtiments dans le NPA i de la strate h . Les nombres de ménages et de personnes sont basés sur le nombre moyen de ménages et de personnes par bâtiment dans le NPA.

Le choix de m_{var} a bien entendu une grande influence sur les tailles estimées. On note cependant qu'une augmentation de φ permet en général d'augmenter très légèrement le nombre espéré de ménages car les strates avantageuses - sur-représentées - ont souvent une plus forte densité de population par bâtiment que les strates difficiles (voir exemple dans la figure 2).

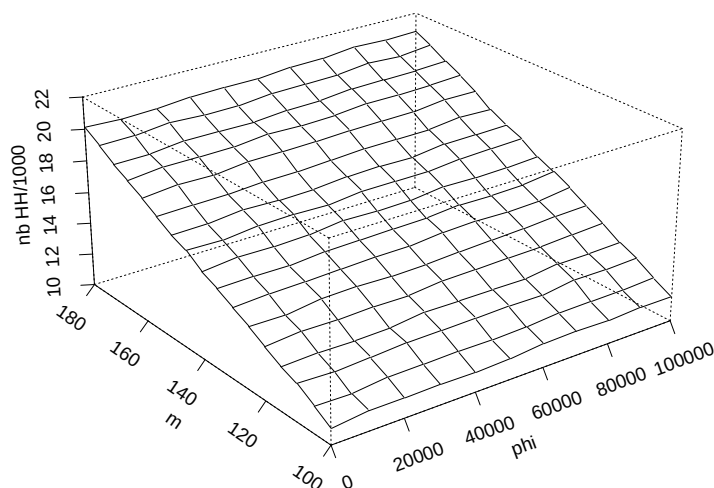


FIG. 2 – Nombre estimé de ménages dans la variante TRANSIT en fonction du choix de φ (phi) et de m_{var} (m).

La figure 3 illustre l'effet du paramètre φ sur les poids d'échantillonnage des trois groupes

de strates de la variante TRANSIT ; voir équation (13). On note que les poids du groupe G_3 sont plus sensibles à la valeur de φ que ceux du groupe G_1 car $B_{G_3}^+$ est inférieur à $B_{G_1}^+$. Notons encore que l'effet du facteur ρ_{ih} ne peut pas être considéré ici car les informations sur les tournées ne sont pas encore connues. La valeur de ρ_{ih} est fixée à 1.

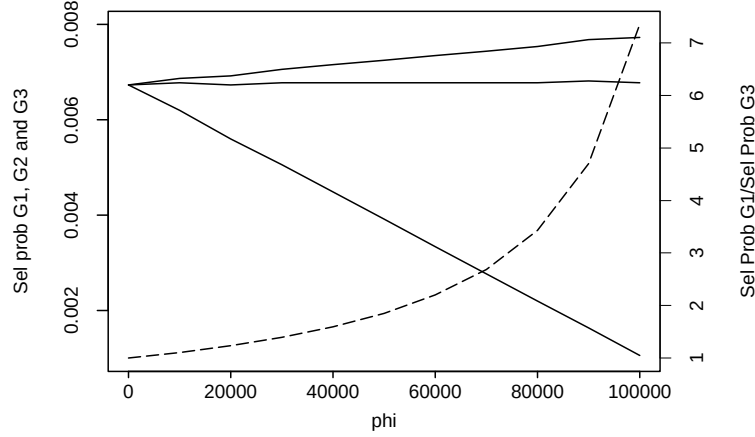


FIG. 3 – Probabilités d’inclusion π_k en fonction du paramètre φ (échelle de gauche et lignes continues : G_1 en haut, G_2 au milieu et G_3 en bas) et facteur $\gamma = \text{Sel Prob G1/Sel Prob G3}$ (échelle de droite et ligne traitillée). Allocation de 160 NPA, variante TRANSIT et $TCS = 60$.

Les choix finaux sont notés dans la Table 10. On note la différence entre le rapport théorique γ et le rapport réel $\tilde{\gamma} = \max w_k / \min w_k$. Ces différences sont dues aux arrondis au niveau des allocations dans les strates et surtout à la limite inférieure de 2 NPA par strate.

TAB. 10 – Résumé de l’allocation selon la méthode- φ dans les variantes de relevé. Nombre M_{var} de NPA et mesures de taille $B_{G_1}^+, B_{G_2}^+, B_{G_3}^+$ et B^+ , choix de m_{var} , de φ et du facteur théorique γ . Poids minimum $\min w_k$ et maximum $\max w_k$ des bâtiments (arrondis) et rapport réel $\tilde{\gamma} = \max w_k / \min w_k$.

Variante	M_{var}	$B_{G_1}^+$	$B_{G_2}^+$	$B_{G_3}^+$	B^+	m_{var}
CLASSIC	737	37402	43277	27169	107848	80
SEMI-CLASSIC	266	28919	20104	14230	63253	37
TRANSIT	2670	710162	605973	118381	1434516	166
FUTURE	26	-	-	-	15433	4

Variante	φ	γ	$\min w_k$	$\max w_k$	$\tilde{\gamma}$
CLASSIC	10'000	2.01	17	39	2.29
SEMI-CLASSIC	4'000	1.58	16	37	2.31
TRANSIT	50'000	1.85	118	248	2.10
FUTURE	-	-	64	64	1

Résultats de l'allocation

Les valeurs de TCS , m_{var} et φ déterminent l'allocation des NPA dans les différentes strates. Avant de définir l'allocation finale des NPA, on contrôle que les probabilités d'inclusion de tous les NPA sont inférieures ou égales à 1 pour le tirage PPS. Cette contrainte n'est en effet pas respectée dans une strate h dès qu'un de ses NPA i a la propriété : $B_{ih}^+ > 1/m_h \sum_{i \in h} B_{ih}^+$.

Des probabilités d'inclusion supérieures à 1 sont observées dans les strates 2 et 4 (variante CLASSIC, moyennes communes de difficultés avantageuses et intermédiaires) et nécessitent donc un éloignement de la méthode- φ :

- strate 2 : diminution du nombre alloué de 18 à 16 NPA. Les probabilités d'inclusion des bâtiments de la strate 2 restent égales mais s'éloignent légèrement de celles de la strate du même groupe (strate 1) ;
- strate 4 : partage de la strate en deux parties et éloignement de l'allocation proportionnelle (1) strate 41 : choix de 15 NPA sur les 55 NPA avec moins de 500 bâtiments (2) strate 42 : choix de 3 NPA sur les 8 NPA avec au moins 500 bâtiments.

L'allocation finale comporte 78 NPA CLASSIC, 37 NPA SEMI-CLASSIC, 166 NPA TRANSIT et 4 NPA FUTURE. Le nombre de NPA varie entre 2 et 60 par strate (voir détails dans la Table 11). Une réserve n'est pas prévue car un taux de réponse à 100% est attendu au niveau de la participation des offices de Poste (travail rémunéré et soutien de la Poste).

TAB. 11 – Allocation des NPA dans les 22 strates (stradap), avec M_h le nombre de NPA total, $\sum B_{ih}^+$ la somme des mesures de taille et m_h le nombre de NPA alloués.

Stradap	M_h	$\sum B_{ih}^+$	m_h	Stradap	M_h	$\sum B_{ih}^+$	m_h
1	174	17734	17	13	25	8874	4
2	78	19668	16	14	33	1980	2
3	145	18631	14	15	1126	490689	60
5	56	7752	4	16	165	219473	27
6	28	7837	4	17	778	401411	46
7	193	11580	5	18	138	204562	24
8	38	4762	3	19	226	104161	7
9	73	24157	15	20	237	14220	2
10	28	4299	2	21	26	15433	4
11	43	15805	9	41	55	15165	15
12	26	3376	2	42	8	9481	3
Total				Total	3699	1621050	285

4.3.4 Tirage

Le tirage aléatoire PPS sans remise des NPA est réalisé avec la procédure SURVEYSELECT de SAS (method=pps³). La mesure de taille est la variable B_{ih}^+ décrite dans la section 4.2.

Les 285 NPA de l'échantillon sont répartis dans 23 des 25 cantons et demi-cantons du plan NORD ; voir la liste dans l'Annexe D. Seuls les demi-cantons de Appenzell Rhodes intérieurs (AI) et Obwald (OW) ne sont pas touchés. Les 285 NPA sont répartis dans les langues officielles comme suit : 195 allemand, 84 français et 6 italien (GR italophone). La répartition dans les classes de taille des communes est la suivante : 50 dans les petites communes, 182 dans les moyennes communes et 53 dans les grandes communes.

Les tailles espérées aux niveaux inférieur de l'échantillonnage sont respectivement : 16'792 bâtiments, 25'986 ménages et 62'319 personnes. Les deux dernières valeurs étant probablement sous-estimées (voir section 4.1).

Les estimations du nombre de personne dans les différents groupes d'intérêt nous permet d'espérer atteindre la taille minimale de 11'800 personnes dans les groupes suivants : variante CLASSIC (en comptant sur une sous-estimation des tailles), variante TRANSIT, moyennes communes (total et TRANSIT), grandes communes (global et TRANSIT), communes francophones (total et TRANSIT) et communes alémaniques (total et TRANSIT) ; voir Table 12.

TAB. 12 – Nombre estimé de personnes atteintes dans le plan NORD, selon la variante, la langue et la taille de la commune.

Variante	Langue			Taille commune			Total
	D	F	I	Pet.	Moy.	Grand	
CLASSIC	6912	3800	462	4875	6298		11173
SEMI-CLASSIC	3145	2646	75	999	4867		5866
TRANSIT	31275	13051		224	23260	20842	44326
FUTURE	953				371	582	953
Total	42285	19497	536	6098	34796	21425	62319

4.4 Plan au niveau des tournées

Le travail principal pour le plan au niveau des tournées se trouve dans la création de la liste des tournées et dans la détermination de la liste des bâtiments qu'elles contiennent.

³La procédure de tirage PPS de SAS utilise l'algorithme Hanurav-Vijayan, introduit par Hanurav (1967) et généralisé par Vijayan (1968). Cet algorithme permet de calculer des probabilités d'inclusion du premier ordre, ainsi que celles du deuxième ordre assurant la non-négativité et la stabilité de l'estimateur de la variance selon Sen-Yates-Grundy.

4.4.1 Base de sondage

La construction de la base de sondage au niveau des circonscriptions de distribution (ci-dessous : tournées) commence par la recherche de la tournée et du numéro postal d'acheminement des 247'972 bâtiments du REAB se trouvant dans les 285 NPA de l'échantillon $s(NPA)$, voir la procédure dans la section 4.1).

On note que les tournées n'ont pas été trouvées pour l'ensemble des bâtiments de 128 NPA (régions sans adresses postales ou tournées pas définies dans le Botenfile). Dans les 157 NPA restant, on compte entre 1 et 41 tournées, comportant de 1 à 515 bâtiments. La proportion de bâtiments pour lesquels on a trouvé une tournée varie aussi fortement, avec des valeurs entre 13% et 100% selon le NPA.

La base de sondage des tournées est créée à partir de la liste des circonscriptions assignées aux 247'972 bâtiments du REAB. On crée une unique tournée fictive dans (1) les 128 NPA sans tournées, (2) les 17 NPA avec seulement 1 ou 2 tournées et (3) les 12 NPA avec ≥ 3 tournées mais moins de 20% des bâtiments avec tournées. Dans la construction de la base, on ignore l'existence des toutes petites tournées (avec moins de 10 bâtiments). La base de sondage est formée de $C = 1489$ tournées d'échantillonnage (réelles ou fictives).

Chaque bâtiment de la liste est assigné à une tournée d'échantillonnage. Dans 95.6% des cas, les bâtiments font clairement partie d'une tournée de la liste. Dans les 4.4% restant, ils sont distribués dans les tournées de leur NPA selon une procédure aléatoire et proportionnellement à la taille des tournées. Une partie des bâtiments ont donc un numéro de tournée d'échantillonnage ne correspondant pas à la tournée auxquels ils sont assignés dans la réalité (que le vrai numéro soit connu ou inconnu). Ce numéro est utilisé uniquement pour l'échantillonnage.

La taille B_{jih} de la tournée d'échantillonnage j du NPA i de la strate h est égale au nombre de bâtiments réellement ou fictivement assignés à cette tournée. De par la construction de la base de sondage, le nombre C_{ih} de tournées d'échantillonnage dans le NPA i de la strate h est égal à 1 ou ≥ 3 .

Notons que seuls 26 NPA sur les 285 NPA de l'échantillon ont un facteur $\rho_{ih} < 1$ avec une valeur minimale atteinte par le NPA 461300 de la strate 17 ($\rho_{ih} = 291/299 = 0.97$). Le facteur ρ_{ih} a donc bien un effet négligeable sur la variabilité de la probabilité d'inclusion π_k au niveau des bâtiments; voir équation (13).

4.4.2 Allocation et tirage

On alloue $c_{ih} = \min(C_{ih}, 3)$ tournées d'échantillonnage dans le NPA i de la strate h . Le nombre de tournées tirées dans chaque NPA est égal à 1 dans 157 NPA et égal à 3 dans 128 NPA.

Le tirage aléatoire PPS sans remise des tournées est effectué avec la procédure PROC SURVEYSELECT de SAS (method=pps). La mesure de taille est définie par $B_{jih}^+ = \max(B_{jih}, TCS/c_{ih})$ (voir section 4.2), c'est-à-dire :

$$B_{jih}^+ = \begin{cases} \max(B_{jih}, 60) & \text{si } c_{ih} = 1 \\ \max(B_{jih}, 20) & \text{si } c_{ih} = 3 \end{cases} \quad (15)$$

Un éloignement du tirage PPS doit être effectué dans deux groupes de NPA. Le premier groupe comporte 16 NPA avec $C_{ih} = c_{ih} = 3$. Les probabilités d'inclusion sont toutes trois fixées à 1 au lieu de prendre les valeurs $\pi_{j|ih} = c_{ih}B_{jih}^+ / \sum_j B_{jih}^+$ prévues dans la méthode d'échantillonnage; voir l'équation (3). Le deuxième groupe de 14 NPA contient $C_{ih} > 3$ tournées mais certaines d'entre elles ont des $\pi_{j|ih}$ supérieures à 1 ($c_{ih} \max_j (B_{jih}^+) / \sum_j B_{jih}^+ > 1$). Dans ces NPA, on prend les δ_{ih} tournées concernées avec probabilité 1 et on tire les $3 - \delta_{ih}$ tournées restantes selon la procédure PPS (processus itératif).

L'éloignement du tirage PPS, et donc du plan décrit dans la section 4.2, est relativement important dans le premier cas particulier ($C_{ih} = c_{ih} = 3$). Une procédure spéciale est donc appliquée au niveau du choix des bâtiments, voir section 4.5, pour réduire l'éloignement de la méthode prévue et limiter ainsi la variabilité des π_k .

L'échantillon des tournées $s_{(tour)}$ est de taille $c = 541$.

Un ensemble de 157 tournées sont fictives, 128 parce qu'elles se trouvent dans des NPA sans tournées dans le Botenfile et 29 parce qu'elles se trouvent dans des NPA avec moins de trois tournées ou dans lesquels moins de 80% des bâtiments pouvaient être associés à une tournée réelle. Les tournées de $s_{(tour)}$ comportent entre 4 et 2353 bâtiments avec une moyenne de 228 bâtiments.

La sélection des petites tournées avec une probabilité plus faible que les grandes permet d'éviter une surcharge de travail des facteurs à temps partiel ou ayant des tournées avec des bâtiments très dispersés.

4.5 Plan au niveau des bâtiments

Dans le plan d'échantillonnage au niveau des bâtiment, seule l'allocation est un peu plus complexe que prévu dans la section 4.2. Une adaptation est en effet effectuée afin de corriger une partie de l'éloignement du tirage PPS au niveau des tournées.

4.5.1 Base de sondage

La base de sondage des bâtiments est formée de la liste des bâtiments du REAB dans les $c = 541$ tournées échantillonnées. Elle comporte $B = 123'276$ bâtiments.

4.5.2 Allocation

L'allocation est faite en deux partie.

Dans les 493 tournées ne se trouvant pas dans les 16 NPA concernés par le tirage de $c_{ih} = C_{ih} = 3$ tournées, on alloue $b_{jih} = \min(B_{jih}, TCS/c_{ih})$ bâtiments dans la tournée j du NPA i de la strate h , c'est-à-dire :

$$b_{jih} = \begin{cases} \min(B_{jih}, 60) & \text{si } c_{ih} = 1 \\ \min(B_{jih}, 20) & \text{si } c_{ih} = 3 \end{cases} \quad (16)$$

Dans les 48 tournées des 16 NPA dans lesquels on a tiré $c_{ih} = C_{ih} = 3$ tournées, on applique la procédure spéciale suivante :

1. allocation de $TCS = 60$ bâtiments proportionnellement à la taille des tournées $b_{jih}^* = TCS \cdot B_{jih} / \sum_j B_{jih}$, $j = 1, 2, 3$.
2. limitation à 25 bâtiment par tournée $b_{jih} = \min(b_{jih}^*, 25)$
3. allocation des bâtiments restants dans les tournées avec $b_{jih}^* < 25$ bâtiments.

Cette procédure permet de limiter le nombre de bâtiments dans les grandes tournées et de ne pas prendre trop de petites tournées entières. Si les b_{jih}^* sont tous inférieurs à 25, les probabilités d'inclusion des bâtiments π_k sont à nouveau indépendantes de la tournée et dépendantes du NPA uniquement par un facteur $\rho_{ih}^* = B_{ih}^+ / B_{ih}$:

$$\pi_k = m_h \frac{B_{ih}^+}{\sum_{i \in h} B_{ih}^+} \frac{TCS \cdot B_{jih} / \sum_j B_{jih}}{B_{jih}} = m_h \frac{TCS}{\sum_{i \in h} B_{ih}^+} \rho_{ih}^* \quad (17)$$

Si certains b_{jih}^* sont supérieurs à 25, les probabilités d'inclusion dépendent de la tournée. Elles sont cependant moins variables que dans le cas d'un tirage de $b_{jih} = \min(B_{jih}, TCS/c_{ih})$ bâtiments.

4.5.3 Tirage

Le tirage aléatoire simple sans remise des bâtiments est réalisé avec la procédure PROC SURVEYSELECT de SAS (method=SRS⁴).

L'échantillon final des bâtiment $s_{(bat)}$ contient $b = 16'792$ bâtiments.

Environ 50% des bâtiments sont assignés à une tournée réelle mais la plupart des autres bâtiments se trouvent dans les NPA sans adresses postales ou sans tournées officielles. Seuls 39 bâtiments ont un numéro postal d'acheminement différent du numéro postal de l'adresse (trois NPA concernés).

Le ratissage large au niveau des bâtiments dans le REAB devrait permettre une bonne couverture de la population mais implique certaines difficultés dans le travail de l'enquête sur le terrain. On note notamment que 14% des bâtiments ont un code OBJET supérieur à 1 et un peu plus de 2% sont annoncés comme inconnus (OBJET=8) ou démolis (OBJET=9), voir définitions dans la Table 4, p. 13.

De plus, certains bâtiments avec codes d'adresse ADR=0 ("adresse postale complète", voir définitions dans la Table 5, p.20) se sont en fait avérés avoir des adresses postales incomplètes (numéro d'assurance incendie ou un numéro de parcelle dans le champs "numéro de maison"). Ce problème a été détecté après le tirage des bâtiments et touche principalement les NPA CLASSIC. Le plan d'échantillonnage au niveau des NPA ne prend donc que partiellement en compte les problèmes d'adresses.

Les codes d'adresse ADR=0 sont revus pour les 2208 bâtiments de l'échantillon (13%) se trouvant dans les numéros postaux dont la commune majoritaire n'a pas de registre

⁴La procédure de tirage SRS de SAS utilise le "Floyd's ordered hash table algorithm", rapide et efficace dans le cas de gros jeux de données; référence dans Bentley and Floyd (1987) et Bentley and Knuth (1986).

de rues (liste de 595 communes fournie par la section RAUM) et pour les bâtiments des numéros postaux avec une numérotation partielle des rues (critère : au moins 5 bâtiments de l'échantillon ont des numéros de maison égaux à un numéro d'assurance ou de parcelle).

Une nouvelle variable "ADRCOR" est créée pour identifier les vrais problèmes d'adresse. Le nombre de bâtiments nécessitant des cartes (ADRCOR=1) passe de 6.7% à 16.3% de l'échantillon et le nombre des bâtiments sans adresse ni coordonnées (ADRCOR=2) passe de 2.4% à 6.0%. Sur les 16'792 bâtiments de l'échantillon, 2722 bâtiments nécessitent des cartes (dans 129 NPA) et 1009 bâtiments n'ont ni adresse postale ni coordonnées et doivent être identifiés sur la base d'autres informations (seuls 12 n'ont pas de nom de personne de contact).

4.6 Sous-échantillonnage des bâtiments

La liste des ménages dans les bâtiments de $s_{(bat)}$ est construite selon la procédure d'enquête décrite dans la section 2 : listes provisoires (voir Annexe C), contrôles par les facteurs, et saisie des corrections.

Les listes des ménages ne sont pas disponibles dans les NPA de Founex et Mayens-de-Riddes. Founex a été exclu de l'échantillon avant l'envoi aux offices de Poste car la commune avait participé à la répétition générale du recensement en automne 1999 (GP99) et à l'enquête pilote de l'EC2000 au printemps 2000. La Poste des Mayens-de-Riddes n'a pas fait les contrôles des listes de ménages. Ces valeurs manquantes ("non-réponse") au niveau des NPA seront prises en compte dans la pondération.

Plusieurs doubles ont été détectés dans les listes de ménages (même numéro de téléphone et même nom). Dans chaque cas, un choix a été fait entre les deux (ou trois) entrées présentes. Notons que le choix peut provoquer des erreurs de localisation dans le cas de doubles dans deux bâtiments différents. La procédure d'appariement avec le recensement devrait cependant permettre de retrouver les personnes même si elles ont été énumérées dans un bâtiment proche de celui où elles ont été recensées.

Parmi les 16'672 bâtiments restants - échantillon $s_{(bat,resp)}$ - 14'050 sont habités par un total de 30'224 ménages. La répartition des ménages dans les variantes, langues et tailles de communes correspond aux proportions estimées dans la section 4.3.4, mais les tailles sont 10 à 12% supérieures.

Un sous-échantillon de bâtiments $s_{(bat,sub)}$ est tiré parmi les bâtiments de $s_{(bat,resp)}$ en suivant les principes suivants :

- réduction globale de environ 12% des ménages ;
- réduction plus forte dans la variante TRANSIT que dans la variante CLASSIC ;
- réduction plus importante dans les NPA comportant une grande proportion de logements habités temporairement et sans téléphone (NPA touristiques avec maisons de vacances).

On choisit une méthode de sous-échantillonnage simple afin d'éviter d'ajouter à la complexité de l'enquête. Aucune correction n'est nécessaire au niveau de la proportion de ménages sans téléphone car cette dernière, égale à 11.7%, est inférieure à la valeur de 15% fixée dans le contrat avec IHA-GfM.

4.6.1 Base de sondage et stratification

La base de sondage des bâtiments est formée des $B_{sub} = 16'672$ bâtiments de $s_{(bat,resp)}$.

Une sous-stratification est appliquée dans 5 NPA ayant une unique tournée d'échantillonnage. La sous-stratification consiste à séparer les bâtiments avec un unique ménage, sans téléphone et noté comme "habité temporairement" ("chalets de vacances") du reste des bâtiments du NPA, voir la Table 13. Elle a pour but de limiter le nombre de ménages à contacter dans des "chalets de vacances".

TAB. 13 – Sous-stratification strasub dans 5 NPA touristiques, avec B_ℓ le nombre de bâtiments, H_ℓ le nombre de ménages, et la strate stradap au niveau des NPA.

Strasub	B_ℓ	H_ℓ	Stradap	Lieu
1923000001	37	39	11	Les Marécottes (pas chalets)
1923000002	23	23	11	Les Marécottes (chalets vacances)
1943000001	47	55	41	Praz-de-Fort (pas chalets)
1943000002	13	13	41	Praz-de-Fort (chalets vacances)
6541000001	48	19	1	Sta Maria (pas chalets)
6541000002	12	12	1	Sta Maria (chalets vacances)
6543000001	40	33	1	Arvigno (pas chalets)
6543000002	20	20	1	Arvigno (chalets vacances)
7423000001	31	39	1	Sarn (pas chalets)
7423000002	29	29	1	Sarn (chalets vacances)

On définit également le groupe des 9 NPA touristiques comme suit : Marécottes (192300), Praz-de-Fort (194300), Sta Maria (654100) Arvigno (654300), Sarn (742300), Grächen (392500), Glurigen (398108), Arosa (705000) et Schmitten (Albula) (749300). Ces NPA ont au moins 14 bâtiments avec des logements occupés temporairement, au moins 10 ménages temporaires sans téléphone et au moins 18% de logements occupés temporairement sans téléphone. Ils ont tous une unique tournée d'échantillonnage.

La construction en plusieurs niveaux et la sous-stratification dans les 5 NPA spéciaux conduisent à un ensemble de 542 strates dans lesquelles sont répartis les $B_{sub} = 16'672$ bâtiments. La strate $\ell = \{1, \dots, 542\}$ est de taille $B_{sub, \ell j i h}$ dans la tournée j du NPA i de la strate h .

Notons que nous sommes dans une situation d'échantillonnage en deux phase (two-phase sampling) car des informations recueillies sur les bâtiments de $s_{(bat,resp)}$ sont utilisées pour le plan d'échantillonnage. Ces informations permettent notamment de sous-stratifier les NPA touristiques en "chalets de vacances" ou "pas chalets". Nous sommes dans un cas d'un tirage aléatoire simple parmi des bâtiments et les probabilités d'inclusion globale des bâtiments sous-échantillonnés sont égales au produit entre la probabilité d'être dans l'échantillon et la probabilité d'être choisi dans le sous-échantillon. Le calcul de la variance est cependant plus complexe, voir par exemple Särndal et al. (1992). Nous en reparlerons dans la conclusion.

4.6.2 Allocation

L'allocation des bâtiments dans la strate ℓ de la tournée j du NPA i de la strate h est définie comme suit (" $\lfloor \cdot \rfloor$ " pour arrondi à l'entier inférieur) :

$$b_{\ell j i h} = \lfloor \beta B_{sub, \ell j i h} \rfloor$$

avec le facteur β défini par :

$$\beta = \begin{cases} 0.5 & \text{si strate "chalet vacances"} \\ 0.8 & \text{sinon et si strate "touristique"} \\ 0.88 & \text{sinon et si TRANSIT dans les strates des groupes } G_1 \text{ et } G_2 \\ 0.9 & \text{sinon et si SEMI-CLASSIC ou FUTURE} \\ 0.95 & \text{sinon} \end{cases} \quad (18)$$

Le plus fort sous-échantillonnage est donc réalisé dans les strates "chalets de vacances", puis viennent les NPA touristiques, ensuite TRANSIT, SEMI-CLASSIC et FUTURE. La majorité des NPA CLASSIC sont très peu sous-échantillonnés.

4.6.3 Tirage

Le tirage aléatoire simple sans remise des bâtiments est réalisé avec la procédure PROC SURVEYSELECT de SAS (method=SRS).

L'échantillon final de $b_{sub} = 14'697$ bâtiments comporte un total de 26'369 ménages, dont 3021 sans téléphone (11.5%). Les tailles en terme de ménages et de population (Tables 14 et 15) confirment que les buts de l'échantillonnage sont atteints. Le nombre de personnes est suffisant (*i.e.* limite de 11'800) pour la variante CLASSIC et la variante TRANSIT (total, moyennes et grandes communes, parties alémanique et romande).

TAB. 14 – Nombre de ménages dans les bâtiments de l'EC2000, selon la variante, la langue et la taille de la commune.

Variante	Langue			Taille commune			Total
	D	F	I	Pet.	Moy.	Grand	
CLASSIC	3075	1429	228	2229	2503		4732
SEMI-CLASSIC	1326	1086	39	445	2006		2451
TRANSIT	13603	5189		116	9332	9344	18792
FUTURE	394				124	270	393
Total	18398	7704	267	2790	13965	9614	26369

Le sous-échantillonnage a un effet non négligeable sur les poids (Table 16, voir par ex. les strate 1, 11 et 41). Les poids de la variante CLASSIC ont des valeurs entre 15.8 et 55.4 (facteur 3.5 entre les extrêmes), ceux de la variante SEMI-CLASSIC entre 4.4 et 61.2 (facteur 13.9), et ceux de la variante TRANSIT entre 45.0 et 325.8 (facteur 7.2). Les poids sont fixes pour la variante FUTURE.

TAB. 15 – Nombre estimé de personnes dans les bâtiments de l’EC2000, selon la variante, la langue et la taille de la commune.

Variante	Langue			Taille commune			Total
	D	F	I	Pet.	Moy.	Grand	
CLASSIC	8138	3835	583	5844	6713		12556
SEMI-CLASSIC	3597	2721	98	1165	5250		6416
TRANSIT	31274	11996		302	23245	19723	43270
FUTURE	924				340	584	924
Total	43933	18552	681	7312	35548	20307	63166

TAB. 16 – Distribution des poids d’échantillonnage finaux des ménages par strate stradap (percentiles P_x , $x = \{0, 25, 50, 75, 100\}$) et nombre H_h de ménages dans l’échantillon NORD.

Stradap	H_h	P_0	P_{25}	P_{50}	P_{75}	P_{100}
1	1148	18.3	18.3	18.3	18.3	36.0
2	1244	21.6	21.6	21.6	21.6	25.6
3	681	23.3	23.3	23.3	23.3	23.3
5	255	34.0	34.0	34.0	34.0	40.4
6	177	34.4	34.4	34.4	34.4	34.4
7	145	40.7	41.4	41.6	42.5	42.7
8	162	29.4	29.4	29.4	29.4	29.4
9	1220	4.4	29.8	29.8	29.8	36.5
10	133	39.8	39.8	39.8	39.8	39.8
11	571	31.3	32.5	32.5	32.5	61.2
12	150	31.3	31.3	31.3	31.3	31.3
13	159	41.1	41.1	41.1	41.1	41.1
14	56	18.3	18.3	18.3	18.3	19.3
15	5412	76.0	157.3	159.7	160.4	235.5
16	5053	156.3	159.4	159.4	159.8	161.5
17	3430	45.0	167.8	171.1	171.1	325.8
18	4291	163.9	167.1	167.1	167.1	168.4
19	568	275.6	275.6	275.6	275.6	310.0
20	38	134.8	134.8	134.8	134.8	158.0
21	394	71.4	71.4	71.4	71.5	71.5
41	871	15.8	17.7	17.7	17.7	36.5
42	211	55.4	55.4	55.4	55.4	55.4
Total	26369	4.4	39.8	159.4	167.1	325.8

La variabilité dans les poids d'une même variante est due au choix de $\varphi > 0$, à l'éloignement de la méthode de la section 4.2 dans le cas de tirages PPS inapplicables pour quelques NPA et tournées, et à la sous-stratification dans le sous-échantillonnage des bâtiments.

5 Plan d'échantillonnage TICINO

Le plan d'échantillonnage TICINO est construit de façon à atteindre environ 1'000 ménages dans 20 communes du Tessin. Les contraintes sur les probabilités d'inclusion sont moins strictes que pour le plan NORD mais on évitera une trop grande variabilité des poids.

5.1 Données

Le plan d'échantillonnage utilise plusieurs sources de données pour définir les bases de sondages, la stratification au niveau des communes et les allocations aux deux niveaux de l'échantillonnage (communes et bâtiments).

5.1.1 Liste des bâtiments de la Stima

Le registre des bâtiments du Tessin est créé sur la base d'une statistique fiscale ("Dichiarazione per la stima ufficiale della sostanza immobiliare", ci-dessous Stima) avec un relevé en 1998 pour un premier groupe de 175 communes et en 1999 pour le reste des 70 communes. Le relevé portait sur tous les fonds immobiliers (données du cadastre, types de constructions et logements) mais ne contient pas l'adresse postale du fond. Le lien avec les données du géocodage de l'OFS (bâtiments de 1990) permet d'assigner des coordonnées et des adresses postales à un sous-ensemble de ces bâtiments (lien par la parcelle effectué par l'Ufficio di statistica du Tessin).

Au moment du tirage des unités primaires (communes), nous disposons de données *provisaires* pour toutes les communes du Tessin dans l'état du 27 octobre 2000. Ces données ne contiennent pas tous les types de bâtiments. Il manque par exemple les : "struttura scolastica", "ospedale", "clinica", et "casa per anziani". Au moment du tirage des bâtiments, nous avons un jeu de données dans l'état du 12 décembre 2000. Il est plus complet car il contient tous les types mais il est restreint aux communes de l'échantillon (données *finale*s). Les nouvelles constructions (après 1998, resp. 1999) ne sont pas disponibles pour toutes les communes et les bâtiments fédéraux sont exclus (ex. douanes).

Les données au niveau des bâtiments du Tessin ne sont pas consolidées au moment de leur utilisation dans le cadre de l'EC2000. La couverture de la population cible sera évaluée par comparaison avec les données des bâtiments du recensement 2000. Elle est probablement moins bonne que dans le plan NORD.

5.1.2 Liste des communes, ménages et estimation de la population

Le canton du Tessin comporte 245 communes politiques. Dans 3 communes politiques, les bâtiments sont partagés en secteurs cadastraux (Lugano : 3 secteurs, Chiasso : 2 secteurs, et Onsernone : 3 secteurs). Les 250 communes politiques et secteurs forment l'ensemble des unités d'échantillonnage au premier niveau. Pour simplifier la présentation, on parlera ci-dessous de communes (que cela soit une commune politique ou un secteur).

Les estimations de la population résidante sont basées sur les données de la statistique de l'état annuel de la population (ESPOP ; valeurs état fin 1998). Le nombre de personnes par secteur cadastral est estimé à partir de la population communale et du nombre moyen de personnes par bâtiments sur le territoire de la commune politique.

Les listes de ménages et de personnes sont créées dans le cadre de l'EC2000 (voir la procédure d'enquête dans la section 2).

5.2 Méthode d'échantillonnage

La procédure d'échantillonnage du plan TICINO utilise un tirage proportionnel à la taille sans remplacement (PPS sampling) au premier niveau (communes), puis aléatoire simple sans remplacement au niveau des bâtiments. La mesure de taille du tirage au premier niveau est le nombre de bâtiments dans la commune.

Pour une répartition équitable du travail, on choisit un nombre fixe $TCS = 60$ de bâtiments par commune. Le regroupement du travail au niveau communal, et non pas postal, ne pose pas de gros problèmes car la majorité des communes tessinoises ont un unique office de Poste.

5.2.1 Niveau des communes

L'ensemble $U_{(com)} = \{1, \dots, i, \dots, M\}$ des unités au premier niveau est formée de l'ensemble des communes politiques croisées avec les secteurs cadastraux (ci-dessous : communes). Les M unités sont réparties dans H strates de tailles M_h , $h = 1, \dots, H$. Le nombre de bâtiments dans la commune i de la strate h est noté B_{ih}^* . Il est calculé sur la base des données *provisoires* de la Stima.

Au premier niveau d'échantillonnage, on tire m_h communes dans la strate h . L'échantillon des communes est noté $s_{(com)}$. Il est de taille $m = \sum m_h$.

La probabilité d'inclusion π_{ih} de la commune i de la strate h est donnée par :

$$\pi_{ih} = P(i \in s_{(com)}) = m_h \frac{B_{ih}^*}{\sum_{i \in h} B_{ih}^*} \quad (19)$$

5.2.2 Niveau des bâtiments

Soit $U_{(batTI,ih)} = \{1, \dots, k, \dots, B_{ih}\}$ l'ensemble des bâtiments de la commune $i = 1, \dots, M_h$ de la strate $h = 1, \dots, H$ (données *finales*).

Au deuxième niveau d'échantillonnage, on tire $b_{ih} = \min(B_{ih}, TCS)$ bâtiments dans la commune i de la strate h . L'échantillon des bâtiments est noté $s_{(batTI)}$. Il est de taille $b = \sum \sum b_{ih}$.

La probabilité d'inclusion $\pi_{k|ih}$ du bâtiment k dans la commune i de la strate h , sachant que la commune i fait partie de l'échantillon $s_{(com)}$, est donnée par :

$$\pi_{k|ih} = P(k \in s_{(batTI)} | i \in s_{(com)}) = \frac{b_{ih}}{B_{ih}} = \min(1, \frac{TCS}{B_{ih}}) \quad (20)$$

5.2.3 Probabilité d'inclusion globale des bâtiments

La probabilité d'inclusion π_k du bâtiment k de la commune i de la strate h dans l'échantillon des bâtiments $s_{(batTI)}$ est donnée par :

$$\pi_k = P(k \in s_{(batTI)}) = \pi_{ih} \pi_{k|ih} = m_h \frac{b_{ih}}{\sum_{i \in h} B_{ih}^*} \frac{B_{ih}^*}{B_{ih}} \quad (21)$$

Dans les unités i de taille $B_{ih} > TCS$ ($b_{ih} = TCS$), la probabilité d'inclusion dépend de la commune i par le nouveau facteur $\psi_{ih} = B_{ih}^*/B_{ih}$, c'est-à-dire le rapport entre les tailles provisoires et finales. Le facteur ψ_{ih} n'est pas contrôlable et implique une variabilité dans les poids d'échantillonnage. Dans les unités i de taille $B_{ih} \leq TCS$ ($b_{ih} = B_{ih}$), les probabilités d'inclusion des bâtiments sont égales à celle de leur commune. Ce dernier cas est cependant peu probable dans les données du Tessin car la majorité des communes ont plus de $TCS = 60$ bâtiments.

Les probabilités d'inclusion des ménages et personnes sont égales aux probabilités d'inclusion des bâtiments dans lesquels ils résident car nous effectuons un relevé exhaustif dans les bâtiments de l'échantillon.

5.3 Plan au niveau des communes

Le plan d'échantillonnage au niveau des communes est stratifié. L'allocation est basée sur une allocation proportionnelle à la taille de la strate, avec léger éloignement pour tenir compte des difficultés attendues sur le terrain. La méthode d'échantillonnage est plus simple que dans le tirage des NPA du plan NORD. Les probabilités d'inclusion sont en effet moins contrôlables que dans le plan NORD car une grande variabilité provient de l'utilisation de données de bâtiments provisoires puis finales.

5.3.1 Base de sondage

La base de sondage est formée de 242 communes politiques et 8 secteurs cadastraux. Elle comporte $M = 250$ unités (ci-dessous : communes) avec $B^* = 101'621$ bâtiments.

Les informations auxiliaires considérées pour le plan d'échantillonnage sont le nombre d'habitants estimé dans la commune, le nombre de bâtiments dans les données *provisoire* de la Stima et la proportion de bâtiments avec des problèmes d'adresse 1 et 2; voir Table 17. Parmi les 250 communes, seule 15 ont moins de $TCS = 60$ bâtiments.

TAB. 17 – Définition des codes d'adresse au Tessin.

ADR	Description
0	adresse complète ("nom de rue" et "numéro de maison")
1	adresse pas complète mais coordonnées
2	adresse pas complète et pas de coordonnées métriques

5.3.2 Stratification et allocation

Les 250 communes sont distribués dans 6 strates selon leur population estimée et les problèmes d'adresses dans la commune; voir la Table 18.

Les classes de taille sont au nombre de trois : petite (moins de 1000 habitants), moyenne (entre 1000 et 8000 habitants) et grande (au moins 8000 habitants).

TAB. 18 – Définition de la stratification des communes.

Strate	Taille	Prob. adresse	M_h
1	petite	$< 70\%$ ADR=1 et $< 50\%$ ADR=2	75
2	petite	$\geq 70\%$ ADR=1 et $< 50\%$ ADR=2	43
3	petite	$\geq 50\%$ ADR=2	50
4	moyenne	$< 50\%$ ADR=1 et $< 50\%$ ADR=2	33
5	moyenne	$\geq 50\%$ ADR=1 ou $\geq 50\%$ ADR=2	46
6	grande		3
Total			250

L'allocation des $m = 20$ unités est donnée dans la Table 19. Elle se base sur l'allocation proportionnelle au nombre de communes avec un léger éloignement en faveur des unités avec peu de problèmes d'adresses. Les valeurs des poids "standardisés" $\omega_k = (\pi_k/\psi_{ih})^{-1}$ sont également données dans la Table 19 (hypothèse : $b_{ih} = TCS = 60$). On note que les poids sont semblables pour les petites et grandes communes. Ils sont environ deux fois plus grand dans les communes moyennes car elles comportent plus de bâtiments.

5.3.3 Tirage

Le tirage aléatoire des communes est réalisé avec la procédure SURVEYSELECT de SAS (method=pps). La mesure de taille est le nombre provisoire B_{ih}^* de bâtiments dans la commune.

Les 20 communes de l'échantillon comportent un total de 16'127 bâtiments, dont 24% avec code d'adresse ADR=1 et 34% avec code ADR=2 (Table 20). Nous avons 19 communes politiques entières et le secteur 2 de la commune politique de Lugano.

Si 60 bâtiments sont tirés dans chaque commune, on obtient un total de 1'200 bâtiments, ce qui correspond à un total de environ 2'700 personnes et 1'200 ménages. Les échan-

TAB. 19 – Allocation des communes dans les strates h (m_h parmi M_h), total du nombre provisoire de bâtiment dans la strate $\sum B_{ih}^*$ et poids "standardisés" $\omega_k = (\pi_k/\psi_{ih})^{-1}$.

Strate	M_h	m_h	$\sum B_{ih}^*$	ω_k
1	75	6	21013	58.4
2	43	3	10866	60.4
3	50	3	10850	60.3
4	33	3	24948	138.6
5	46	3	26529	147.4
6	3	2	7415	61.8
Total	250	20	101'621	

tillons finaux de ménages et personnes seront cependant plus petits car la liste finale des bâtiments comportera plus de bâtiments vides.

5.4 Plan au niveau des bâtiments

Le plan au niveau des bâtiment demande un effort particulier pour la construction de la base de sondage à partir des données *finales* de la Stima. A l'inverse du plan NORD, la liste des bâtiments doit être travaillée car les données disponibles comportent de multiples entrées ne correspondant pas à des bâtiments. Les choix sont toutefois fait de façon à garder un éventail large de bâtiments et donc de maximiser la couverture de la population.

5.4.1 Base de sondage

La base de sondage des bâtiments est un sous-ensemble de la liste des 28'013 bâtiments dans les 20 communes tirées (données *finales*). Elle est construite sur la base des codes ZDESPRI (voir Table 21) et CDESEDI01 (une partie des codes dans la Table 22) disponibles au niveau des bâtiments.

Notons que, parmi les 20 communes touchées par l'EC2000, seules 9 communes ont des données mises à jour (ajout des construction à partir de 1998/1999, nettoyage des données). Les 11 communes restantes ont des données encore partielles.

Le choix des exclusions a été fait avec l'aide des italophones du service linguistique de l'OFS, en répondant à la question "Ce code peut-il cacher un bâtiment habitable ou pas ?".

On commence par exclure les bâtiments "Non definito" (ZDESPRI=0) si aucun code n'est donné dans la variable CDESEDI01 et s'il n'y a pas d'indication sur le propriétaire. On exclu également les bâtiments avec les codes correspondants aux "casa di vacanza (monti)", "grotto" et "capannone (piu piani)" (CDESEDI01= 105, 114 et 116). Parmi les bâtiments restant, on conserve les "Abitativa" (ZDESPRI=1), ainsi que les codes CDESEDI01 entre 101 et 121 et entre 127 et 130. On retire ensuite 30 codes parmi les 90 codes CDESEDI01 ≥ 500 avec ZDESPRI > 1 ayant plus de 5 occurrences, voir Table 23. Ces

TAB. 20 – Information sur les unités sélectionnées. Nombre provisoire de bâtiments B_{ih}^* et répartition dans les problèmes d'adresse ADR=0, ADR=1 et ADR=2, population estimée P_{ih} , et densité de population par bâtiment P_{ih}/B_{ih}^* .

Strate	Commune	Unité	B_{ih}^*	ADR=0	ADR=1	ADR=2	P_{ih}	P_{ih}/B_{ih}^*
1	Marolta	50420	74	0	38	36	56	0.76
1	Prugiasco	50450	191	0	101	90	150	0.79
1	Vogorno	51350	478	0	251	227	309	0.65
1	Bissone	51540	267	184	0	83	791	2.96
1	Avegno	53020	362	0	187	175	513	1.42
1	Cavergno	53080	477	0	333	144	505	1.06
2	Aquila	50310	592	0	416	176	500	0.84
2	Bidogno	51500	232	0	166	66	355	1.53
2	Brusino-Arsizio	51600	318	3	231	84	444	1.4
3	San Nazzaro	51270	628	0	200	428	608	0.97
3	Vaglio	52280	272	0	24	248	487	1.79
3	Peccia	53200	206	0	27	179	206	1
4	Losone	51150	1618	1329	0	289	5912	3.65
4	Lugano	51922	1085	929	3	153	4211	3.88
4	Riva San Vitale	52630	793	513	51	229	2293	2.89
5	Malvaglia	50410	1121	0	575	546	1218	1.09
5	Melano	51970	322	0	228	94	1078	3.35
5	Biasca	52810	1574	1	901	672	5937	3.77
6	Bellinzona	50020	3146	2227	109	810	16759	5.33
6	Locarno	51130	2371	1584	3	784	14312	6.04
	Total		16127	6770	3844	5513	56644	3.51

types sont inscrits dans le champ "Specificare", *i.e.* pas préimprimés sur le questionnaire. Pour des raisons de temps, les environ 300 codes avec moins de 5 occurrences n'ont pas été triés. Ils ont cependant peu de chance de sortir dans l'échantillon.

Le choix de l'exclusion des "casa di vacanza (monti)" est un peu risqué mais il y a peu de chance que des personnes habitent ces bâtiments et leur exclusion permet de diminuer grandement les problèmes pratiques de la localisation de ces bâtiments dans la procédure d'enquête.

La base de sondage finale comporte 24'728 bâtiments. On a 15% de bâtiments avec le code ADR=1 et 54% avec le code ADR=2. Toutes les communes ont plus de $TCS = 60$ bâtiments. Les problèmes de localisation des bâtiments sont donc importants.

On note que la liste finale des bâtiments correspond a une moyenne de 2.3 habitants/bâtiment dans les 20 communes. Cette moyenne est inférieure à celle du plan NORD (4.2 habitants/bâtiment) et également à celle des NPA CLASSIC du plan NORD (2.4 habitants/bâtiment). La liste finale des bâtiments au Tessin devrait donc fournir une couverture acceptable de la population.

TAB. 21 – Liste des codes ZDESPRI de la Stima.

ZDESPRI	DESPRI
0	Non definito
1	Abitativa
2	Commerciale/ristorazione
3	Industriale/artigianale
4	Agricola
5	Altre destinationi
6	missing

TAB. 22 – Liste des codes CDESEDI01 notés explicitement sur le questionnaire de la Stima.

CDESEDI01	DESEDI01
101	CASA MONOFAMIGLIARE O BIFAMIGLIARE
102	ABITAZIONE DA 3 A 10 APPARTAMENTI
103	ABITAZIONE DA 11 O PIU' APPARTAMENTI
104	RUSTICO TRASFORMATO
105	CASA DI VACANZA (MONTI)
106	CASA A SCHIERA
107	STABILE AMMINISTRATIVO
108	ISTITUTO BANCARIO
109	CENTRO COMMERCIALE
110	NEGOZIO
111	ALBERGO
112	GARNI - PENSIONE
113	BAR - RISTORANTE
114	GROTTO
115	DEPOSITO (UN SOLO PIANO)
116	CAPANNONE (PIU' PIANI)
117	INDUSTRIA PESANTE
118	INDUSTRIA LEGGERA
119	LABORATORIO, OFFICINA
120	CASA COLONICA
121	RUSTICO AGRICOLO
122	STALLA PER BOVINI
123	STALLA PER CAPRINI,OVINI
124	FIENILE
126	SERRA
127	STRUTTURA SCOLASTICA
128	OSPEDALE - CLINICA
129	CASA PER ANZIANI
130	STAZIONE DI BENZINA
131	SERBATOIO

TAB. 23 – Codes exclus parmi les CDESEDI01 ≥ 500 , ZDESPRI > 1 avec au moins 5 occurrences.

CDESEDI01	nb	DESEDI01
514	5	LAVATOIO
515	29	LEGNAIA
520	29	POLLAIO
521	6	PORCILE
522	137	RIPOSTIGLIO
533	7	POSTEGGIO
552	6	PORTICO
560	6	CABINA
563	89	STALLA
593	75	DIROCCATO
598	29	CAPPELLA
602	5	DEPOSITO LEGNA
625	6	FONTANA
666	11	TETTOIA
670	9	APIARIO
694	5	DEPOSITO ATTREZZI
718	5	PIAZZALE
735	31	GARAGE
821	9	SPOGLIATOIO
841	9	WC
859	7	SEMIDIROCCATO
1009	8	SOSTRA
1101	8	CABINA DI TRASFORMAZIONE
1248	8	BOX
1541	5	IN DISUSO
20210	5	EVACUAZIONE E TRATTAMENTO DELLE ACQUE LU
20340	16	COSTRUZIONI DI ALTRE STRADE, PARCHEGGI
20791	5	SERBATOI PER OLIO, ALTRI SERBATOI
20793	9	POSTEGGI - GARAGES
20810	6	AGRICOLTURA

5.4.2 Allocation et tirage

On tire $b_{ih} = TCS = 60$ bâtiments dans chaque commune avec la procédure SURVEYSELECT de SAS (method=SRS). L'échantillon des bâtiment $s_{(batTI)}$ est de taille $b = 1'200$.

A l'inverse du plan NORD, les probabilités d'inclusion des bâtiments localisés dans les communes d'une même strate ne sont pas égales car le facteur B_{ih}^*/B_{ih} varie d'une commune à l'autre, voir équation (21) et Table 24.

TAB. 24 – Informations sur les communes de l'échantillon. Tailles provisoires B_{ih}^* et finales B_{ih} , probabilités d'inclusion π_{ih} des communes, nombre de bâtiments tirés b_{ih} et poids d'échantillonnage $\omega_k = 1/\pi_k$ (arrondis).

Strate	Commune	Unité	B_{ih}^*	π_{ih}	B_{ih}	B_{ih}^*/B_{ih}	b_{ih}	ω_k
1	Marolta	50420	74	0.0211	90 ^c	0.82	60	71
1	Prugiasco	50450	191	0.0545	224 ^c	0.85	60	68
1	Vogorno	51350	478	0.1365	953	0.5	60	116
1	Bissone	51540	267	0.0762	326	0.82	60	71
1	Avegno	53020	362	0.1034	602 ^c	0.6	60	97
1	Cavergho	53080	477	0.1362	1235	0.39	60	151
2	Aquila	50310	592	0.1635	1034	0.57	60	105
2	Bidogno	51500	232	0.0641	425	0.55	60	111
2	Brusino-Arsizio	51600	318	0.0878	540 ^c	0.59	60	103
3	San Nazzaro	51270	628	0.1736	865	0.73	60	83
3	Vaglio	52280	272	0.0752	393 ^c	0.69	60	87
3	Peccia	53200	206	0.057	427 ^c	0.48	60	125
4	Losone	51150	1618	0.1946	2156	0.75	60	185
4	Lugano	51922	1085	0.1305	1356	0.8	60	173
4	Riva San Vitale	52630	793	0.0954	1167 ^c	0.68	60	204
5	Malvaglia	50410	1121	0.1268	2495	0.45	60	328
5	Melano	51970	322	0.0364	485 ^c	0.66	60	222
5	Biasca	52810	1574	0.178	2678	0.59	60	251
6	Bellinzona	50020	3146	0.8486	4239 ^c	0.74	60	83
6	Locarno	51130	2371	0.6395	3038	0.78	60	79

^c : données finales nettoyées, avec les bâtiments construits après 1998/1999.

Dans les 1'200 bâtiments tirés, il reste 16 constructions qui ne sont pas habitables (codes pas contrôlé dans la construction de la base de sondage). Ces bâtiments sont retirés de la liste des bâtiments à envoyer aux offices de Poste. Le nombre de ménages est mis a priori à zéro dans ces bâtiments.

Notons encore que certaines adresses codées ADR=0 se sont avérées en fait des adresses pas complètes (par ex. rue="1" et numéro de maison ="0"). Les 8 codes d'adresses concernés ont donc modifiés (nouvelle variable : ADRCOR).

5.4.3 Identification des bâtiments vides a priori

Avant l'envoi aux offices de Poste, les 20 communes ont été mises à contribution pour localiser sur des cartes les 729 bâtiments de l'échantillon sans adresse et sans coordonnées (ADRCOR=2). Les communes ont indiqué l'emplacement des bâtiments sur une carte fournie par l'OFS et/ou donné une description des bâtiments.

Sur les 729 bâtiments avec ADRCOR=2 envoyés aux communes, nous avons gardé 163 bâtiments pour l'envoi aux offices de Poste. Les 566 bâtiments restants sont supposés vides (ex. poulailler, granges, abris). Le résumé de l'opération est résumé dans la Table 25. Il aboutit à un ensemble de 618 bâtiments pour lesquels on élabore une liste de ménages.

TAB. 25 – Information sur le contrôle des listes des bâtiments tessinois. Nombre b_{ih} de bâtiments tirés, nombre out de bâtiments retirés a priori, nombres $cor0$ et $cor1$ de bâtiments avec ADRCOR=0 et ADRCOR=1 après avoir retiré les bâtiments out , nombre $cor2$ de bâtiments avec ADRCOR=2 contrôlés par les communes, nombre $cor2ok$ de bâtiments habitables parmi les $cor2$, et nombre total $listes = cor0 + cor1 + cor2ok$ de bâtiments pour lesquels on élabore des listes de ménages.

Strate	Commune	unité	b_{ih}	out	$cor0$	$cor1$	$cor2$	$cor2ok$	$listes$
1	Marolta	50420	60	2	0	25	33	5	30
1	Prugiasco	50450	60	0	0	20	40	5	25
1	Vogorno	51350	60	0	0	17	43	2	19
1	Bissone	51540	60	0	42	0	18	7	49
1	Avegno	53020	60	1	0	19	40	5	24
1	Cavergno	53080	60	3	0	11	46	0	11
2	Aquila	50310	60	1	1	20	38	3	24
2	Bidogno	51500	60	0	0	16	44	8	24
2	Brusino-Arsizio	51600	60	1	1	18	40	11	30
3	San Nazzaro	51270	60	0	0	11	49	29	40
3	Vaglio	52280	60	0	0	1	59	22	23
3	Peccia	53200	60	3	0	7	50	8	15
4	Losone	51150	60	1	37	0	22	4	41
4	Lugano	51922	60	1	45	2	12	4	51
4	Riva San Vitale	52630	60	1	29	3	27	14	46
5	Malvaglia	50410	60	0	0	14	46	8	22
5	Melano	51970	60	0	0	24	36	8	32
5	Biasca	52810	60	0	1	21	38	7	29
6	Bellinzona	50020	60	0	35	1	24	6	42
6	Locarno	51130	60	2	33	1	24	7	41
	Total		1200	16	224	231	729	163	618

Dans chaque commune, on a deux groupes de bâtiments : (1) les bâtiments considérés a priori comme habitables et pour lesquels on construit des listes de ménages, et (2) les bâtiments considérés comme vides et n'ayant aucune chance d'être habités (poulailler, etc.).

Les bâtiments considérés comme vides sont identifiés à différents niveaux de la procédure :

- exclusion a priori juste après le tirage des bâtiments sur la base des informations de la Stima (voir variable *out* dans la Table 25, 16 cas) ;
- exclusion des cas clairement identifiés comme non habitables par les communes (*cor2 – cor2ok* dans la Table 25, 566 cas) ;
- exclusion de 5 cas de "monti" à Aquila sur la base des listes de ménages (voir ci-dessous).

5.5 Sous-échantillonnage de bâtiments

Les listes de ménages sont élaborées dans la grande majorité des cas (607 bâtiments) selon la procédure d'enquête utilisée pour le plan NORD : listes provisoires sur la base de l'annuaire de téléphone électronique, contrôle par les facteurs et saisie des corrections. Elles comportent 987 ménages répartis dans 538 bâtiments. Pour des raisons de délais, les listes dans les 11 bâtiments gardés parmi les 40 bâtiments ADRCOR=2 de Brusino-Arsizio sont effectuées par un interviewer de l'institut IHA et non pas par l'office de Poste. Ces 11 bâtiments abritent 6 ménages, répartis dans 5 bâtiments.

Après avoir mis 5 bâtiments de Aquila dans l'ensemble des "monti" (1 unique ménage explicitement noté "Ferienhaus", habités temporairement et sans téléphone), il reste 988 ménages répartis dans 538 bâtiments parmi les 613 bâtiments considérés comme habitables ou pas vides a priori.

Un sous-échantillonnage de bâtiments est appliqué afin de légèrement diminuer le nombre de ménages sans téléphone et notés comme logement temporaires ("chalets de vacances"). Il est minimal car le nombre de ménages au Tessin est déjà petit.

5.5.1 Base de sondage et stratification

La base de sondage est formée des $B_{subTI} = 1'200 = b$ bâtiments du plan TICINO.

Dans les 5 communes ayant beaucoup de bâtiments comportant un seul ménage sans téléphone et noté comme habité temporairement ou vide (ci-dessous : "chalets de vacances"), on crée deux sous-strates : (1) chalets de vacances, (2) autres bâtiments. Ces communes sont 50310 Aquila, 50420 Marolta, 50450 Prugiasco, 51500 Bidogno, et 51270 San Nazzaro.

Dans les 15 communes restantes, aucune sous-stratification n'est requise. On a une unique sous-strate comportant les bâtiments échantillonnés dans la commune.

Les $B_{subTI} = 1'200$ bâtiments sont donc répartis dans 25 strates $\ell = \{1, \dots, 25\}$ de taille $B_{sub, \ell i h}$ dans la commune i de la strate h .

La stratification dans les 5 communes touristiques est basée sur des informations recueillies lors de l'enquête. On se trouve donc, comme dans le plan NORD, dans un cas d'échantillonnage en deux phases. Les informations pour la stratification sont connues uniquement sur les bâtiments échantillonnés et non pas sur tous les bâtiments de la commune et le nombre de bâtiments dans chaque sous-strate est un nombre aléatoire. Les probabilités d'inclusion globales des bâtiments sous-échantillonnés sont égales au produit entre la probabilité d'être dans l'échantillon et la probabilité d'être choisi dans le sous-échantillon, mais le calcul de la variance est plus complexe.

5.5.2 Allocation

Soit $B_{sub,\ell ih}$ le nombre de bâtiments dans la sous-strate ℓ de la commune i de la strate h . On alloue $b_{\ell ih}$ bâtiments comme suit (" $\lfloor \cdot \rfloor$ " pour l'arrondi à l'entier inférieur) :

$$b_{\ell ih} = \lfloor \beta B_{sub,\ell ih} \rfloor$$

avec le facteur de réduction β défini par :

$$\beta = \begin{cases} 0.5 & \text{si strate "chalet vacances"} \\ 1 & \text{sinon} \end{cases} \quad (22)$$

On fait donc une réduction de l'échantillon uniquement dans les 5 sous-strates "chalets de vacances" des communes touristiques. Aucun sous-échantillonnage n'est effectué dans les bâtiments vides a priori.

5.5.3 Tirage

Le tirage aléatoire simple sans remplacement est réalisé avec la procédure SURVEYSELECT de SAS (method=SRS).

Le sous-échantillon contient $b_{sub} = 1'180$ bâtiments, dont 587 bâtiments sont considérés comme vides a priori et 593 bâtiments sont considérés habitables et contiennent un total de 968 ménages.

Les poids varient entre 68 et 328 (facteur 4.8, voir Table 26). Seuls les poids d'échantillonnage des 5 communes avec sous-stratification ont subi des variations.

TAB. 26 – Distribution des poids d'échantillonnage finaux des ménages (percentiles P_x , $x = \{0, 25, 50, 75, 100\}$) avec le nombre H_h de ménages échantillonné par strate (sans les 6 ménages spéciaux de Brusino-Arsizio, strate 2).

Strate	H_h	P_0	P_{25}	P_{50}	P_{75}	P_{100}
1	153	68.5	71.0	71.3	97.1	154.0
2	63	102.5	102.5	105.4	110.6	221.2
3	82	83.0	83.0	87.1	124.9	166.1
4	337	173.2	173.2	173.2	184.7	204.0
5	83	222.0	222.0	250.8	328.0	328.0
6	244	79.2	79.2	79.2	83.3	83.3
Total	962	68.5	83.0	116.4	173.2	328.0

6 Échantillon final de l'EC2000

L'ensemble des échantillons de ménages des plans NORD et TICINO forme l'échantillon des 27'331 ménages de l'EC2000⁵.

Les ménages contactés dans le cadre de l'EC2000 se répartissent dans tous les cantons sauf Appenzell Rhodes intérieurs et Obwald. La répartition géographique des unités d'échantillonnage au premier niveau (NPA et communes) est présentée dans la Figure 4. La répartition linguistique (langue officielles) est 67% allemand, 28% français et 4.5% italien.

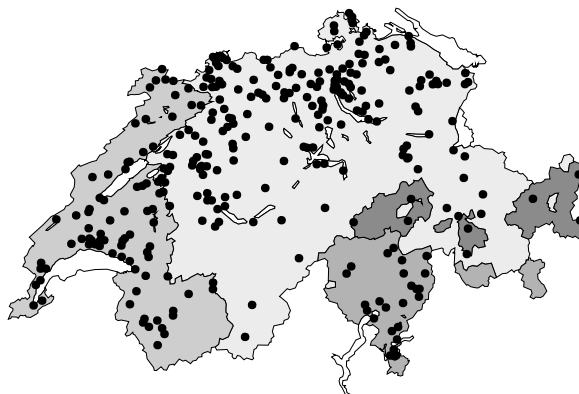


FIG. 4 – Répartition géographique des NPA et communes de l'échantillon de l'EC2000. Les différents tons de gris représentent les régions linguistiques (du plus clair au plus foncé : allemand, français, italien et romanche). Source : Fond de carte ©OFS GEOSTAT/S+T.

Globalement, 11.85% des ménages de l'échantillon n'ont pas de téléphone et devront donc être contactés en face-à-face. Parmi les contacts en face-à-face, 75% ont des adresses complètes, 21% nécessitent des cartes et 4% ne peuvent être identifiés que sur la base du nom (135 ménages).

La part de ménages habitants dans des logements indiqués comme habités temporairement est de 4%, et environ 0.5% des logements sont notés comme vides mais comportent le nom d'un ménage. L'information sur l'occupation des logements n'est cependant pas très fiable et ces valeurs ne sont qu'indicatives.

Il y a en tout 62 ménages collectifs. Ces derniers ne comportent pas forcément des résidents (ex. Hôtel). Ils ont été identifiés sur la base des indications du REAB et des facteurs. La qualité de cette identification n'est cependant probablement pas très bonne.

Les objectifs concernant les tailles d'échantillonnage sont atteints dans les différents groupes d'intérêt. On a env. 27'000 ménages avec suffisamment de personnes dans la variante CLASSIC et la variante TRANSIT, tout comme dans les différentes classes de taille des communes ; voir notamment la Table 15, p. 33. Le nombre de ménages au Tessin est proche bien que légèrement inférieur aux 1'000 prévus.

⁵Dans la commune tessinoise de Brusino-Arsizio un groupe de 6 ménages fait partie de l'échantillon mais ne sont pas interrogés par l'institut de sondage car ils ont été listés après le début de interviews. Ces ménages ne sont pas considérés dans les informations données dans cette section.

Les poids d'échantillonnage des ménages de l'EC2000 $\omega_k = 1/\pi_k$ varient entre 4 et 328, avec des coefficients de variation entre 0% et 45% selon la variante de relevé, voir Table 27.

TAB. 27 – Nombre de ménages H par variante de relevé et statistiques sur les poids d'échantillonnage des ménages (moyenne *mean*, coefficient de variation *cv*, maximum *max* et minimum *min*).

Variante	H	<i>mean</i>	<i>cv</i> [%]	<i>max</i>	<i>min</i>
CLASSIC	4732	24.04	37.56	55.44	15.83
SEMI-CLASSIC	2451	31.86	14.95	61.20	4.43
TRANSIT	18792	166.32	15.74	325.77	44.98
FUTURE	394	71.46	0.01	71.47	71.45
TICINO	962	136.39	44.69	328.03	68.45
Total	27331	127.20	52.00	328.03	4.43

L'ensemble des ménages a été envoyé à l'institut IHA après avoir détecté 5 doubles rescapés des contrôles. L'institut a donc reçu un ensemble de 27'326 ménages pour les interviews.

7 Conclusion

Le plan d'échantillonnage NORD est construit de façon à atteindre un nombre minimal de personnes dans différents sous-groupes de la population en contrôlant la variabilité des poids d'échantillonnage dans chaque variante de relevé du recensement. La méthode proposée permet d'obtenir des poids de ménages dépendant essentiellement du groupe de strate dans lequel ils se trouvent (régions avantageuses, intermédiaires ou difficiles). La mise en application de la méthode a cependant nécessité quelques adaptations pour contrer l'éloignement du tirage PPS dans quelques numéros postaux et tournées.

Le plan d'échantillonnage TICINO est influencé par l'utilisation de jeux de données provisoires puis finaux au niveau des bâtiments. La conséquence principale est la variabilité des poids d'échantillonnage au niveau des bâtiments, ménages et personnes.

Les plans NORD et TICINO reposent sur les listes de bâtiments et les listes de ménages utilisées comme référence. La qualité des listes des bâtiments du REAB et de la Stima sera évaluée par comparaison avec les données du recensement. La qualité des listes de ménages, construites dans le cadre de l'EC2000, est plus difficile à évaluer car aucun relevé auxiliaire n'est disponible pour une comparaison. Malgré le bon déroulement des opérations sur le terrain divers problèmes ont été détectés. Ces derniers auront probablement des conséquences sur la qualité de l'appariement avec le recensement et les estimations de la couverture.

Un poids important a été mis sur les côtés pratiques de l'enquête et son acceptabilité parmi les personnes touchées. On a ainsi regroupé les bâtiments, non seulement dans un nombre fixe d'offices de Poste mais aussi dans des tournées pour limiter le travail des facteurs. On a aussi essayé de minimiser le nombre d'interviews CAPI dans des régions

touristiques (chalets de vacances) par une procédure ciblée de sous-échantillonnage. Ces choix ont semble-t'il été apprécié sur le terrain mais ils compliquent la méthode d'échantillonnage et ne pourront probablement pas être tous pris en compte dans les estimations de la variance des estimateurs de la méthode duale. Avec un peu de recul, nous ne referions probablement pas de stratification dans le sous-échantillonnage des bâtiments de communes touristiques car elle provoque de la variabilité dans les poids et correspond à un tirage en deux phases. Le niveau des tournées ne serait également utilisé que dans les numéros postaux comprenant beaucoup de tournées (par exemple au moins 5 tournées), afin d'éviter les éloignements du tirage PPS et les problèmes du traitement des bâtiments sans tournées.

Les 27'332 ménages de l'échantillon final de l'EC2000 sont répartis dans 305 localités (NPA ou communes). Ils permettent d'atteindre le nombre désirés de personnes dans les différents sous-groupes d'intérêt.

Annexes

A Recensement de la population de l'an 2000

Le recensement fédéral de la population de l'an 2000 (RFP2000) est également nommé relevé structurel de la Suisse. Il a pour but de récolter des informations sur les 7.1 millions de personnes, 3.3 millions de ménages et 1.5 millions de bâtiments en Suisse. Le jour de référence du RFP2000 est le 5 décembre 2000.

Le recensement de la population est régi par la loi sur le recensement fédéral de la population du 26 juin 1998 et l'ordonnance sur le recensement fédéral de la population de l'an 2000 du 13 janvier 1999. L'office fédéral de la statistique est responsable de l'organisation et de la bonne marche du recensement mais la collecte des données est sous la responsabilité des communes politiques du pays.

Dans le cadre du RFP2000, les communes avaient le choix entre les 4 variantes de relevé suivantes :

CLASSIC : des agents recenseurs distribuent et collectent les questionnaires auprès de la population, comme lors des recensements précédents ;

SEMI-CLASSIC : les questionnaires sont préimprimés sur la base du registre des habitants et envoyés par la Poste à la population. Des agents recenseurs sont ensuite chargés de collecter les questionnaires auprès des ménages ;

TRANSIT : les questionnaires sont préimprimés sur la base du registre des habitants. L'envoi et le retour des questionnaires passent par la Poste ;

FUTURE : idem que TRANSIT avec lien entre les ménages et les logements déjà réalisé dans le registre des habitants.

La plupart des communes TRANSIT ont profité d'une organisation centralisée pour la préimpression des questionnaires, la distribution par la Poste et le contrôle des retours.

Le canton du Tessin a organisé le relevé dans son canton de façon centralisée et indépendante du reste de la Suisse. La variante de relevé au Tessin est similaire à la variante TRANSIT. Un résumé des choix est donné dans la Table 28.

TAB. 28 – Information sur les variantes de relevé du recensement de l'an 2000, état au 15 novembre 2000.

Variante	Communes	Population [%]
CLASSIC	690	3.8%
SEMI-CLASSIC	223	2.5%
TRANSIT	1717	88.6%
FUTURE	21	0.8%
TESSIN	245	4.3%
Total	2896	100%

Deux types de ménages coexistent dans le recensement : les ménages privés et les ménages collectifs. Les ménages privés se composent d'une ou de plusieurs personnes qui vivent dans

un même logement. Les familles, les couples, les personnes seules ou encore les personnes vivant en colocation forment des ménages privés. Les ménages collectifs sont des groupes de personnes qui ne constituent pas un ménage privé. Les membres des établissements d'exécution de peines et mesures, des internats, des établissements de soins, des cloîtres ou des pensions forment des ménages collectifs.

On définit les domiciles civils, de séjour et économiques comme suit⁶ :

- Une personne a son domicile civil dans la commune où elle a déposé ses papiers si elle est de nationalité suisse ou dans la commune qui lui a fourni son autorisation de séjour si elle est de nationalité étrangère.
- Le domicile de séjour est différent du domicile civil dans les cas où la personne a deux domiciles. On estime qu'une personne donnée a aucun ou un seul domicile de séjour différent du domicile civil.
- Le domicile économique est identique au domicile civil ou au domicile de séjour. Il est identique au domicile civil si la personne n'a qu'un domicile. Il est en principe identique au domicile dans lequel la personne vit la majeure partie de la semaine si elle a deux domiciles.

Selon l'ordonnance sur le recensement de l'an 2000, la population résidente d'une commune comprend toutes les personnes qui ont leur domicile économique dans cette commune.

B Enquête pilote de l'EC2000

L'enquête pilote de l'EC2000 (ECtest) a eu lieu au printemps 2000 dans 4 communes ayant participé à la répétition générale du recensement (GP99 ; jour de référence 30 novembre 1999). Elle avait pour but de tester l'élaboration des listes de ménages provisoires, les formulaires de bâtiments, le contact avec la Poste et les interviews (questionnaires, acceptabilité, charge de travail, temps d'interviews).

Les communes de Aarau, Pratteln, Founex et Nendaz sont choisies pour l'ECtest. Ce choix permet d'atteindre des communes de langues allemande et française, de différents types, de densité différentes et avec des adresses de bâtiments bien définies, très partielles ou même inexistantes. Un ensemble de 282 bâtiments ont été choisis dans le répertoire des adresses de bâtiments (REAB) contrôlé par les communes pour la GP99. Le choix tient compte, d'une part, du besoin de grande diversité de situations et, d'autre part, des besoins pratiques tels que la répartition équitable parmi les facteurs tout en limitant le nombre de circonscriptions. Le nombre de bâtiments et de ménages choisis dans chaque commune est donné dans la Table 29.

Des interviews ont eu lieu avec 317 ménages en CATI et 38 ménages en CAPI, avec un taux de réponse global de 77%.

Notons que les listes provisoires et finales de ménages de l'ECtest ont été construites par l'entreprise Data Care AG (DCL). Les listes provisoires utilisaient la base de données des déménagements ; formée sur la base des annonces des déménagements recueillies par la Poste. Ces listes avaient été envoyées aux offices de Poste et complétées par la saisie des

⁶selon l'ordonnance et le document de F.Rais "E-17-A Domicile : définition" du 10 mai 1999.

TAB. 29 – Information sur les échantillons de l'ECtest. Nombre *bat* de bâtiments, nombre *tour* de tournées, nombre *tel* de ménages avec téléphone et nombre *notel* de ménages sans téléphone.

Commune	<i>bat</i>	<i>tour</i>	<i>tel</i>	<i>notel</i>
Aarau	59	3	134	14
Pratteln	77	4	134	15
Founex	74	4	87	15
Nendaz	72	≥ 2	68	15
Total	282	≥ 13	409	59

corrections et la recherche des numéros de téléphone. Pour l'EC2000, l'élaboration des listes de ménages a été reprise par l'OFS. Seul le contact avec les offices de Poste est encore sous la responsabilité de DCL.

C Listes provisoires de ménages

Les listes provisoires des ménages habitant dans les bâtiments de l'échantillon de l'EC2000 sont construites pour tous les bâtiments avec une adresse postale composée d'un nom de rue, d'un numéro de maison et d'un numéro postal.

Les listes provisoires des ménages habitant dans les bâtiments de l'échantillon de l'EC2000 sont construites par étapes :

1. préparation des adresses des bâtiments ;
2. recherche dans l'annuaire de téléphone électronique (ETV) des entrées correspondant aux adresses des bâtiments ;
3. nettoyage des données trouvées dans ETV.

La préparation des adresses des bâtiments consiste à adapter le champ "nom de rue" donné dans le REAB ou la Stima pour optimiser la recherche dans ETV. A l'aide d'un programme SAS, on exclu par exemple les mots "le", "rue", "route", "r.", ou "rte" du nom de rue. La simplification est effectuée selon un ordre de priorité prédéfini, en assurant l'existence de au moins un mot dans le nom (ex. "rue du Pont" devient "PONT" et "Pont de la ville" devient "VILLE").

La recherche dans ETV est effectuée en deux parties successives à l'aide du logiciel ETV Batchman de l'entreprise NOVA ESPRIT. La première recherche est basée sur les champs "nom de rue", "numéro de maison" et "numéro postal". La deuxième recherche prend également en compte la localité. Les deux recherches se limitent aux entrées privées dans ETV (pas les entreprises) avec un taux de concordance de 80%.

Le nettoyage des données trouvées dans ETV est effectué en partie automatiquement à l'aide d'un programme SAS et en partie manuellement, après partage des données dans les groupes suivants :

1. adresses sans aucune entrée dans ETV ;

2. entrées trouvées dans ETV mais sans nom ni prénom (ex. records partiels d'entreprises) ;
3. entrées complètes mais avec une erreur dans le numéro de la maison (ex. 13A au lieu de 13) ou numéro de fax ou aucun numéro de téléphone fourni ("liste rouge") ;
4. entrées sans erreur au niveau du numéro de maison mais avec un nom de rue pas identique. Le cas se présente lors de doubles adresses pour une même personne, ex. Jean Martin, rue de la Gare 10 à Lausanne et rue du Pont 2 à Vevey ;
5. entrées sans erreur au niveau du numéro de maison et du nom de rue.

Les entrées dans le groupe 4 sont contrôlées manuellement pour enlever les entrées effectivement erronées et garder celles qui sont correctes. L'ensemble des données correctes du groupe 4 et celles du groupe 5 sont ensuite traitées de façon à enlever les doubles au niveau des noms et des numéros de téléphone, respectivement. Si une même personne a deux numéros de téléphone, le deuxième téléphone est conservé en vue des interviews.

Exemple :

Nom	Prénom	Numéro de téléphone
Martin	Pierrette	01/4444444
Martin	Pierre	01/4444444
Mueller	Jean	021/2222222
Mueller	Jean	079/3333333
Schmidt	Heidi	021/2222222

Devient :

Nom	Prénom	Numéro 1	Numéro 2
Martin	Pierrette	01/4444444	.
Mueller	Jean	021/2222222	079/3333333

Un contrôle final permet d'exclure les cas erronés restants (ex. entrées pour la "rue du Pont" au lieu de la "place du Pont"). Des listes provisoires sont ainsi créées pour les 14'918 bâtiments ayant des adresses complètes sur les 16'792 bâtiments de l'échantillon $s_{(bat)}$. Les listes provisoires finales, envoyées aux offices de Poste, comportent 22'389 ménages, voir Table 30.

TAB. 30 – Listes provisoires. Nombre *bat* de bâtiments avec des adresses complètes, nombre *rien* de bâtiments pour lesquels aucun ménage n'a été trouvé dans ETV et nombre *nprov* de ménages dans les listes provisoires finales.

Région	<i>bat</i>	<i>rien</i>	<i>nprov</i>
Suisse alémanique	10'714	3'226	16'761
Suisse romande	3'637	1'665	5'298
Suisse GR italophone	343	338	3
Tessin	224	87	327
Total	14'918	5'316	22'389

D Échantillon des numéros postaux (plan NORD)

Légende :

- NPA : numéro postal à 6 chiffres ;
- ComNr : numéro officiel OFS de la commune majoritaire dans le NPA ;
- Cant : numéro du canton selon la classification NUTS3 ;
- Lang : langue officielle de la commune majoritaire dans le NPA (d : allemand et romanche, f : français et i : italien) ;
- Nom : appellation du numéro postal ;
- Stradap : strate au premier niveau, voir la définition dans la Table 8, p. 22.

NPA	ComNr	Cant	Lang	Nom	Stradap						
100026	5586	22	f	Lausanne 26	18	186500	5411	22	f	Les Diablerets	15
100400	5586	22	f	Lausanne	18	186800	6152	23	f	Collombey	15
100802	5585	22	f	Jouxens-Mézery	15	187300	6157	23	f	Val-d'Illiez	19
101000	5586	22	f	Lausanne	18	190800	6139	23	f	Riddes	9
101200	5586	22	f	Lausanne	16	191800	6139	23	f	Mayens-de-Riddes	13
102300	5583	22	f	Crissier	17	192000	6136	23	f	Martigny	18
102400	5635	22	f	Ecublens VD	17	192100	6137	23	f	Martigny-Croix	13
104103	5471	22	f	Bettens	12	192201	6218	23	f	Granges (Salvan)	13
105300	5516	22	f	Cugy VD	15	192300	6218	23	f	Les Marécottes	11
106300	5664	22	f	Chapelle-s-Moudon	3	193300	6035	23	f	Sembrancher	19
107000	5607	22	f	Puidoux-Gare	20	193700	6034	23	f	Orsières	6
112300	5621	22	f	Aclens	12	194300	6034	23	f	Praz-de-Fort	41
112800	5644	22	f	Reverolle	10	196500	6265	23	f	Savièse	19
120400	6621	25	f	Genève	18	201600	6408	24	f	Cortailod	15
125300	6642	25	f	Vandoeuvres	15	203400	6412	24	f	Peseux	15
126000	5724	22	f	Nyon	16	203500	6407	24	f	Corcelles NE	15
127100	5720	22	f	Givrins	41	207200	6459	24	f	St-Blaise	15
127403	5722	22	f	Grens	1	211400	6506	24	f	Fleurier	15
129000	6644	25	f	Versoix	18	233600	6742	26	f	Les Bois	17
129700	5717	22	f	Founex	9	234500	6743	26	f	Les Breuleux	17
132300	5761	22	f	Romainmôtier	8	252500	6455	24	f	Le Landeron	15
134300	5873	22	f	Les Charbonnières	15	255400	390	2	d	Meinisberg	15
137300	5749	22	f	Chavornay	15	255500	733	2	d	Brügg BE	17
140000	5938	22	f	Yverdon-les-Bains	16	257700	499	2	d	Siselen BE	41
140402	5935	22	f	Villars-Epeney	7	273200	703	2	f	Reconvilier	17
142801	5566	22	f	Provence	5	274000	700	2	f	Moutier	17
143500	5915	22	f	Essert-Pittet	7	274600	691	2	f	Crémines	2
152200	5675	22	f	Lucens	15	280000	6711	26	f	Delémont	18
153000	5822	22	f	Payerne	9	285400	6701	26	f	Bassecour	11
156300	2014	10	f	Dompierre FR	2	290000	6800	26	f	Porrentruy	17
156400	2013	10	f	Domdidier	42	290400	6776	26	f	Bressaucourt	3
160701	5795	22	f	Palézieux-Village	9	293500	6773	26	f	Beurnevésin	3
160802	5781	22	f	Bussigny-sur-Oron	7	294600	6794	26	f	Miécourt	3
160903	2327	10	f	Fiaugères	1	294700	6779	26	f	Charmoille	3
161400	2328	10	f	Granges (Veveyse)	15	301100	351	2	d	Bern	18
164100	2161	10	f	Vuippens	1	301300	351	2	d	Bern	18
166800	2144	10	f	Neirivue	1	301400	351	2	d	Bern	18
166903	2121	10	f	Sciernes-d'Albeuve	14	303200	360	2	d	Hinterkappelen	17
169502	2223	10	f	Rueyres-St-Laurent	1	303800	354	2	d	Kirchlindach	17
172600	2192	10	f	Farvagny	15	304300	360	2	d	Uettligen	17
172900	2176	10	f	Bonnefontaine	2	307200	363	2	d	Ostermundigen	18
176200	2197	10	f	Givisiez	11	307600	627	2	d	Worb	18
176300	2198	10	f	Granges-Paccot	17	312200	870	2	d	Kehrsatz	17
178300	2243	10	f	Pensier	2	312700	876	2	d	Mühlethurnen	41
178403	2283	10	f	Wallenried	1	315400	853	2	d	Rüschegg Heubach	9
178500	2257	10	f	Cressier FR	2	320800	665	2	d	Gurbrü	1
180200	5883	22	f	Corseaux	9	321500	2260	10	d	Gempnach	1
181500	5886	22	f	Clarens	16	328000	2275	10	d	Murten	17
183602	5842	22	f	La Tine	20	328300	304	2	d	Kallnach	15
184400	5414	22	f	Villeneuve VD	15	329800	391	2	d	Überwil b. Büren	15
185400	5407	22	f	Leysin	15	331701	545	2	d	Mülchi	3
						332201	543	2	d	Mattstetten	2

332202	551	2	d	Urtenen	17	564700	4237	19	d	Oberrüti	2
336000	979	2	d	Herzogenbuchsee	17	570300	4209	19	d	Seon	9
337602	972	2	d	Berken	7	572200	4006	19	d	Gränichen	21
337700	991	2	d	Walliswil b. Wangen	15	572600	4146	19	d	Unterkulm	15
341200	407	2	d	Heimiswil	17	574600	2585	11	d	Walterswil SO	17
341800	956	2	d	Rüegsbach	15	600400	1061	3	d	Luzern	16
343500	955	2	d	Ramsei	15	600600	1061	3	d	Luzern	16
361200	939	2	d	Steffisburg	16	601400	1060	3	d	Littau	18
361600	945	2	d	Schwarzenegg	15	616600	1005	3	d	Hasle LU	13
362600	929	2	d	Hünibach	15	619700	906	2	d	Schangnau	15
363300	921	2	d	Amsoldingen	2	621500	1081	3	d	Beromünster	19
366300	867	2	d	Gurzelen	15	624700	1143	3	d	Schötz	19
366500	886	2	d	Wattenwil	15	631302	1704	9	d	Edlibach	17
375300	762	2	d	Oey	2	633000	1702	9	d	Cham	16
375500	762	2	d	Horboden	2	637000	1509	7	d	Stans	17
376500	766	2	d	Überwil i. S.	41	637400	1502	7	d	Buochs	17
381400	577	2	d	Gsteigwiler	3	646600	1204	4	d	Bauen	5
381800	576	2	d	Grindelwald	15	649000	1202	4	d	Andermatt	9
386300	781	2	d	Gadmen	3	653500	3834	18	i	Roveredo GR	2
392000	6300	23	d	Zermatt	19	654100	3810	18	i	Sta. Maria	1
392500	6285	23	d	Grächen	2	654300	3801	18	i	Arvigo	1
396000	6248	23	f	Sierre	18	655700	3831	18	i	Cama	3
397200	6241	23	f	Miège	17	656200	3823	18	i	Soazza	1
398108	6060	23	d	Gluringen	5	656300	3822	18	i	Mesocco	11
405100	2701	12	d	Basel	18	700000	3901	18	d	Chur	16
405200	2701	12	d	Basel	18	705000	3921	18	d	Arosa	19
410200	2765	13	d	Binningen	16	714900	3606	18	d	Vrin	8
410300	2767	13	d	Bottmingen	15	715800	3616	18	d	Waltensburg/Vuorz	15
410800	2481	11	d	Witterswil	17	730700	3952	18	d	Jenins	15
412300	2762	13	d	Allschwil	18	732300	3297	17	d	Wangs	15
412700	2766	13	d	Birsfelden	18	742300	3666	18	d	Sarn	1
413200	2770	13	d	Muttenz	18	743101	3503	18	d	Mutten	8
415300	2773	13	d	Reinach BL	16	744701	3681	18	d	Am Bach (Avers)	7
420400	2618	11	d	Himmelried	17	746400	3536	18	d	Parsonz	15
420800	2621	11	d	Nunningen	17	749300	3514	18	d	Schmitten (Albula)	1
422200	2793	13	d	Zwingen	17	753700	3843	18	d	Müstair	11
422700	2614	11	d	Büsserach	11	754200	3744	18	d	Susch	1
423200	2616	11	d	Fehren	17	755800	3753	18	d	Strada	10
424300	2784	13	d	Dittingen	11	800300	261	1	d	Zürich	16
425201	2611	11	d	Bärschwil Dorf	41	800600	261	1	d	Zürich	16
433300	4172	19	d	Münchwilen AG	41	800800	261	1	d	Zürich	16
440200	2824	13	d	Frenkendorf	17	803700	261	1	d	Zürich	16
441000	2829	13	d	Liestal	18	804600	261	1	d	Zürich	16
443700	2895	13	d	Waldenburg	41	805200	261	1	d	Zürich	16
449700	2860	13	d	Rünenberg	41	811200	94	1	d	Uetelfingen	17
450000	2601	11	d	Solothurn	18	815300	97	1	d	Rümlang	17
452500	2541	11	d	Balm b. Günsberg	3	815500	90	1	d	Niederhasli	15
453900	987	2	d	Rumisberg	1	819500	70	1	d	Wasterkingen	41
455600	2511	11	d	Aeschi SO	41	819700	67	1	d	Rafz	17
455802	2521	11	d	Heinrichswil	2	820000	2939	14	d	Schaffhausen	16
460000	2581	11	d	Olten	18	821300	2904	14	d	Neunkirch	17
461300	2582	11	d	Rickenbach SO	17	822500	2953	14	d	Sibingen	6
465700	2573	11	d	Dulliken	42	823500	2917	14	d	Lohn SH	41
471400	2421	11	d	Aedermannsdorf	11	824200	2913	14	d	Bibern SH	3
491400	337	2	d	Roggwil BE	17	824300	2911	14	d	Altdorf SH	3
493802	339	2	d	Rohrbachgraben	1	825300	4541	20	d	Diessenhofen	15
500000	4001	19	d	Aarau	21	825500	4536	20	d	Schlattingen	15
502600	4004	19	d	Densbüren	9	827402	4651	20	d	Gottlieben	1
507300	4165	19	d	Gipf-Oberfrick	15	830400	69	1	d	Wallisellen	21
507400	4161	19	d	Eiken	9	830900	64	1	d	Nürensdorf	15
521300	4122	19	d	Villnachern	9	831000	176	1	d	Kemptthal	17
522300	4111	19	d	Riniken	9	834200	117	1	d	Wernetshausen	17
522400	4119	19	d	Unterbözingen	2	836000	4724	20	d	Eschlikon TG	17
532400	4307	19	d	Full-Reuenthal	21	840000	230	1	d	Winterthur	18
541300	4024	19	d	Birmenstorf AG	15	842200	224	1	d	Pfungen	17
545300	4039	19	d	Remetschwil	15	842600	63	1	d	Lufingen	41
560400	4198	19	d	Hendschiken	9	846700	41	1	d	Truttikon	3
562200	4240	19	d	Waltenschwil	9	847102	214	1	d	Dägerlen	14
563000	4236	19	d	Muri AG	17	847400	216	1	d	Dinhard	11
564200	4235	19	d	Mühlau	17	850000	4566	20	d	Frauenfeld	16

852300	220	1	d	Hagenbuch ZH	41	888400	3295	17	d	Überterzen	15
853200	4621	20	d	Warth	41	890503	4061	19	d	Arni AG	15
855500	4831	20	d	Müllheim Dorf	15	890900	2	1	d	Zwillikon	15
856600	4666	20	d	Neuwilen	2	891200	10	1	d	Obfelden	15
856602	4666	20	d	Dotnacht	6	895400	244	1	d	Geroldswil	15
860000	191	1	d	Dübendorf	16	896200	4023	19	d	Bergdietikon	15
860300	197	1	d	Schwerzenbach	15	901400	3203	17	d	St. Gallen	16
861000	198	1	d	Uster	16	905300	3024	15	d	Teufen AR	17
862700	116	1	d	Grünigen	17	906300	3005	15	d	Stein AR	17
870000	154	1	d	Küsnacht ZH	16	910000	3001	15	d	Herisau	18
870200	161	1	d	Zollikon	16	911300	3401	17	d	Degersheim	15
870800	155	1	d	Männedorf	15	920400	3441	17	d	Andwil SG	15
871200	158	1	d	Stäfa	16	923000	3402	17	d	Flawil	15
875000	1609	8	d	Glarus	42	942702	3038	15	d	Zelg (Wolfhalden)	15
875500	1607	8	d	Ennenda	17	944500	3255	17	d	Rebstein	15
876200	1627	8	d	Schwanden GL	17	945000	3251	17	d	Altstätten SG	16
876202	1628	8	d	Schwändi GL	3	947200	3273	17	d	Grabs	15
876203	1629	8	d	Sool	5	954800	4591	20	d	Matzingen	15
876700	1605	8	d	Elm	6	962100	3375	17	d	Überhelfenschwil	17
880800	1322	5	d	Pfäffikon SZ	16	964200	3352	17	d	Ebnat-Kappel	15
885300	1344	5	d	Lachen SZ	2	965200	3354	17	d	Neu St. Johann	9

E Programmes SAS et S-PLUS

Les fichiers suivants contiennent les programmes SAS (versions 8 et 8.2 sur UNIX) et S-PLUS (version 6 sur UNIX) utilisés pour obtenir les résultats présentés dans le présent rapport.

Plan NORD

Construction de la base de sondage au niveau des NPA :

- ... \ProgSAS\lecframe.sas
- ... \ProgSAS\constframe2.sas
- ... \ProgSAS\framefinal.sas

Stratification des NPA :

- ... \ProgSAS\strafinal.sas
- ... \ProgSAS\stratirage.sas

Allocation des NPA (simulation en S-PLUS) :

- ... \ProgSplus\allocfin.s
- ... \ProgSplus\allocfin.xls

Tirage des NPA :

- ... \ProgSAS\stratirage.sas

Échantillonnage aux niveaux des tournées et des bâtiments et correction des codes d'adresse :

- ... \ProgSAS\Tournee\frametour.sas
- ... \ProgSAS\Tournee\framebat3.sas
- ... \ProgSAS\Tournee\fileoutput.sas

Lecture des fichiers de la saisie, statistiques de base, et sous-échantillonnage :

- ... \ProgSAS\HHBATsais2fin\lecHHBATsais.sas
- ... \ProgSAS\HHBATsais2fin\prepsub.sas
- ... \ProgSAS\HHBATsais2fin\subbat3.sas

Plan TICINO

Échantillonnage des communes :

- ...\\ProgSAS\\Tessin\\frameTioct00_2.sas

Échantillonnage des bâtiments, correction codes d'adresse ;

- ...\\ProgSAS\\Tessin\\lecTidec00.sas
- ...\\ProgSAS\\Tessin\\frameTidec00.sas
- ...\\ProgSAS\\Tessin\\fileoutTI.sas

Lecture des fichiers de la saisie, stats de base, frame et sous-échantillonnage :

- ...\\ProgSAS\\HHBATsais2fin\\lecHHBATsaisTI.sas
- ...\\ProgSAS\\HHBATsais2fin\\prepsubTI.sas

Autres

Programmes de traitement des listes de ménages provisoires (programmes développés par Markus Eichenberger (BIT) et adaptés à l'OFS) :

- ...\\ProgSAS\\listeprovec2000\\etvpgm1_final.sas
- ...\\ProgSAS\\listeprovec2000\\etvpgm2_final_part.sas
- ...\\ProgSAS\\listeprovec2000\\etvmac4.sas
- ...\\ProgSAS\\listeprovec2000\\etvword3.txt

Combinaison NORD et TICINO, tests doubles, output pour institut et divers :

- ...\\ProgSAS\\HHBATsais2fin\\outallHH.sas
- ...\\ProgSAS\\HHBATsais2fin\\diversall.sas

References

- ABS (1997). The 1996 census of population and housing. Annual report 1996-97, Australian Bureau of Statistics.
- Bentley, J.L. and Floyd, R. (1987). A sample of brilliance. *Communications of the Association for Computing Machinery*, 30 :754–757.
- Bentley, J.L. and Knuth, D. (1986). Literate programming. *Communications of the Association for Computing Machinery*, 29 :364–369.
- Brown, J. J., Diamond, I. D., Chambers, R. L., Buckner, L. J., Teague, A. D. (1999). A methodological strategy for a one-number census in the UK. *J. R. Statist. Soc. A*, 162(2) :247–267.
- Childers, D. (2000). Accuracy and coverage evaluation : The design document. DSSD Census 2000 Procedures and Operation Memorandum Series, Chapter S-DT-1, U.S. Census Bureau.
- Davis, P. (2001). Accuracy and coverage evaluation : Dual system, estimation results. DSSD Census 2000 Procedures and Operations Memorandum Series B-9*, U.S. Census Bureau.
- Fienberg, S. E. (1992). Bibliography on capture-recapture modelling with application to census undercount adjustment. *Survey methodology*, 18(1) :143–154.
- Griffin, R. (2000). Accuracy and coverage evaluation : Dual system estimation. DSSD Census 2000 Procedures and Operation Memorandum Series #Q-20, U.S. Census Bureau.
- Hanurav, T. (1967). Optimum utilization of auxiliary information : π_{ps} sampling of two units from a stratum. *Journal of the Royal Statistical Society, Series B*, 29 :374–391.
- Hogan, H. (1992). The 1990 post-enumeration survey : An overview. *The American Statistician*, 46(4) :261–269.
- Hogan, H. (1993). The post-enumeration survey : Operations and results. *Journal of the American Statistical Association*, 88(423) :1047–1060.
- McLennan, W. (1996). Census population and housing : Data quality - undercount. Information paper 2940.0, Australian Bureau of Statistics.
- Navarro, A. (2000). Accuracy and coverage evaluation : Targeted extended search plans. DSSD Census 2000 Procedures and Operation Memorandum Series #Q-18.
- Renaud, A. (2002). Methodology of the swiss census 2000 coverage survey. In *ASA, 2001 Proceedings of the Section on Survey Research*. American Statistical Association, to appear.
- Särndal, C. E., Swensson, B. and Wretman, J. (1992). *Model Assisted Survey Sampling*. Springer-Verlag, New-York.
- Statistique Canada (1999). Couverture. Rapports techniques du recensement de 1996. 92-370-XIF, Statistique Canada.
- Vijayan, K. (1968). An exact π_{ps} sampling scheme : A generalization of a method of Hanurav. *Journal of the Royal Statistical Society, Series B*, 30 :556–566.
- Wolter, K. M. (1986). Some coverage error models for census data. *Journal of the American Statistical Association*, 81(394) :338–346.

Methodenberichte des Dienstes Statistische Methoden des BFS Rapports de méthodes du Service de méthodes statistiques de l'OFS Methodological reports of the Statistical Methods Unit of SFSO

- Hulliger, B., Peters, R. (1994). Enquête sur la structure des salaires en Suisse: stratégie d'échantillonnage pour le secteur privé.
- Peters, R., Hulliger, B. (1994). La technique de pondération des données: application à l'enquête suisse sur la santé.
- Schwendener, P. (1995). Verbrauchserhebung 1990 - Vertrauensintervalle.
- Hulliger, B. (1995). Konjunktuelle Mietpreiserhebung: Stichprobenplan und Schätzverfahren.
- Peters, R. (1995). Enquête de structure sur les loyers: établissement d'un plan d'échantillonnage stratifié.
- Salamin, P.A. (1995). Estimation des flux pour le module II des comptes globaux du marché de travail.
- Hulliger, B. (1996). Gütertransportstatistik 1993: Schätzverfahren mit Kompensation der Antwortausfälle.
- Hulliger, B., Peters, R. (1996). Enquête sur le comportement de la population suisse en matière de transport en 1994: plan d'échantillonnage et pondération.
- Peters, R. (1996). Schéma de pondération des indices PAUL.
- Peters, R., Hulliger, B. (1996). Schätzverfahren für die Lohnstruktur-Erhebung 1994 / Procédure d'estimation pour l'enquête de 1994 sur la structure des salaires.
- Hulliger, B. (1996). Haushalterhebung Familie 1994: Stichprobenplan, Stichprobenziehung und Reservestichproben.
- Comment, T., Hulliger, B., Ries, A. (1996). Gewichtungungsverfahren für die Schweizerische Arbeitskräfteerhebung (1991-1995).
- Peters, R. (1996). Pondération des données de l'enquête sur la famille en Suisse.
- Peters, R. (1997). Enquête 1996 sur la structure des salaires en Suisse: établissement du plan d'échantillonnage.
- Peters, R. (1997). Etablissement du plan d'échantillonnage pour l'enquête 1996 sur la recherche et le développement dans l'économie privée en Suisse.
- Salamin, P.A. (1997). Evaluation de la Statistique de l'emploi.
- Bender, A., Hulliger, B. (1997). Enquête suisse sur la population active: rapport de pondération pour 1996.
- Peters, R. (1998). Extrapolation des données de l'enquête de structure sur les loyers.
- Salamin, P.A. (1998). Etablissement d'une clef de passage pondérée entre l'ancienne (NGAE 85) et la nouvelle nomenclature (NOGA 95) générale des activités économiques.

- Hulliger, B., Kassab, M. (1998). Evaluation of Estimation Methods for the Survey on Environment Protection Expenditures of Swiss Communes.
- de Rossi, F.-X. (1998). Méthodes statistiques pour le compte routier suisse.
- Hulliger, B., Eichenberger, P. (2000). Stichprobenregister für Haushalterhebungen: Umstellung auf Telefonnummern ohne Namen und Adressen, Abläufe für Erstellung und Stichprobenziehung. Bestellnummer: 338-0000
- Renaud, A. (2001). Statistique suisse des bénéficiaires de l'aide sociale. Plan d'échantillonnage des communes. Numéro de commande: 338-0001
- Hüsler, J., Müller, S. (2001). Schlussbericht Betriebszählung 1995 (BZ 95), Mehrfach imputierte Umsatzzahlen. Bestellnummer: 338-0002
- Graf, M. (2001). Désaisonnalisation. Aspects méthodologiques et application à la statistique de l'emploi. Numéro de commande: 338-0003
- Potterat, J., Hulliger, B. (2001). Schätzung der Sägereiproduktion mit der Sägerei-Erhebung PAUL. Bestellnummer: 338-0004
- Peters, R., Renfer, J.-P. et Hulliger, B. (2001). Statistique de la valeur ajoutée 1997-1998. Procédure d'extrapolation des données. Numéro de commande: 338-0005
- Salamin, P.A. (2001). La technique de la double enquête pour la statistique du transport routier de marchandise. Numéro de commande: 338-0006
- Graf, M. (2002). Passage du concept établissement au concept entreprise. Numéro de commande: 338-0007
- Kilchmann, D., Hulliger, B. (2002). Stichprobenplan für die Obstbaumzählung 2001. Bestellnummer: 338-0008
- Renaud, A., Eichenberger P. (2002). Estimation de la couverture du recensement de la population de l'an 2000. Procédure d'enquête et plan d'échantillonnage de l'enquête de couverture. Numéro de commande: 338-0009

Publikationsprogramm BFS

Das Bundesamt für Statistik (BFS) hat – als zentrale Statistikstelle des Bundes – die Aufgabe, statistische Informationen breiten Benutzerkreisen zur Verfügung zu stellen.

Die Verbreitung der statistischen Information geschieht gegliedert nach Fachbereichen (vgl. Umschlagseite 2) und mit verschiedenen Mitteln:

Programme des publications de l'OFS

En sa qualité de service central de statistique de la Confédération, l'Office fédéral de la statistique (OFS) a pour tâche de rendre les informations statistiques accessibles à un large public.

L'information statistique est diffusée par domaine (cf. verso de la première page de couverture); elle emprunte diverses voies:

<i>Diffusionsmittel</i>	<i>Kontakt N° à composer</i>	<i>Moyen de diffusion</i>
Individuelle Auskünfte	032 713 60 11	Service de renseignements individuels
Das BFS im Internet	http://www.statistik.admin.ch	L'OFS sur Internet
Medienmitteilungen zur raschen Information der Öffentlichkeit über die neusten Ergebnisse	032 713 60 11	Communiqués de presse: information rapide concernant les résultats les plus récents
Publikationen zur vertieften Information (zum Teil auch als Diskette)	032 713 60 60	Publications: information approfondie (certaines sont disponibles sur disquette)
Online-Datenbank	032 713 60 86	Banque de données (accessible en ligne)

Nähere Angaben zu den verschiedenen Diffusionsmitteln liefert das alle 2 Jahre nachgeführte **Publikationsverzeichnis**. Es ist auf dem Internet abzurufen oder kann gratis über Tel. 032 713 60 60 bezogen werden.

La **Liste des publications**, qui est mise à jour tous les deux ans, donne davantage de détails sur les divers moyens de diffusion. Elle est proposée sur Internet et peut être obtenue gratuitement en téléphonant au n° 032 713 60 60.

Methodenberichte des Dienstes Statistische Methoden Rapports de méthodes du Service de méthodes statistiques Methodological reports of the Statistical Methods Unit

Die Methodenberichte beschreiben die mathematischen und statistischen Methoden, die den Resultaten und Analysen der öffentlichen Statistik zu Grunde liegen. Sie enthalten ausserdem die Evaluation und Entwicklung von neuen Methoden im Hinblick auf eine zukünftige Anwendung. Diese Publikationen sollen einerseits die verwendeten Methoden dokumentieren, um Transparenz und Wissenschaftlichkeit sicher zu stellen, und sie sollen andererseits die Zusammenarbeit mit Hochschule und Wissenschaft fördern.

Zur Illustration der beschriebenen mathematischen Konzepte werden im Bericht numerische Resultate aufgeführt. Diese sind allerdings nicht als offizielle Resultate der betreffenden Erhebungen zu verstehen. Ebenfalls können die tatsächlich angewendeten Methoden leicht von den hier beschriebenen abweichen.

Les rapports de méthodes décrivent les méthodes mathématiques et statistiques à la base des résultats et des analyses de la statistique publique. Ils présentent également l'évaluation et le développement de nouvelles méthodes en vue d'une application future. Ces publications visent d'une part à documenter les méthodes utilisées ou envisagées dans un souci de transparence et de rigueur scientifique, et d'autre part à favoriser la collaboration avec le monde scientifique et universitaire.

Les résultats numériques présentés dans les rapports de méthodes illustrent les concepts mathématiques décrits, mais ne sont pas des résultats officiels des enquêtes concernées. De même, les méthodes réellement appliquées peuvent différer légèrement de celles décrites dans ces rapports.

Les défauts de couverture d'un recensement de la population en Suisse sont estimés pour la première fois dans le cadre du recensement de l'an 2000 (RFP2000).

L'enquête post-censitaire nécessaire pour les estimations selon le système dual est nommée enquête de couverture (EC2000). Elle est basée sur un plan d'échantillonnage à plusieurs niveaux avec deux bases d'échantillonnage indépendantes au premier niveau (communes au Tessin et numéros postaux dans le reste de la Suisse). La méthode d'échantillonnage permet de contrôler les tailles des échantillons aux différents niveaux, tout en limitant la variabilité des poids d'échantillonnage et en tenant compte des difficultés pratiques de l'enquête sur le terrain.

Durant l'enquête, on crée la liste des ménages dans un échantillon d'environ 18'000 bâtiments. Le sous-échantillon final de 15'800 bâtiments comporte 27'000 ménages pour lesquels on recueille les informations de base du recensement. Cet échantillon de ménages permet d'atteindre le nombre désiré de personnes dans les groupes pour lesquels on désire une estimation de la couverture.