

Affaires économiques

et sociales

Manuel de vérification des recensements de la population et de l'habitation



Nations Unies

Département des affaires économiques et sociales
Division de statistique

Etudes méthodologiques

Série F n° 82

Manuel de vérification des recensements de la population et de l'habitation



**Nations Unies
New York, 2002**

NOTE

Les appellations utilisées et la présentation des données correspondantes n'impliquent de la part du Secrétariat de l'Organisation des Nations Unies aucune prise de position quant au statut juridique des pays, des territoires, villes ou zones, ou de leurs autorités, ni quant au tracé de leurs frontières ou limites.

Le terme « pays » utilisé dans la présente publication s'entend également, suivant le cas, des territoires ou zones.

Les appellations « régions développées » et « régions en développement » sont employées à des fins statistiques et n'expriment pas nécessairement une opinion quant au stade de développement de tel ou tel pays ou de telle ou telle zone.

Les cotes de documents de l'Organisation des Nations Unies se composent de lettres majuscules et de chiffres. La simple mention d'une cote dans un texte signifie qu'il s'agit d'un document de l'Organisation.

ST/ESA/STAT/SER.F/82

<http://www.un.org/Depts/unsd/>

PUBLICATION DES NATIONS UNIES

Numéro de vente : 00.XVII.9

ISBN : 92-1-261183-4

PRÉFACE

L'organisation des Nations Unies élabore depuis plusieurs années une série de manuels et de rapports techniques destinés à aider les pays à concevoir et à réaliser des recensements de la population et de l'habitation en améliorant leur qualité et leur efficacité. Ces manuels et ces rapports ont été révisés à plusieurs reprises et mis à jour de façon à tenir compte des nouvelles évolutions et des problèmes se faisant jour en matière de recensement. La présente publication fait partie d'une série de manuels qui ont été mis au point afin d'aider les pays à préparer le recensement de l'an 2000 et les recensements futurs. Parmi les autres manuels de la série figurent :

a) *Handbook on Census Management for Population and Housing Censuses*, Series F, n° 83 (publication des Nations Unies, numéro de vente : E.00.XVII.15);

b) *Handbook on Geographic Information Systems and Digital Mapping*, Series F, n° 79 (publication des Nations Unies, numéro de vente : E.00.XVII.12);

c) *Guide for the collection of Economic Characteristics in Population Censuses* (à paraître).

La publication intitulée *Principes et recommandations concernant les recensements de la population et de l'habitation, première révision* (Nations Unies, 1998) examine les avantages d'un système de contrôle et d'amélioration de la qualité, mis en place dès le début d'un processus de recensement, et dont l'importance est décisive pour le succès de l'ensemble des opérations. Elle étudie en outre le rôle d'un plan de vérification qu'il convient d'élaborer dans le cadre du programme de recensement et d'intégrer aux différents plans et procédures mis en œuvre à cet effet. Les utilisateurs du *Manuel de vérification des recensements de la population et de l'habitation* trouveront particulièrement utile de se référer à la publication *Principes et recommandations* qui fournit une grande quantité de renseignements d'ordre général concernant les procédures de vérification présentées aux chapitres III, IV et V.

La présente publication a pour objet de donner aux pays un vaste aperçu de la méthodologie de vérification des données des recensements et des enquêtes et de fournir aux responsables concernés des indications quant à la mise en œuvre des différentes approches en matière de vérification des recensements. Un autre but consiste par ailleurs à inciter les pays à conserver un historique de leurs expériences de vérification, à promouvoir les échanges de vues entre spécialistes du sujet et informaticiens et à documenter les acti-

vités menées dans le cadre du recensement ou de l'enquête du moment afin d'éviter la répétition inutile des mêmes tâches lors de la prochaine enquête ou du prochain recensement.

Le manuel passe en revue les avantages et les inconvénients des vérifications manuelles et informatisées. Pour les grands recensements, la correction manuelle des données est, d'un point de vue économique, rarement envisageable. Les conditions d'exécution de ces corrections sont généralement spécifiées dans des logiciels spécialisés qui décèlent et corrigent automatiquement les erreurs et imputent des valeurs calculées à partir d'autres renseignements sur la personne ou le ménage en cause ou bien sur d'autres personnes ou d'autres ménages. La plus grande partie du manuel traite de la correction automatique des données.

Les vérifications informatisées jouent un rôle majeur en matière de détection et de correction des erreurs. Au stade des vérifications informatisées, des contrôles de cohérence détaillés peuvent être définis en se concertant avec les spécialistes du sujet. Les erreurs décelées peuvent être corrigées soit en se référant au questionnaire d'origine, soit de façon automatique. Bien que les vérifications automatisées accélèrent le traitement, elles exigent un contrôle rigoureux de la qualité des données d'entrée.

La présente publication comprend une introduction et cinq chapitres. L'introduction décrit le processus du recensement, ainsi que les différents types d'erreurs observées dans ce cadre. Le chapitre I traite des règles fondamentales de vérification des recensements. Les chapitres II à V décrivent les procédures et techniques de vérification des données de recensement, aux différents stades de leur traitement. Enfin, des considérations techniques, concernant notamment la programmation, figurent dans les annexes.

Bien que ce manuel soit axé sur la vérification des recensements de la population et de l'habitation, les concepts et les techniques dont il est question sont pour la plupart également applicables aux enquêtes.

Nous tenons tout particulièrement à remercier le United States Bureau of the Census pour le concours apporté par M. Michael J. Levin à l'élaboration du manuel. Nous remercions également les nombreux spécialistes du sujet et programmeurs de toutes les régions du monde, dotés par ailleurs d'une vaste expérience en matière de recensement et d'enquête qui ont revu le projet de publication et contribué à sa mise au point finale.



TABLE DES MATIÈRES

	<i>Page</i>
Préface.....	iii
Abréviations.....	xv
INTRODUCTION.....	1
A. — Objet du manuel.....	1
B. — Le processus de recensement.....	1
C. — Erreurs affectant les opérations de recensement.....	2
1. Erreurs de couverture.....	2
2. Erreurs matérielles.....	2
a) Erreurs dues à la conception du questionnaire.....	2
b) Erreurs dues à l'agent recenseur.....	2
c) Erreurs dues aux personnes interrogées.....	3
d) Erreurs de codage.....	3
e) Erreurs de saisie de données.....	3
f) Erreurs dues aux vérifications automatiques.....	3
g) Erreurs d'exploitation des données.....	3
D. — Structure du manuel.....	4
I. — OPÉRATIONS DE VÉRIFICATION CONCERNANT LES RECENSEMENTS ET LES ENQUÊTES.....	5
A. — Rétrospective historique.....	5
B. — L'équipe de vérification.....	6
C. — Méthodes de vérification : données vérifiées et non vérifiées.....	6
D. — Principes de base de vérification.....	8
1. Comment la survérification est-elle préjudiciable.....	10
a) Respect des délais.....	10
b) Charges financières.....	10
c) Distorsion des valeurs vraies.....	10
d) Un sentiment erroné de sécurité.....	10
2. Traitement des données inconnues.....	10
3. Modifications parasites.....	11
4. Détermination de tolérances.....	11
5. Enseignements du processus de vérification.....	11
6. Assurance qualité.....	11
7. Coût des opérations de vérification.....	11
8. Imputation.....	12
9. Archivage.....	12
II. — MISE EN ŒUVRE DES OPÉRATIONS DE VÉRIFICATION.....	13
A. — Vérification manuelle et vérification automatisée.....	13
B. — Principes de correction des données.....	15
C. — Contrôle de validité et de cohérence.....	17
1. Méthode de vérification descendante.....	17
2. Méthode de vérification portant sur plusieurs variables.....	18
D. — Problèmes de codage.....	21

E. — Méthodes de correction et d'imputation des données	22
1. Technique de l'imputation statique ou <i>cold deck</i>	22
2. Technique d'imputation dynamique ou <i>hot deck</i>	22
3. Problèmes posés par l'imputation dynamique (méthode <i>hot deck</i>)	25
a) Considérations géographiques	25
b) Utilisation des rubriques connexes	25
c) Effet de l'ordre des variables sur les matrices	25
d) Complexité des matrices d'imputation	25
e) Etablissement de la matrice d'imputation	26
f) Matrices d'imputation normalisées	26
g) Cas de non-utilisation de la méthode d'imputation dynamique	29
h) Quelle doit être la taille des matrices d'imputation ?	30
i) Problèmes posés par une taille excessive de la matrice d'imputation	30
ii) Connaissance précise des tâches effectuées au moyen de la matrice d'imputation	30
iii) Problèmes posés par une matrice d'imputation de taille insuffisante	30
iv) Rubriques difficiles à corriger au moyen de matrices d'imputation	31
4. Contrôle des matrices d'imputation	31
a) Constitution de la matrice statique initiale	31
b) Messages relatifs aux erreurs	31
c) Etablissement à la demande de listes d'erreurs	32
d) Nombre de passages du programme de vérification	33
5. Indicateurs d'imputation	33
F. — Autres systèmes de vérification	34
III. — VÉRIFICATIONS STRUCTURELLES	36
A. — Vérifications géographiques	36
1. Emplacement des locaux d'habitation (localité)	36
2. Résidences urbaines et rurales	37
B. — Contrôles de couverture	37
1. Dénombrement de fait et de droit	37
2. Hiérarchie des ménages et des unités d'habitation	37
3. Fragments de questionnaires	37
C. — Structure des enregistrements sur l'habitation	37
D. — Correspondance entre les enregistrements relatifs à l'habitation et à la population	38
1. Habitations inoccupées et occupées	38
a) Décision de laisser une unité d'habitation inoccupée	38
b) Réexamen des unités d'habitation à plusieurs reprises afin de compléter les questionnaires	38
c) Substitution d'une autre unité de logement pour les personnes manquantes	38
2. Ménages et unités d'habitation figurant en double	38
3. Ménages manquants et unités d'habitation	39
4. Correspondance entre le nombre d'occupants et la somme des occupants	39
a) Quand le nombre d'occupants est supérieur au décompte des occupants	39
b) Contrôle du nombre de personnes par sexe	39
c) Numérotation séquentielle	39
5. Correspondance entre les occupants et le type d'immeuble/de ménage	40
E. — Enregistrements en double	40
F. — Populations spéciales	40
1. Personnes vivant en collectivité	40
a) Cas dans lesquels les enregistrements collectifs constituent un type d'enregistrement différent	40

b)	Cas dans lesquels une variable distingue les enregistrements collectifs des autres enregistrements.	40
c)	Cas dans lesquels le code « type de collectivité » est omis.	41
d)	Cas dans lesquels le code « type de collectivité » est présent, mais où il existe un lien entre toutes les personnes	41
e)	Distinction entre différents types de collectivité	41
2.	Populations sans habitat	41
a)	Migrations saisonnières.	41
b)	Personnes sans abri	41
c)	Réfugiés.	41
G. —	Détermination du chef de ménage et du conjoint.	41
1.	Détermination de la variable chef de ménage	41
a)	Ordre des liens avec le chef de ménage	41
b)	Cas dans lesquels le chef de ménage n'est pas la première personne.	42
i)	Attribution d'un pointeur à l'enregistrement du chef de ménage.	42
ii)	Désignation du chef de ménage en tant que personne n°1	42
iii)	Réattribution de codes de lien avec le chef de ménage de façon à placer celui-ci en première position	43
c)	Plusieurs chefs de ménage.	43
d)	Absence de chef de ménage	43
2.	Vérification concernant le conjoint	43
a)	Cas d'un conjoint unique dans des sociétés monogames	43
b)	Cas de plusieurs conjoints dans des sociétés monogames	44
c)	Conjoints dans des sociétés polygames.	44
H. —	Age et date de naissance	44
1.	Indication de la date de naissance, mais omission de l'âge.	44
2.	Défaut de concordance de l'âge et de la date de naissance.	44
I. —	Décompte des données non valides	44
IV. —	VÉRIFICATIONS CONCERNANT LES DONNÉES SUR LA POPULATION	45
A. —	Caractéristiques démographiques.	46
1.	Lien avec le chef de ménage (P2A).	46
a)	Vérifications du lien avec le chef de ménage	46
b)	Cas dans lesquels le chef de ménage doit figurer en premier	46
c)	Cas dans lesquels les liens avec le chef de ménage sont codés « à l'envers »	47
d)	Présence de plusieurs conjoints (polygamie)	47
e)	Présence de plusieurs parents	47
f)	Cas dans lesquels les recensements recueillent les données de lien avec le chef de ménage propres à un sexe	47
g)	Défaut de concordance entre le lien avec le chef de ménage et l'état matrimonial	47
2.	Sexe (P3A).	47
a)	Cas dans lequel le code du sexe est valide, mais où le chef de ménage et le conjoint sont du même sexe	48
b)	Cas dans lequel une donnée sur la fécondité est associée à un homme ou ne figure pas en ce qui concerne une femme adulte	48
c)	Code de sexe non valide et présence d'un conjoint.	48
d)	Non-validité du code de sexe du conjoint.	48
e)	Code du sexe non valide et présence d'indication quant au sexe féminin	48
f)	Code du sexe non valide, lorsqu'il s'agit du mari du conjoint	48
g)	Code du sexe non valide en l'absence d'indications suffisantes pour déterminer le sexe	48
h)	Remarques concernant les rapports des valeurs imputées de la variable sexe	49

3.	Date de naissance et âge (P3B)	49
	a) Age et date de naissance	49
	b) Relation entre âge et date de naissance	49
	c) Cas dans lequel l'âge calculé dépasse la valeur limite supérieure fixée	49
	d) Vérification de l'âge	50
	e) Vérification de l'âge en présence du chef de ménage et du conjoint	50
	f) Vérification de l'âge du chef de ménage en l'absence de conjoint, mais en présence d'un enfant	50
	g) Vérification de l'âge du chef de ménage en présence d'un parent de ce dernier	50
	h) Vérification de l'âge du chef de ménage en présence d'un petit-enfant de ce dernier	51
	i) Vérification de l'âge du chef de ménage en l'absence de toute autre donnée d'âge	51
	j) Vérification de l'âge du conjoint après détermination de l'âge du chef de ménage	51
	k) Vérification de l'âge des autres couples mariés du ménage, lorsque l'âge de l'une des personnes est connu	51
	l) Vérification de l'âge d'un enfant après détermination de l'âge du chef de ménage	51
	m) Vérification de l'âge d'un parent après détermination de l'âge du chef de ménage	52
	n) Vérification de l'âge d'un petit-enfant après détermination de l'âge du chef de ménage	52
	o) Vérification de l'âge de toutes les autres personnes	52
4.	Etat matrimonial (P3C)	52
	a) Vérification de l'état matrimonial	52
	b) Attribution de la variable état matrimonial sans imputation dynamique	52
	c) Attribution de l'état matrimonial avec imputation dynamique	52
	d) Le conjoint devrait être marié	53
	e) Conjoint dans un couple marié	53
	f) S'il a la qualité de conjoint, le chef de ménage doit être marié	53
	g) Chef de ménage, sans conjoint, ni enfant	53
	h) Imputation en cas d'échec de toutes les autres tentatives	53
	i) Relation entre l'âge et l'état matrimonial des jeunes	53
5.	Age au premier mariage (P4F)	53
	a) L'état matrimonial des personnes qui n'ont jamais été mariées doit être laissé en blanc	54
	b) Une valeur doit être indiquée dans le cas des personnes qui ont déjà été mariées	54
6.	Fécondité : nombre d'enfants (P4A) et nombre d'enfants vivants (P4B)	54
	a) Informations recueillies sur la fécondité	54
	b) Principes généraux applicables à la vérification des données sur la fécondité	54
	c) Relation entre le nombre total d'enfants vivants et le nombre d'enfants	55
	d) Vérification lorsque seul le nombre d'enfants est indiqué	55
	e) Vérification en cas d'indication du nombre d'enfants et du nombre d'enfants vivants	55
	f) Vérification en cas de mention du nombre d'enfants, du nombre d'enfants vivants et du nombre d'enfants décédés	56
	i) Indication des trois informations	56
	ii) Indication de deux informations	57
	iii) Indication d'une information	57
	iv) Indication d'aucune information	57
	g) Vérification en cas d'indication du nombre d'enfants, du nombre d'enfants qui vivent à la maison, du nombre d'enfants qui vivent à l'extérieur et du nombre d'enfants décédés	57
	i) Indication des quatre informations	57
	ii) Indication de trois des quatre informations	58
	iii) Indication de deux informations sur quatre	58
	iv) Indication d'une seule information	59
	v) Indication d'aucune information	59

<i>h)</i> Importance de l'utilisation d'un donneur unique pour toutes les données sur la fécondité . . .	59
7. Fécondité : date de naissance du dernier enfant né vivant (P4C)	60
8. Fécondité : âge à la première naissance (P4G)	60
9. Mortalité (P4D)	61
10. Orphelins de père ou de mère (P4E) et numéro de ligne de la mère	61
B. — Caractéristiques des migrations	61
1. Lieu de naissance (P1C)	62
<i>a)</i> Liens entre les données concernant le lieu de naissance et le nombre d'années passées dans la circonscription administrative	62
<i>b)</i> Attribution de la valeur « inconnu » aux données non valides concernant le lieu de naissance.	62
<i>c)</i> Utilisation de la méthode d'imputation statique pour le lieu de naissance	62
<i>d)</i> Utilisation de la méthode d'imputation dynamique pour le lieu de naissance	62
<i>e)</i> Attribution du lieu de naissance lorsque la mère de la personne figure parmi les membres du ménage	62
<i>f)</i> Attribution du lieu de naissance dans le cas d'un enfant du chef de ménage	63
<i>g)</i> Attribution d'un lieu de naissance à un enfant autre qu'un enfant du chef de ménage	63
<i>h)</i> Attribution d'un lieu de naissance pour des femmes adultes mariées	63
<i>i)</i> Attribution d'un lieu de naissance pour des femmes adultes non mariées	63
<i>j)</i> Attribution d'un lieu de naissance à des individus de sexe masculin	63
2. Nationalité (P3D)	63
<i>a)</i> Vérification de la nationalité	64
<i>b)</i> Relation entre groupe ethnique/race et nationalité	64
<i>c)</i> Relation entre naturalisation et nationalité	64
<i>d)</i> Rapport entre durée de résidence et nationalité	64
3. Durée de résidence (P1D)	64
<i>a)</i> Vérification de la durée de résidence	64
<i>b)</i> Résidence de fait/de droit et durée	64
<i>c)</i> Relation entre l'âge et durée de résidence	64
<i>d)</i> Relation entre lieu de naissance et durée de résidence	65
<i>e)</i> Pour les personnes qui ont toujours vécu au même endroit	65
<i>f)</i> Durée de résidence d'une personne d'après celle de la mère	65
<i>g)</i> Durée de résidence d'une personne d'après la durée de résidence d'un enfant	65
<i>h)</i> Durée de résidence d'une personne en l'absence de toute autre information disponible	65
4. Lieu de résidence antérieure (P1E)	66
<i>a)</i> Vérification du lieu de résidence antérieure	66
<i>b)</i> Lieu de résidence antérieure lorsque les frontières ont changé	66
<i>c)</i> Situation dans laquelle la personne n'a pas déménagé depuis sa naissance	66
<i>d)</i> Utilisation d'autres personnes faisant partie de l'unité d'habitation	66
<i>e)</i> Aucune autre personne appropriée pour la vérification du lieu de résidence antérieur	66
5. Lieu de résidence à une date déterminée du passé (P1F)	66
C. — Caractéristiques sociales	67
1. Aptitude à lire et à écrire (P5A)	67
2. Fréquentation scolaire (P5B)	67
<i>a)</i> Vérification de la fréquentation scolaire	67
<i>b)</i> Fréquentation scolaire à temps plein ou à temps partiel	67
<i>c)</i> Cohérence entre la fréquentation scolaire et l'activité économique	68
<i>d)</i> Attribution d'une valeur en cas de réponse non valide ou incohérente concernant la « fréquentation scolaire »	68
3. Niveau de formation (nombre d'années d'étude ou niveau d'étude atteint) [P5C]	68

a)	Vérification du niveau de formation	68
b)	Age minimum compatible avec le niveau de formation	68
c)	Relation entre âge et niveau de formation	68
4.	Domaine d'étude et diplôme de fin d'études (P5D)	68
5.	Religion (P3E)	68
a)	Vérification concernant la religion	69
b)	Indication de religion omise dans le cas du chef de ménage, mais présente pour une autre personne de l'unité d'habitation	69
c)	Identification de religion omise pour le chef de ménage ou pour toute autre personne de l'unité d'habitation	69
d)	Personne autre que le chef de ménage, sans mention de religion	69
6.	Langues (P3F)	69
a)	Vérification de la langue	69
b)	Vérification de la langue : chef de ménage	69
c)	Vérification concernant la langue : personnes autres que le chef de ménage	69
d)	Vérification concernant la langue : utilisation de l'origine ethnique ou du lieu de naissance	69
e)	Vérification concernant la langue : langue maternelle	69
f)	Vérification concernant la langue : aptitude à parler une langue déterminée	70
7.	Groupe ethnique (P3G)	70
a)	Vérification concernant l'appartenance ethnique	70
b)	Vérification concernant le groupe ethnique : pour le chef de ménage	70
c)	Vérification concernant le groupe ethnique : personnes autres que le chef de ménage	70
d)	Vérification concernant le groupe ethnique : utilisation de la langue et du lieu de naissance	70
8.	Incapacité (P8A)	71
a)	Vérification concernant les incapacités	71
b)	Incapacités multiples	71
9.	Déficiences et désavantages (P8B)	71
a)	Vérification concernant les déficiences et les désavantages	71
10.	Causes de l'incapacité (P8C)	71
a)	Vérification concernant les causes de l'incapacité	71
D. —	Caractéristiques économiques	72
1.	Situation d'activité (P6A)	72
a)	Catégories liées à la situation d'activité	73
i)	Population au chômage (P6A1)	73
ii)	A la recherche d'un emploi (P6A2)	73
iii)	Population inactive du moment (P6A3)	73
iv)	Causes de la non recherche d'emploi (P6A4)	73
b)	Vérification concernant la situation au regard de l'activité économique	74
i)	Personnes pourvues d'un emploi	74
ii)	Activité économique des personnes sans emploi	74
iii)	Activité économique des étudiants et des retraités	74
iv)	Activité économique non valide et déclaration de variables concernant l'emploi	74
v)	Activité économique non valide et déclaration de variables concernant le chômage	74
vi)	Activité économique non valide et aucune déclaration concernant les variables économiques	74
2.	Temps de travail (P6B)	74
3.	Profession (P6C)	75
4.	Activité économique (P6D)	75
5.	Situation dans l'emploi (P6E)	75
6.	Revenu (P6F)	76

7. Secteur institution (P6G).....	76
8. Lieu de travail (P6H).....	76
V. — VÉRIFICATIONS CONCERNANT L'HABITATION.....	78
A. — Sujets de base.....	79
1. Immeuble : description de l'immeuble (H01).....	79
2. Matériaux de construction des murs extérieurs (H02).....	79
3. Immeubles : année ou période de construction (H03).....	80
4. Locaux à usage d'habitation : emplacement (H04).....	80
5. Locaux à usage d'habitation : type de locaux (H05).....	80
6. Locaux à usage d'habitation : statut d'occupation (H06).....	80
7. Locaux à usage d'habitation : type de propriété (H07).....	81
8. Locaux à usage d'habitation : nombre de pièces (H08).....	81
9. Locaux à usage d'habitation : espace plancher (H09).....	81
10. Locaux à usage d'habitation : système d'alimentation en eau (H10).....	82
11. Locaux à usage d'habitation : lieux d'aisances et égouts (H11).....	82
12. Locaux à usage d'habitation : salles d'eau (H12).....	83
13. Locaux à usage d'habitation : installations destinées à la préparation des repas (H13).....	83
14. Locaux à usage d'habitation : mode d'éclairage (H14).....	83
15. Locaux à usage d'habitation : type d'évacuation des déchets solides (H15).....	84
16. Locaux à usage d'habitation : occupation par un ménage ou plus (H16).....	84
17. Locaux à usage d'habitation : nombre d'occupants (H17).....	84
18. Occupants : caractéristiques du chef de ménage (H18).....	84
19. Occupants : modalités de jouissance (H19).....	84
20. Occupants : loyer et coût de logement des propriétaires occupants (H20).....	85
B. — Sujets supplémentaires.....	85
1. Immeuble : nombre de logements (A01).....	85
2. Immeuble : ascenseur (A02).....	85
3. Bâtiments : bâtiment à usage agricole (H03).....	85
4. Matériaux utilisés pour la construction de la toiture (A04a).....	86
5. Immeuble : matériaux de construction des planchers (A04b).....	86
6. Immeuble : état d'entretien (A05).....	86
7. Locaux à usage d'habitation : nombre de chambres (A06).....	86
8. Locaux à usage d'habitation : type de combustible utilisé pour la préparation des repas (A07).....	87
9. Locaux à usage d'habitation : type de chauffage et d'énergie utilisés à cette fin (A08).....	87
10. Locaux à usage d'habitation : alimentation en eau chaude (A09).....	87
11. Locaux à usage d'habitation : gaz sous conduite (A10).....	88
12. Locaux à usage d'habitation : téléphone (A11).....	88
13. Locaux à usage d'habitation : utilisation de l'unité d'habitation (A12).....	88
14. Nombre de voitures (A13).....	88
15. Occupants : équipements ménagers (A14).....	88
16. Occupants : espaces extérieurs disponibles (A15).....	89
C. — Unités d'habitation occupées et inoccupées.....	89

ANNEXES

I. — VARIABLES CALCULÉES.....	93
A. — Variables calculées associées aux données sur l'habitation.....	93
1. Revenu du ménage.....	93
2. Revenu familial.....	93
3. Type de famille.....	94
4. Personnes apparentées.....	94

5. Travailleurs de la famille	95
6. Installations complètes de plomberie	95
7. Installation complète de préparation des aliments.	95
8. Loyer brut	95
B. — Variables calculées associées aux données sur la population.	95
1. Situation d'activité économique	95
2. Numéro de sous-famille et personnes apparentées à l'intérieur d'une sous-famille.	96
3. Enfants dont la mère est vivante	97
4. Parents vivant à la maison.	98
5. Année de scolarité du moment	98
II. — LIEN ENTRE PRÉSENTATION DU QUESTIONNAIRE ET INTRODUCTION DES DONNÉES.	99
III. — LA SAISIE MANUELLE DES DONNÉES	103
A. — Introduction des données	103
1. Lecture optique	103
2. Saisie manuelle continue.	103
a) Introduction continue sans saut de champ	104
b) Introduction continue avec sauts de champ	104
3. Saisie manuelle discontinue	104
B. — Expérimentation des instructions de saisie.	105
C. — Contrôle	105
1. Contrôle par comparaison	105
2. Contrôle indépendant.	105
IV. — EXEMPLE D'ORGANIGRAMMES.	107
V. — MÉTHODES D'IMPUTATION	113
VI. — PROGICIELS DE VÉRIFICATION	117
Glossaire	118
Références	124

TABLEAUX

	<i>Page</i>
1. Répartition d'un échantillon de population par groupe d'âge de 15 ans et par sexe, avant et après vérification des données.	7
2. Population et évolution de la population par groupe d'âge, avec mention des données inconnues : 1990 et 2000 .	8
3. Population et évolution de la population par groupe d'âge, sans mention de données inconnues : 1990 et 2000	8
 Figures	
Figure 1. Exemple de ménage fictif, avec indication du lien avec le chef de ménage, du sexe et de la fécondité de ses membres	15
Figure 2. Exemple de ménage : chef de ménage et conjoint de même sexe	16
Figure 3. Exemple de ménage avec indication des âges de certains membres	17
Figure 4. Exemple de ménage présentant des incohérences potentielles des âges déclarés	17
Figure 5. Exemple de règles de vérification à variables multiples de certaines caractéristiques démographiques .	19
Figure 6. Exemple comportant un chef de ménage et un conjoint de même sexe dans un jeu de données non vérifiées.	19
Figure 7. Exemple de spécification de vérification visant à corriger les valeurs indiquées pour la variable sexe (en langage symbolique).	20
Figure 8. Exemple d'analyses de vérification à plusieurs variables, dans le cas d'une veuve très jeune avec trois enfants.	20
Figure 9. Exemples de codes communs attribués à certaines rubriques	21
Figure 10. Exemple de ménage utilisé pour fournir les données d'entrée d'une imputation dynamique	23
Figure 11. Matrice initiale d'imputation statique de l'âge d'après le sexe et le lien avec le chef de ménage	23
Figure 12. Exemple de matrice d'imputation dynamique à la suite d'une modification	24
Figure 13. Exemple de matrice d'imputation dynamique à la suite de plusieurs modifications	24
Figure 14. Exemple de défauts d'indication de la langue du chef de ménage et du père du chef de ménage	26
Figure 15. Valeurs initiales d'une matrice d'imputation dynamique de la langue.	27
Figure 16. Exemple de membres d'un ménage dans lequel aucune langue n'a été attribuée	27
Figure 17. Exemple de données concernant un chef de ménage et un enfant avec omission de l'âge de l'enfant .	28
Figure 18. Exemple de données concernant un chef de ménage et un enfant avec omission de l'âge et du niveau de scolarité de l'enfant	28
Figure 19. Exemple de jeu de valeurs utilisées dans un tableau d'imputation statique et codes imputés correspondants	30
Figure 20. Exemple de rapport indiquant le nombre d'imputations par erreur	31
Figure 21. Exemple de rapport d'erreurs contenues dans un questionnaire.	32
Figure 22. Exemple de liste imprimée d'erreurs supplémentaires portant sur plusieurs variables	32
Figure 23. Exemple de données sur la population comportant des indicateurs associés aux valeurs imputées ...	34
Figure 24. Exemple d'indicateurs correspondant à une jeune femme dont l'enregistrement comporte des données sur la fécondité laissées en blanc, avec ajout d'un indicateur	34
Figure 25. Exemple de ménage dont le chef figure en premier sur la liste	42
Figure 26. Exemple de ménage dont le chef est classé en cinquième position	43
Figure 27. Exemple de ménage comportant des données sur la fécondité.	55
Figure 28. Valeurs initiales pour la détermination du nombre d'enfants vivants avec des données valides d'âge et de nombre d'enfants vivants.	56
Figure 29. Exemple de matrice d'imputation à définir pour des couples de valeurs connues.	59

Encadrés

1. Objectifs des opérations de vérification des recensements	6
2. Principes importants de correction des données	15
3. Principes directeurs concernant les vérifications structurelles	36
4. Vérification et imputation de l'âge	50

Figures de l'annexe

Figure A.I.1. Illustration d'une famille étendue	96
Figure A.I.2. Exemple de ménage comprenant deux sous-familles	97
Figure A.II.1. Exemple de questionnaire constitué de pages personnelles	99
Figure A.II.2. Exemple de séquence de collecte de l'information sur les pages personnelles d'un questionnaire	99
Figure A.II.3. Exemple de questionnaire, page concernant un ménage, sur laquelle figurent toutes les personnes du ménage	100
Figure A.II.4. Exemple de séquence de traitement des informations inscrites sur des pages de questionnaire concernant un ménage, avec plusieurs personnes par page	100
Figure A.II.5. Exemple de feuille concernant un ménage à plusieurs personnes, ne posant pas de problème de saisie	101
Figure A.II.6. Exemple de page concernant un ménage à plusieurs personnes, susceptible de poser des problèmes de saisie	101
Figure A.IV.1. Exemple d'organigramme de détermination du chef de ménage (CM)	108
Figure A.IV.2. Exemple d'organigramme de détermination de la présence d'un conjoint dans le ménage	109
Figure A.IV.3. Exemple d'organigramme de vérification de la variable sexe concernant le chef de ménage et le conjoint	110

Abréviations

ED	Enfants décédés
EV	Enfants vivants
EVE	Enfants vivant à l'extérieur
EVF	Enfants vivant au foyer
ICIDH	Classification internationale des handicaps, déficiences, incapacités et désavantages
NMI	Nouvelle méthodologie d'imputation
NTE	Nombre total d'enfants
OIT	Organisation internationale du Travail
OMR	Lecture optique de marques
PES	Enquête postdénombrement
ROC	Reconnaissance optique de caractères
SAS	Système d'analyse statistique
SCN	Système de comptabilité nationale
SD	Secteur de dénombrement
SIG	Systèmes d'information géographique
SPSS	Progiciel de statistiques pour les sciences sociales
UNESCO	Organisation des Nations Unies pour l'éducation, la science et la culture



INTRODUCTION

A. — Objet du manuel

1. Une enquête ou un recensement judicieusement conçu, dont les résultats comportent des erreurs réduites au minimum, est une source d'informations infiniment précieuse pour un pays. Or, l'exactitude des résultats exige l'utilisation de données exemptes dans toute la mesure possible d'erreurs et d'incohérences, en particulier à la suite de la phase de traitement des données. L'édition ou la vérification des recensements de la population et de l'habitation, désigne la procédure qui consiste à détecter erreurs et incohérences dans chaque enregistrement et d'un enregistrement à l'autre, pendant et après la phase de collecte et de saisie, et par ailleurs à ajuster les différents éléments d'information.

2. Les données des recensements ou des enquêtes ne sont jamais parfaites. Ayant constaté depuis longtemps que les données des recensements et des enquêtes étaient sujettes à caution, les pays ont mis en œuvre différentes procédures pour remédier aux lacunes et aux défauts de cohérence des réponses recueillies. Toutefois, vu l'espace important des recensements successifs, les procédures utilisées pour vérifier les données font rarement l'objet de descriptions suffisamment détaillées. De ce fait, lors de l'organisation d'un nouveau recensement ou d'une nouvelle enquête, les pays ont dû réinventer les méthodes employées à l'occasion d'activités similaires antérieures.

3. Le *Manuel de vérification des recensements de la population et de l'habitation* entend remédier à cette lacune des connaissances dans ce domaine, et fournir aux responsables concernés des indications quant à l'utilisation de différentes méthodes de vérification des recensements; l'objectif poursuivi consiste en outre à encourager les pays à conserver un historique de leurs expériences en la matière, à améliorer la communication entre les spécialistes du sujet et les informaticiens, ainsi qu'à documenter les activités menées à bien dans le cadre des recensements ou des enquêtes en cours afin d'éviter à l'avenir les répétitions inutiles de certaines tâches.

4. Ce manuel est un ouvrage de référence, pour les spécialistes du sujet¹ comme pour les informaticiens, puisqu'ils travaillent en équipes à la définition de règles et de programmes de vérification applicables aux recensements et aux enquêtes. Sa conception s'apparente à celle d'un « livre de recettes », qui permet aux pays d'adopter les vérifications les mieux adaptées à la situation statistique du moment au niveau national. La présente publication est conçue en outre pour améliorer la communication entre spécialistes du sujet

¹ Selon la définition retenue dans le présent manuel, les spécialistes du sujet désignent les personnes dont les activités sont axées sur la population, l'habitation et les domaines connexes.

et informaticiens au fur et à mesure de l'élaboration et de la mise en place de leur programme de vérification.

5. L'introduction décrit le processus du recensement, les divers types d'erreurs rencontrés dans ce cadre, et les principes fondamentaux des opérations de vérification. Les chapitres suivants décrivent les procédures et les techniques de vérification des données du recensement aux différentes étapes du traitement. Bien que ce manuel soit axé sur la vérification des recensements de la population et de l'habitation, les concepts et les techniques dont il est question sont pour la plupart également applicables aux enquêtes.

B. — Le processus de recensement

6. Un recensement de la population et/ou de l'habitation recouvre l'ensemble des opérations qui consistent à recueillir, à exploiter, à évaluer, à analyser et à diffuser des données démographiques et/ou relatives à l'habitation, économiques et sociales concernant l'ensemble des personnes et des logements (Nations Unies, 1998). Le recensement est réalisé à un moment donné, sur la totalité ou une partie bien déterminée du territoire national. En tant que tel le recensement réalise un instantané de la situation de la population ou de l'habitation à un moment donné.

7. Un recensement a essentiellement pour objet de fournir des informations sur l'effectif, la répartition et les caractéristiques de la population d'un pays. Les données du recensement servent à l'élaboration des politiques, à la planification et à l'administration, comme à la gestion et à l'évaluation des programmes dans différents domaines : éducation, main-d'œuvre, planification familiale, logement, santé, transport et développement rural. La délimitation des circonscriptions électorales et la répartition des participants aux organes directeurs en constituent une utilisation administrative fondamentale. Le recensement est en outre une source précieuse d'informations à des fins de recherche : il fournit en effet des données utilisées pour l'analyse scientifique de la composition et de la répartition de la population, comme pour la modélisation statistique à des fins de prévision de sa croissance future. Le recensement fournit par ailleurs à l'industrie et aux entreprises les données de base nécessaires à l'évaluation de la demande de logements, d'écoles, de mobilier, de nourriture, de vêtements, d'équipements de loisir, de fournitures médicales et de différents produits et services.

8. Tous les recensements et toutes les enquêtes ont en commun un certain nombre de caractéristiques importantes notamment en matière de : a) travaux préparatoires; b) dénombrement; c) traitement des données, notamment saisie (introduction au clavier), vérification et mise en tableau; d) établissement de bases de données et diffusion des résultats; e) évaluation; et f) analyse des résultats.

9. Le travail de préparation comprend de nombreuses tâches : détermination du fondement juridique du recensement; établissement du budget, élaboration du calendrier; organisation administrative; cartographie; création d'une liste des unités d'habitation; mise au point du programme d'exploitation; préparation du questionnaire et élaboration de plans et formation du personnel, pour le dénombrement, les tests préalables, le traitement de données et la diffusion des résultats.

10. Les opérations de dénombrement dépendent de la méthode choisie à cet effet, de leur calendrier et de leur durée, du niveau de contrôle et de l'utilisation éventuelle d'un échantillon. Une fois les données recueillies, elles doivent être codées, saisies, vérifiées et exploitées. Le traitement des données produit des bases de microdonnées et de macrodonnées. Les bureaux nationaux de recensement/statistique utilisent ces bases de données pour réaliser des mises en tableaux, des analyses de séries chronologiques, des graphiques et des cartes; ils font appel par ailleurs aux systèmes d'information géographique (SIG) pour les opérations de cartographie thématique, ainsi qu'à différentes techniques de diffusion. La teneur et la couverture des résultats sont évaluées par différentes méthodes, notamment l'analyse démographique et les études postdénombrement. Enfin, les résultats font l'objet de différents types d'analyses, notamment de résumés descriptifs, d'analyses opérationnelles, et d'études détaillées d'un ou plusieurs aspects de la situation démographique et sociale du pays.

C. — Erreurs affectant les opérations de recensement

11. Les données du recensement sont entachées d'erreurs dont les causes peuvent être multiples et que l'on peut classer généralement en erreurs de couverture et en erreurs matérielles.

1. ERREURS DE COUVERTURE

12. Les erreurs de couverture sont dues à des omissions ou à des doubles comptages de personnes ou d'unités d'habitation lors du dénombrement. Parmi les causes des erreurs de couverture figurent, *notamment*, le caractère inexact ou incomplet des cartes ou des listes des secteurs de dénombrement, l'absence d'inspection directe par les agents de recensement de toutes les unités des secteurs qui leur ont été assignés, les doubles comptages, l'omission des personnes qui ne souhaitent pas être recensées, le traitement erroné de certaines catégories de personnes telles que les visiteurs et les étrangers non résidents et la perte ou la destruction d'enregistrements de recensement après dénombrement. Les erreurs de couverture doivent, dans toute la mesure possible, être résolues sur le terrain. Le processus de vérification administrative supprime les données effectivement dédoublées. Toutefois, il faut prendre soin d'établir s'il s'agit d'un double enregistrement de ménages ou de personnes. Les jumeaux, par exemple, peuvent donner lieu à des informations identiques, à l'exception du numéro séquentiel. Par conséquent, les règles de vérification appliquées dans le cadre de ce processus consistent à déterminer quand accepter et quand rejeter des informations qui font apparemment dou-

ble emploi et quand introduire des modifications par imputation.

13. Les vérifications structurelles décrites au chapitre III consistent à contrôler si les ménages correspondent au nombre adéquat d'enregistrements personnels correctement classés et si des personnes ont pu être comptées deux fois.

2. ERREURS MATÉRIELLES

14. Les erreurs matérielles sont imputables à la notification ou à l'enregistrement inexact des caractéristiques des personnes, des ménages et des logements. La cause de ce type d'erreur peut être un défaut de conception des questions ou le choix d'un ordre séquentiel inapproprié ou encore un défaut de communication entre recensé et recenseur; il peut s'agir également d'erreurs de codage et de saisie, et d'erreurs commises lors des contrôles manuels et informatisés ou lors de la mise en tableaux des résultats. A chaque étape du processus il importe donc de réaliser soigneusement et de conserver des vérifications rétrospectives (appelées également analyses rétrospectives) pour veiller à ce qu'il n'y ait aucune perte de données. On trouvera dans les sections suivantes une description de chacune des erreurs mentionnées ci-dessus.

a) Erreurs dues à la conception du questionnaire

15. La mauvaise formulation des questions ou des instructions figure parmi les causes d'erreurs matérielles. Le type de questionnaire, sa présentation et sa formulation précise, ainsi que la disposition des différentes rubriques justifient que le plus grand soin soit apporté à cette tâche, en raison de l'impossibilité de remédier aux défauts de conception, une fois les opérations de recensement commencées ou terminées. Il conviendrait de procéder à des essais préalables pour réduire au minimum les erreurs potentielles dues à un défaut de conception. Si par exemple les instructions de saut de champ ne sont pas claires ou ne figurent pas au bon endroit, le recenseur risque de sauter par erreur certaines sections du questionnaire et d'omettre ainsi de recueillir toutes les informations correspondantes.

b) Erreurs dues à l'agent recenseur

16. Il existe des interactions entre recenseurs et recensés, sauf en cas d'utilisation d'un questionnaire à remplir soi-même. L'agent de recensement peut se tromper en posant les questions, en abrégant ou en modifiant leur formulation ou en s'abstenant de bien expliquer leur signification à la personne interrogée; il risque également d'introduire des erreurs en notant les réponses. La qualité des recenseurs et de leur formation contribue de façon décisive à la qualité des données recueillies; ils doivent être convenablement formés à tous les aspects des procédures de recensement. Il faut par ailleurs les amener à comprendre l'importance de leur tâche et comment le dénombrement proprement dit s'inscrit parmi les différentes étapes du recensement. Qui plus est, puisque les agents recenseurs proviennent d'horizons très différents et que leur niveau d'instruction est plus ou moins élevé, il faut assurer une formation propre à garan-

tir qu'ils savent comment poser les questions pour obtenir une réponse appropriée.

c) *Erreurs dues aux personnes interrogées*

17. Des erreurs risquent également d'être introduites dans les données lorsque les personnes interrogées ne comprennent pas certaines questions; elles peuvent aussi résulter d'un enregistrement délibérément erroné ou de réponses approximatives (lorsque quelqu'un d'autre que la personne à laquelle les informations se rapportent fournit les réponses au questionnaire). Il est possible d'améliorer la qualité des réponses individuelles en assurant la publicité du recensement, ainsi que la formation des recenseurs de telle sorte qu'ils puissent expliquer l'objet du recensement et justifier les différentes questions. Certains pays utilisent des questionnaires à remplir soi-même pour supprimer les interactions entre recenseur et recensé. Dans le cas des questionnaires à remplir soi-même, les erreurs sont dues au défaut de compréhension des questions et des instructions par les personnes interrogées.

18. C'est au stade du dénombrement qu'il est le plus facile de remédier aux erreurs imputables aux personnes interrogées et à l'agent de recensement, lorsque formulaires, recensés et recenseur sont encore accessibles. Le personnel d'encadrement doit être en mesure de former les agents de recensement; il doit également pouvoir vérifier régulièrement les données recueillies par ces derniers pendant les opérations de dénombrement, pour assurer qu'aucun biais systématique n'est introduit dans les données. Enfin, les chefs d'équipe doivent corriger sur le terrain les erreurs des agents recenseurs et des personnes interrogées, avant que les questionnaires ne soient envoyés aux bureaux régionaux ou centraux.

d) *Erreurs de codage*

19. Des erreurs peuvent se produire dans le cadre des opérations de codage, la personne affectée à cette tâche étant susceptible de coder l'information de façon erronée. Les erreurs de frappe peuvent être à l'origine d'erreurs lors de l'introduction des données. En règle générale, l'absence de contrôle et de vérification à ce stade retarde la diffusion des données, la détection et la correction des erreurs étant plus délicates par la suite. Les vérifications manuelles interviennent souvent avant ou pendant le codage.

e) *Erreurs de saisie de données*

20. Des contrôles de vraisemblance et certains contrôles de cohérence peuvent être intégrés au logiciel de saisie de données afin d'éviter l'introduction de données incorrectes. Un système intelligent de saisie des données veille à ce que la valeur affectée à chaque champ ou à chaque élément de donnée soit dans le domaine correspondant de valeurs admissibles. Ce système a pour effet de rendre d'autant plus probable l'introduction de données plausibles par l'opérateur de saisie et d'alléger dans une certaine mesure les tâches de vérification aux stades ultérieurs du processus de mise en forme des données. Ces contrôles risquent cependant de ra-

lentir la saisie. Aussi importe-t-il d'évaluer soigneusement l'importance des contrôles de cohérence à ce stade au regard de la nécessité de maintenir une rapidité adéquate. Il faut donc à cet égard définir a priori un compromis, pour que les opérateurs de saisie ne consacrent pas un temps excessif aux tâches de contrôle. La vérification des données introduites au clavier améliore nécessairement leur qualité. Une deuxième introduction au clavier des mêmes informations, souvent par sondage, permet de vérifier les formulaires saisis.

f) *Erreurs dues aux vérifications automatiques*

21. Les vérifications représentent une des étapes décisives du traitement des données du recensement. Le processus de vérification modifie ou corrige les données non valides ou présentant un défaut de cohérence, en imputant des données plausibles aux réponses manquantes ou incohérentes. Paradoxalement, l'une ou l'autre de ces opérations risque d'introduire de nouvelles erreurs.

g) *Erreurs d'exploitation des données*

22. Des erreurs peuvent se produire au stade de la mise en tableaux des données, du fait d'erreurs de traitement ou de l'utilisation d'informations « inconnues » (non spécifiées). Il est difficile de procéder alors à des corrections sans introduire de nouvelles erreurs. Enfin, l'absence de vérifications croisées entre les différents tableaux, et les erreurs d'impression entraînent des erreurs au stade de la publication. Au lieu de chercher à corriger les tableaux proprement dits, il importe de mettre à jour le système de traitement pour effectuer des vérifications supplémentaires dès que des défauts de cohérence apparaissent dans les tableaux. En cas de report des erreurs à tous les stades du processus jusqu'à celui de la publication, celles-ci seront alors manifestes et la valeur des résultats obtenus sera douteuse. Avant de publier des tableaux, il est donc essentiel de procéder à un contrôle rigoureux pour garantir que toutes les tâches d'exploitation de données sont définies pour toutes les unités géographiques prévues. Tandis que les contrôles de vraisemblance et de cohérence réalisés au stade de la vérification permettent de corriger la plus grande partie des erreurs, un contrôle général après dépouillement joue un rôle essentiel. Il y a lieu de confier à des personnes dûment formées et expérimentées l'examen des divers tableaux obtenus, afin de vérifier si les chiffres figurant dans les différentes cases sont compatibles avec la situation locale connue. Dans un nombre limité de cas, un contrôle rapide des feuilles de recensement peut faire apparaître des erreurs de codage. Le calcul de certains rapports et taux d'accroissement et la comparaison avec les chiffres des recensements précédents ou avec d'autres données numériques publiées dans le cadre d'enquêtes par sondage peuvent être également utiles. Toutefois, la comparaison avec les chiffres fondés sur d'autres enquêtes ne doit être entreprise que si les concepts utilisés sont comparables. Si des erreurs sont constatées dans les tableaux de résultats, il convient de corriger en premier lieu les jeux de données utilisés.

23. Il est particulièrement important que les informaticiens évitent de modifier les programmes de tabulation pour corriger les problèmes constatés dans les jeux de données. Ces modifications n'apparaîtront pas dans les microdonnées et ne pourront donc être répétées lors de l'élaboration et de l'exécution d'autres programmes. L'équipe responsable du traitement des données doit apporter toutes les modifications utiles au jeu de microdonnées, notamment pour permettre à d'autres informaticiens du bureau national de recensement/statistique d'établir des tableaux comparables. En outre, puisque les bureaux nationaux de recensement/statistique diffusent parfois une partie des fichiers de microdonnées à des chercheurs et à d'autres utilisateurs du secteur privé et du secteur public, la reproductibilité des tableaux doit être assurée.

24. Tel qu'indiqué ci-dessus, le recensement se compose d'un certain nombre d'opérations successives interdépendantes, dont chacune peut comporter des erreurs. Il importe de ne pas perdre de vue que les vérifications automatisées font partie d'un système d'information en retour et qu'elles influent non seulement sur les opérations en aval de mise en tableaux, mais aussi en amont sur les opérations de collecte et de traitement sur le terrain. Aussi l'intensification des vérifications sur le terrain est-elle la meilleure façon pour un bureau national de recensement/bureau de statis-

tique d'éviter les problèmes liés aux vérifications automatisées. Le bureau national de recensement/bureau de statistique doit en outre veiller à l'exactitude du codage et de la saisie des données, ainsi qu'à l'obtention en permanence d'informations en retour concernant toutes les opérations, notamment de saisie, de vérification et de dépouillement.

D. — Structure du manuel

25. Le chapitre I est consacré au rôle des opérations de vérification dans le cadre des recensements et des enquêtes. Les autres chapitres traitent de sujets spécifiques. Le chapitre II décrit la mise en œuvre concrète des opérations de vérification et d'imputation. Le chapitre III décrit les vérifications de structure, lesquelles concernent simultanément les informations sur l'habitation et sur la population, ainsi que certaines procédures visant à faciliter les autres vérifications, consistant par exemple à déterminer la présence éventuelle d'un seul chef de ménage. Le chapitre IV traite des vérifications touchant à la population et le chapitre V des vérifications qui se rapportent à l'habitation. Enfin, une série d'annexes passe en revue des questions particulières liées à la vérification des données des recensements de la population et de l'habitation et aux opérations d'imputation nécessaires.

I. — OPÉRATIONS DE VÉRIFICATION CONCERNANT LES RECENSEMENTS ET LES ENQUÊTES

A. — Rétrospective historique

26. Avant l'arrivée des ordinateurs, la plupart des recensements comportaient l'embauche d'un grand nombre d'employés semi-qualifiés chargés de vérifier chaque bulletin. Toutefois, en raison de la complexité des relations entre un nombre même relativement réduit de rubriques, de simples contrôles ne pouvaient laisser escompter la détection de tous les défauts de cohérence possibles. En effet, des employés différents interprétaient vraisemblablement les règles de différentes façons et tel ou tel employé pouvait même ne pas être parfaitement cohérent.

27. L'apparition des ordinateurs a profondément modifié les opérations de vérification des recensements. Les ordinateurs ont permis de déceler des défauts de cohérence bien plus nombreux que les vérifications manuelles. Les spécifications de vérification sont devenues de plus en plus complexes et évoluées. Les imputations automatisées ont alors été possibles et différentes règles ont été définies à cet effet (Nordbotten, 1963; Naus, 1975). Simultanément, le processus opératoire autorisait des contacts de plus en plus nombreux avec les personnes interrogées ou du moins une multiplication des rapprochements avec les questionnaires dûment complétés. De nombreuses équipes de vérificateurs ont commencé à penser que la qualité du recensement était directement fonction de l'ampleur du travail de vérification et que les résultats étaient d'autant plus exacts que ce même travail était plus sophistiqué. Les programmes ont donc généré des milliers de messages d'erreur, exigeant un examen manuel des questionnaires d'origine ou pour certaines enquêtes de nouvelles entrevues avec les personnes interrogées.

28. Grâce aux ordinateurs, il est devenu de plus en plus facile d'introduire des modifications dans le jeu de données. Parfois ces modifications ont eu pour effet de corriger certains enregistrements ou certaines rubriques. Nombre d'enregistrements faisaient l'objet de plusieurs traitements, les erreurs et les défauts de cohérence étant examinés à chaque fois par différentes personnes (Boucher, 1991; Granquist, 1997).

29. Toute cette évolution a conduit à l'élaboration de plusieurs programmes généraux de vérification des recensements, dont certains sont encore utilisés aujourd'hui. A l'origine les programmes ont été conçus pour de gros ordinateurs : certains ont par la suite été modifiés pour être utilisés sur des ordinateurs personnels. Au cours de cette période, Fellegi et Holt (1976) ont mis au point une nouvelle méthode générale de vérification et d'imputation, qui n'a pas été immédiatement appliquée, mais qui est actuellement adoptée de plus en plus fréquemment, à la faveur de la com-

plexification des opérations de vérification effectuées par les bureaux nationaux de recensement/statistique.

30. Un progrès majeur en matière de vérification des recensements est apparu dans les années 80 quand les bureaux nationaux de recensement/statistique ont commencé à utiliser les ordinateurs personnels pour saisir, vérifier et exploiter leurs données. Rapidement, le personnel chargé du traitement des données a pu effectuer des vérifications en direct, au stade de la saisie des données ou peu de temps après. Pour les enquêtes, les informaticiens ont été en mesure d'élaborer des programmes de détection des erreurs au cours de la collecte des données ou lors de la saisie directe sur ordinateur. Ainsi, l'informatisation des vérifications a permis de développer des contacts continus avec les personnes interrogées de façon à résoudre les problèmes rencontrés à l'occasion des contrôles effectués (Pierzchala, 1995).

31. Au cours des premières années, la réalisation de vérifications de plus en plus sophistiquées et de plus en plus complètes sur les données des recensements et des enquêtes a donné apparemment des résultats très satisfaisants. Les équipes de vérification ont défini des spécifications de vérification de plus en plus élaborées, et les spécialistes du traitement informatique ont passé des mois à mettre au point des organigrammes et des programmes de calcul. Les progiciels en question ont rarement été évalués par des analystes. Les programmes de vérification avaient apparemment la capacité de remédier à tous les problèmes dus aux phases antérieures de collecte, de codage et de saisie des données. Toutefois, il est également apparu à nombre d'analystes que, dans de nombreux cas, toutes ces opérations supplémentaires de vérification avaient pour conséquence d'altérer les données ou du moins de retarder les résultats ou encore, d'y introduire une distorsion. Parfois, les passages en machine étaient si nombreux pour corriger dans un premier temps un élément d'information, puis un autre, que les résultats n'étaient plus comparables à ceux obtenus à partir des données initiales non vérifiées.

32. Pour nombre de recensements et pour les grandes enquêtes, ces vérifications aussi poussées entraînaient des retards considérables. Les commis passaient beaucoup de temps à rechercher manuellement des questionnaires, tandis que les informaticiens poursuivaient la mise au point d'applications destinées à examiner un très petit nombre de cas. Or, comme le signale Granquist (1997), nombre d'études ont établi qu'une grande partie de ce travail supplémentaire correspondait « à des améliorations qualitatives marginales, inexistantes ou même négatives; de fait, de nombreux types d'erreurs systématiques importantes ne peuvent être mises en évidence par les opérations de vérification ».

33. Puisque les bureaux nationaux de recensement et les organismes de sondage continuent à préparer des recensements et des enquêtes, la réalisation de vérifications automatisées approfondies est possible et même probable. La question suivante se pose donc aux bureaux nationaux de recensement/statistique : quel est le niveau approprié de vérification automatisée ?

B. — L'équipe de vérification

34. Lorsqu'ils préparent un recensement, les bureaux nationaux de statistique doivent envisager d'introduire différents types d'améliorations potentielles dans la qualité de leur travail. Parmi ces dernières figurent la création d'un groupe de vérification. Ainsi, les opérations de vérification doivent être confiées à une équipe de responsables du recensement, de spécialistes du sujet et d'informaticiens. Cette équipe devrait être formée dès le début des préparatifs en vue du recensement, de préférence au cours de la mise au point du questionnaire. L'équipe de vérification joue un rôle important dès le début et conserve son rôle tout au long du processus de vérification. Le soin apporté à la création de l'équipe ainsi qu'à la définition et à la mise en place des règles de vérification et d'imputation garantit pour le recensement une rapidité et une efficacité accrues.

35. Les réunions entre les fonctionnaires chargés du recensement et la communauté des utilisateurs en ce qui concerne les tableaux de résultats et les autres produits peuvent apporter des points de vues intéressants quant aux vérifications nécessaires. Fréquemment, les utilisateurs demandent un tableau ou un type de tableau particulier dont la réalisation exige des vérifications supplémentaires afin d'en éliminer les incohérences potentielles. L'équipe de vérification doit donc prévoir de créer ces tableaux pendant la période initiale de vérification, plutôt que de les créer en tant que tableaux spéciaux, une fois terminé le traitement du recensement. Le fait de définir les règles et les programmes informatiques de vérification dans le cadre d'un essai préalable ou d'une sorte de répétition générale, permet de tester les logiciels proprement dits et de raccourcir les délais d'exécution des différentes phases du processus de vérification et d'imputation. L'équipe de vérification évalue alors l'incidence de ces différents processus et prend les mesures correctives éventuellement nécessaires.

36. Spécialistes du sujet et informaticiens doivent collaborer à la définition des règles de vérification et d'imputation. L'équipe de vérification élabore un programme de recherche et de correction des erreurs au début des préparatifs du recensement. Elle établit ensuite par écrit des séries de règles de contrôle et de vérification de cohérence.

37. Spécialistes du sujet et informaticiens doivent collaborer non seulement lors de la définition des règles de vérification et d'imputation, mais aussi à tous les stades du recensement ou de l'enquête, notamment au stade de l'analyse. Les vérifications excessives comportent autant de risques que les vérifications insuffisantes et que la présence d'informations non vérifiées ou incohérentes dans le jeu de données. Par conséquent, les deux groupes de spécialistes doivent prendre la responsabilité d'entretenir correctement

leurs bases de métadonnées. L'équipe de vérification doit en outre utiliser efficacement les sources administratives et les données d'enquête disponibles pour améliorer la qualité des recensements ou des enquêtes ultérieures.

38. Les échanges entre spécialistes du sujet et informaticiens étaient limités lorsque les bureaux nationaux de statistique/de recensement utilisaient de gros ordinateurs; ils le sont restés pendant un certain temps après l'arrivée des micro-ordinateurs, mais les logiciels informatiques sont devenus plus faciles à utiliser et à présent, nombre de spécialistes du sujet sont effectivement en mesure de mettre au point et d'expérimenter leurs propres programmes de tabulation et leurs propres vérifications. Sans généralement procéder eux-mêmes au traitement proprement dit des données, les spécialistes du sujet sont souvent bien au fait des étapes du traitement des données mis au point par les informaticiens.

C. — Méthodes de vérification : données vérifiées et non vérifiées

39. Les pays procèdent à des vérifications des données de recensement afin d'en améliorer la qualité et la présentation. Dans ce chapitre, le manuel attire l'attention sur un problème qui se pose aux bureaux nationaux de recensement/statistique lors de la publication de données de recensement non vérifiées. Le type de problèmes rencontrés est illustré au moyen d'un jeu de données fictif.

ENCADRÉ 1. Objectifs des opérations de vérification des recensements

La correction des recensements s'attache aux objectifs suivants :

- 1) Fournir aux utilisateurs des données de recensement de qualité supérieure;
- 2) Identifier les types et les causes d'erreurs;
- 3) Fournir des résultats de recensement dûment rectifiés.

40. Le bureau national de recensement/statistique d'un pays fictif doit relever le défi consistant à vouloir répondre aux besoins de multiples utilisateurs. Certains d'entre eux peuvent souhaiter la prise en compte, à des fins d'analyse ou de recherche, des valeurs dites « inconnues », tandis que d'autres souhaitent le cas échéant des données comportant un bruit (erreur possible) réduit au minimum pour leurs propres besoins de planification ou d'action. Si le bureau national de recensement/statistique diffuse un tableau non corrigé, comme celui de la moitié gauche du tableau 1, alors analystes et responsables devront faire des hypothèses lorsqu'ils utiliseront ces données. Le tableau 1 illustre ce problème uniquement dans le cas d'un petit nombre de personnes. Dans le pays en question, les données enregistrées comportent une omission du sexe pour 23 personnes, et de l'âge pour 15 personnes; ces omissions sont probablement imputables à des non-réponses ou à des erreurs d'introduction au clavier; en particulier, il y a eu dans deux cas omission conjointe du sexe et de l'âge.

TABLEAU 1. Répartition d'un échantillon de population par groupe d'âge de 15 ans et par sexe, avant et après vérification des données

Groupe d'âge	Données non vérifiées				Données vérifiées		
	Total	Hommes	Femmes	Non indiqué	Total	Hommes	Femmes
Total	4 147	2 033	2 091	23	4 147	2 045	2 102
Moins de 15 ans	1 639	799	825	15	1 743	855	888
15 à 29 ans	1 256	612	643	1	1 217	603	614
30 à 44 ans	727	356	369	2	695	338	357
45 à 59 ans	360	194	166	0	341	182	159
60 à 74 ans	116	54	59	3	114	53	61
75 ans et plus	34	12	22	0	37	14	23
Non indiqué	15	6	7	2			

41. La plupart des utilisateurs prennent généralement leur propre décision quant au traitement des valeurs inconnues. Une solution logique, parfois naïve, consisterait à répartir les valeurs inconnues dans la même proportion que les valeurs connues. Si le bureau national de recensement/statistique décide d'imputer les données inconnues, l'équipe de vérification peut retenir un effectif de 12 hommes et de 11 femmes, soit une répartition dans des proportions sensiblement égales, mais introduisant un biais, puisque le recensement a dénombré plus de femmes que d'hommes. Les résultats seront alors cohérents avec les données corrigées indiquées dans la partie droite du tableau 1.

42. Il existe d'autres options en matière de traitement des données inconnues. Par exemple, l'équipe de vérification peut décider d'introduire des corrections fondées sur la seule répartition par sexe, en ignorant toutes les autres informations disponibles, telles que la relation entre les conjoints, le fait de savoir si une personne de sexe non spécifié est signalée comme étant la mère d'une autre personne ou si, pour une personne de sexe non spécifié, il y a une indication de nombre d'enfants nés. Une autre stratégie d'imputation consisterait à prendre en compte une ou plusieurs de ces variables.

43. Le bureau national de recensement/statistique pourrait faire appel à une autre solution consistant à décider de fonder l'imputation sur la répartition par âge. Pour l'échantillon de population du tableau 1, l'âge est omis dans 15 cas au total. Ces données pourraient également être réparties dans les mêmes proportions que les valeurs connues, ce qui constitue là encore, une stratégie d'imputation logique. Toutefois, l'équipe de vérification pourrait probablement obtenir de meilleurs résultats en considérant d'autres variables et certaines combinaisons, par exemple les âges comparés du mari et de la femme, des parents et des enfants, ou des grands-parents et des petits-enfants, ou encore la présence d'enfants d'âge scolaire, de retraités et de personnes de la population active.

44. Au tableau 1, les données corrigées de la partie droite sont « plus pures », puisque les données inconnues en

ont été éliminées (voir données figurant sous la rubrique « données corrigées »). Les données inconnues ont disparu de cette partie du tableau, puisqu'elles ont été réparties par le programme dans des colonnes correspondant à différentes réponses. Toutefois, nombre de démographes et de spécialistes ont traditionnellement exprimé le vœu de voir figurer des données inconnues dans tous les tableaux, comme au tableau 1. Selon eux, cette façon de procéder leur permet d'effectuer différents types d'évaluations sur les chiffres présentés afin de mesurer l'efficacité des méthodes de recensement ou de faciliter la préparation des enquêtes et des recensements futurs. Or, les deux objectifs peuvent être atteints — un tableau corrigé destiné aux utilisateurs proprement dits et un tableau non corrigé présenté pour appréciation — en créant des tableaux avec et sans mention des données inconnues.

45. La prise en compte des données inconnues dans les tableaux publiés soulève une autre difficulté; ces données peuvent en effet avoir une incidence quant à l'analyse des tendances. Les nouvelles techniques disponibles facilitent considérablement ce type d'analyse par rapport à la situation passée. Par exemple, le tableau 2 présente une répartition par âge telle qu'elle ressort de deux recensements consécutifs. Le nombre de données inconnues a diminué dans le cas de ce petit pays, passant de 217, soit quelque 6,5 % des réponses rapportées en 1990, à seulement 15, soit moins de 1 % des réponses en 2000.

46. A cet égard, le bureau national de recensement/statistique doit déterminer dans quelle mesure des effectifs incohérents de données inconnues affectent le recensement considéré et l'évolution d'un recensement à l'autre. Par exemple, le chiffre de 6,5 % de données inconnues pour le recensement de 1990 rend difficile la comparaison des répartitions en pourcentage par groupe d'âge de 15 ans dans les résultats des deux recensements. Ainsi, le pourcentage des personnes de 15 à 29 ans augmente apparemment en passant de 27 % seulement à 30 % au cours de la décennie; or, la répartition des données inconnues a pu affecter les conclusions de l'analyse.

TABLEAU 2. Population et évolution de la population par groupe d'âge, avec mention des données inconnues : 1990 et 2000

Groupe d'âge	Effectifs		Variation	Variation en pourcentage	Pourcentage	
	2000	1990			2000	1990
Total	4 147	3 319	828	24,9	100,0	100,0
Moins de 15 ans	1 639	1 348	291	21,6	39,5	40,6
15 à 29 ans	1 256	902	354	39,2	30,3	27,2
30 à 44 ans	727	538	189	35,1	17,5	16,2
45 à 59 ans	360	200	160	80,0	8,7	6,0
60 à 74 ans	116	89	27	30,3	2,8	2,7
75 ans et plus	34	25	9	36,0	0,8	0,8
Non indiqué	15	217	- 202	- 93,1	0,4	6,5

TABLEAU 3. Population et évolution de la population par groupe d'âge, sans mention des données inconnues : 1990 et 2000

Groupe d'âge	Effectifs		Variation	Variation en pourcentage	Pourcentage	
	2000	1990			2000	1990
Total	4 147	3 319	828	24,9	100,0	100,0
Moins de 15 ans	1 743	1 408	335	23,8	42,0	42,4
15 à 29 ans	1 217	952	265	27,8	29,3	28,7
30 à 44 ans	695	578	117	20,2	16,8	17,4
45 à 59 ans	341	230	111	48,3	8,2	6,9
60 à 74 ans	114	109	5	4,6	2,7	3,3
75 ans et plus	37	42	- 5	- 11,9	0,9	1,3

47. Le tableau modifié (tableau 3) est obtenu après répartition des données inconnues, soit proportionnellement, soit par une autre méthode d'imputation. Les variations numériques et en pourcentage apparaissent alors plus nettement, ainsi que la répartition des groupes d'âge dans les deux recensements. Evidemment, pour obtenir des résultats précis et fiables, les équipes de vérification doivent veiller à la cohérence des corrections introduites dans les résultats des deux recensements et/ou des enquêtes, comme à la cohérence interne. La ligne correspondant aux données non indiquées a disparu.

D. — Principes de base de vérification

48. On entend par vérification l'examen et la correction (ou la modification) systématiques des réponses en fonction de règles définies au préalable. Certaines opérations de vérification comportent des corrections manuelles, c'est-à-dire des corrections introduites par des agents, par opposition à d'autres opérations de vérification comportant des corrections électroniques réalisées au moyen d'ordinateurs. Les publications de recensements risquent de contenir une certaine quantité d'informations dénuées de sens si les bureaux nationaux de recensement/statistique omettent de vérifier

les résultats des enquêtes ou des recensements. Cette tâche a pour effet de limiter le nombre des estimations faussées, de faciliter le traitement des données et de renforcer la confiance de l'utilisateur à l'égard des résultats. En outre, de l'avis de Pullum, Harpham et Ozsever (1986), « les opérations de vérification ou d'élimination des erreurs ont essentiellement pour objet de déterminer si les différentes réponses sont cohérentes les unes avec les autres et avec le format de base utilisé pour l'enquête ».

49. Les fichiers de données brutes d'un recensement contiennent des erreurs de nature variée. Il existe deux types d'erreurs du point de vue du traitement des données : celles qui risquent de faire obstacle à la poursuite du traitement et celles qui conduisent à des résultats non valides ou incohérents, sans nécessairement interrompre la séquence logique des traitements ultérieurs. Tel qu'indiqué dans *Principes et recommandations concernant les recensements de la population et de l'habitat, première révision* [Nations Unies, 1998, par. 195]. Les erreurs du premier type doivent toutes être corrigées et celles du second doivent l'être dans la mesure du possible. La vérification des recensements au stade du dépouillement a donc essentiellement pour objet d'iden-

tifier le plus grand nombre possible d'erreurs et de modifier en conséquence le jeu de données, pour assurer la validité et la cohérence des informations recueillies. Néanmoins, les traitements réalisés ne peuvent corriger toutes les erreurs de recensement, notamment les réponses aux questionnaires qui présentent une cohérence interne, mais qui constituent en fait des erreurs de déclaration de la part des personnes interrogées ou des erreurs d'enregistrement de la part des agents recenseurs.

50. Les vérifications relèvent généralement de deux catégories : 1) les **vérifications fatales** qui détectent des erreurs avec certitude; et 2) les **vérifications avec interrogation** qui attirent l'attention sur des données suspectes (Granquist et Kovar, 1997 : 420). Les erreurs fatales, détectées par les vérifications du même nom, comprennent les entrées non valides ou manquantes, ainsi que les erreurs dues à des incohérences. En revanche, les vérifications avec interrogation détectent les données qui se situent en dehors de limites de vérification essentiellement subjectives, les données dont la valeur est relativement élevée ou faible par rapport aux autres données du même questionnaire et différentes données douteuses. Pour préserver la fiabilité du recensement, et en particulier lorsque le bureau national de recensement/statistique décide de diffuser des microdonnées, le processus de vérification doit détecter les erreurs fatales et effectuer les corrections correspondantes. Les vérifications avec interrogation se traduisent difficilement par des corrections, et comportent moins d'avantages que la détection et la correction des erreurs fatales, enfin, elles augmentent le coût de l'ensemble des opérations de dépouillement.

51. Puisque dans un recensement toutes les données recueillies sont spécifiquement prises en compte en fonction des besoins des planificateurs et des dirigeants, les vérifications avec interrogation doivent impérativement, lors des opérations de vérification et d'imputation, être résolues dans une plus forte proportion, par comparaison aux enquêtes. Néanmoins, afin de déterminer les vérifications finales dont un recensement doit faire l'objet, les démographes doivent étudier les vérifications mises au point pour des recensements pilotes et celles élaborées à l'occasion du dépouillement, de manière à s'assurer que chaque vérification envisagée se caractérise par les coûts et les avantages escomptés. Les analyses de ce type font nécessairement partie de l'évaluation du recensement. Comme le font observer Granquist et Kovar (1997, p. 422), les données concernant « les taux de succès, c'est-à-dire la fraction du nombre d'indications qui entraînent des modifications des données de départ, sont rarement mentionnées dans les évaluations ou les études consacrées aux processus de vérification ».

52. Une autre série de techniques et d'expressions se rapportent aux microvérifications et aux macrovérifications. Tel qu'indiqué, les vérifications relatives aux recensements et aux enquêtes détectent les erreurs et incohérences à l'intérieur des enregistrements et d'un enregistrement à l'autre. Le présent manuel décrit les opérations de microvérification, c'est-à-dire les moyens utilisés pour garantir la validité et la cohérence des enregistrements individuels et des relations entre les enregistrements relatifs à un ménage. Une

autre méthode intitulée macrovérification a pour objet de contrôler des données agrégées et de vérifier qu'elles présentent également un caractère plausible. Par exemple, un pays donné peut avoir un pourcentage très important de personnes dont l'âge n'est pas indiqué. Après imputation de valeurs de l'âge afin d'obtenir un jeu complet, les contrôles effectués au niveau des macrodonnées, au niveau agrégé, permettraient de s'assurer que des omissions sélectives de la part des personnes d'un certain âge n'entraînent pas un biais dans les valeurs imputées. Selon les résultats de l'analyse, l'équipe de vérification pourrait convenir de prendre des mesures permettant d'atténuer le risque de biais potentiel.

53. Les opérations de vérification doivent autant que possible protéger la qualité des données initiales. L'équipe de vérification doit non seulement veiller à la qualité des données et à l'élimination des erreurs qu'elles contiennent, mais doit en outre protéger le travail effectué sur le terrain par l'administration. Les données initiales doivent en effet être protégées à tous les stades du traitement informatique, si l'équipe de vérification décide qu'il faut réexaminer le processus de correction. Les données d'origine sont réexaminées lorsque l'équipe détecte une erreur systématique dans le processus de vérification, ou bien lorsqu'on constate l'absence ou le doublement d'une partie du jeu de données, dont il faut recommencer la constitution et la vérification.

54. Parfois, la cause des erreurs est extérieure au service chargé du dépouillement. Banister (1980, 2) fait observer que, « lorsque nous savons qu'une forte proportion des personnes d'un sous-groupe particulier n'ont pas répondu à une question donnée du recensement, cela signifie qu'elles n'ont pas bien compris la question, qu'elles n'ont pas voulu y répondre, ou qu'elles n'ont pas eu envie de coopérer aux opérations de recensement ». Par conséquent, les taux de non-réponse relatifs aux sous-groupes devraient figurer selon elle dans les données enregistrées au titre du recensement et dans les tableaux publiés. Les bureaux nationaux de recensement/statistique ont à présent davantage de possibilités pour conserver ces données sur disque compact ou sur d'autres supports pour les besoins des chercheurs.

55. Il est de plus en plus évident que l'importance plus ou moins grande des opérations de vérification automatique ne saurait se substituer au soin particulier apporté à la collecte des données du recensement. Les bureaux nationaux de recensement/statistique n'ignorent pas qu'à partir d'un certain stade les vérifications automatiques sont non seulement limitées, mais deviennent inefficaces : le processus de vérification ajoute plus d'erreurs au jeu de données qu'il n'en corrige. Le fait de modifier une rubrique de recensement et celui d'y apporter des corrections ne sont nullement équivalents. Par conséquent, l'équipe de vérification doit se concerter pour définir la portée et les limites précises de son action.

56. Les vérifications et les imputations peuvent certes améliorer ou non la qualité des données, mais l'utilisation d'un jeu de données dont les erreurs ont été éliminées facilite considérablement l'analyse. Le processus commence par la conception même du questionnaire du recensement.

Les démographes et les autres spécialistes du sujet déterminent généralement sa teneur, souvent en consultation avec des groupes d'utilisateurs. En définitive, les données des recensements ne sont pas élaborées « en priorité à destination de démographes puristes, mais pour un public beaucoup plus vaste d'universitaires, de responsables et de non-spécialistes » (Banister, 1980, p. 17). Il est néanmoins essentiel de réaliser un recensement dont les données invalides et incohérentes ont été éliminées, lorsque la crédibilité du bureau national de recensement/statistique est en jeu. Comme le signale Banister, « les services de recensement peuvent citer des cas d'articles humoristiques parus dans la presse ou de lettres indignées adressées aux responsables du recensement par des citoyens, au sujet de tableaux statistiques faisant état de grands-pères âgés de 3 ans, et d'usagers de trains de banlieue parfaitement fictifs ».

57. Le problème est de savoir jusqu'où il faut aller pour obtenir un jeu de données de bonne qualité. Tel qu'indiqué plus haut, l'apparition des ordinateurs, d'abord des gros ordinateurs, puis des micro-ordinateurs, a permis une automatisation pratiquement complète des opérations de vérification. Dans nombre de bureaux nationaux de recensement/statistique, les démographes sont en fait devenus de fervents partisans des vérifications. Aussi y réalise-t-on à présent de nombreux contrôles de cohérence autrefois difficiles à mettre en place, en particulier ceux impliquant des contrôles interenregistrements et interménages. Malheureusement, cette possibilité offerte par les micro-ordinateurs a également entraîné nombre de problèmes, au premier rang desquels figure la survérification.

1. COMMENT LA SURVÉRIFICATION EST-ELLE PRÉJUDICIALE

58. La survérification a un effet préjudiciable sur le processus de contrôle des données, et ce à plusieurs égards, notamment en termes de respect des délais, de coût et de distorsions des valeurs vraies. De plus, il en résulte un faux sentiment de sécurité quant à la qualité des données. Ces différentes préoccupations sont passées en revue ci-dessous.

a) *Respect des délais*

59. Le dépouillement est d'autant plus long que les opérations de vérification effectuées par un bureau national de recensement/statistique sont nombreuses. L'essentiel est de déterminer dans quelle mesure la durée supplémentaire augmente la valeur des résultats du recensement. Les équipes de vérification doivent donc évaluer, tant en permanence qu'à posteriori, les avantages nets d'un allongement de la durée de dépouillement et d'une augmentation des moyens affectés au recensement dans son ensemble. Souvent, les gains sont si limités au regard du temps investi, qu'il vaut mieux tolérer de petites « anomalies » dans les données que priver les principaux utilisateurs de l'obtention en temps opportun des informations escomptées.

b) *Charges financières*

60. De manière analogue, les coûts liés au recensement augmentent avec la durée du dépouillement. Il incombe à

chaque bureau national de recensement/statistique de décider, lorsqu'il entreprend des vérifications plus importantes et plus complexes, si l'augmentation des coûts est justifiée par le surcroît de travail et peut être effectivement prise en charge.

c) *Distorsion des valeurs vraies*

61. Bien que le processus de vérification ait pour objectif d'améliorer la qualité des données, la multiplication et la complexité croissante des corrections risquent d'avoir un effet préjudiciable. Parfois, les équipes de vérification modifient par erreur des données pour différentes raisons : défaut de communication entre le démographe et les spécialistes de traitement des données; erreurs d'un programme sophistiqué particulièrement complexe; ou manipulation à plusieurs reprises d'une donnée dans le cadre d'un contrôle. Les bureaux nationaux de recensement/statistique cherchent à éviter ce type de problème dans la mesure du possible. Comme Granquist et Kovar (1997) le fait d'imputer l'âge d'un mari et d'une épouse en supposant entre eux une différence d'âge prédéfinie peut s'avérer extrêmement utile, mais risque d'introduire une distorsion artificielle des données lorsque ces cas sont très nombreux.

d) *Un sentiment erroné de sécurité*

62. La survérification des données confère au personnel des bureaux nationaux de recensement/statistique, ainsi qu'aux autres utilisateurs, un faux sentiment de sécurité, surtout lorsque les services en question s'abstiennent d'adopter et de documenter des mesures d'assurance qualité. En outre, les tableaux obtenus feront apparaître certaines anomalies, indépendamment de l'importance du travail réalisé par l'équipe de vérification; aussi importe-t-il d'avertir les utilisateurs de la présence éventuelle de petites erreurs. Tel est particulièrement le cas maintenant que de nombreux pays diffusent des échantillons de microdonnées. Les bureaux nationaux de recensement/statistique ne souhaitant pas diffuser des données susceptibles de nuire au processus de planification, il faut veiller tout particulièrement à ce que toutes les variables décisives fassent l'objet de vérifications adéquates et soient utilisables à des fins de planification. Par exemple, aucun bureau de recensement/statistique, ne souhaiterait diffuser des microdonnées ou des tableaux comportant des données inconnues de sexe ou d'âge. En revanche, des variables telles que les caractéristiques relatives à l'incapacité ou à l'aptitude à lire ou à écrire sont parfaitement utilisables avec des vérifications moins importantes. Bien que les tableaux croisés puissent faire apparaître certains défauts de cohérence parce que les bureaux nationaux de recensement/statistique ne peuvent vérifier tous les couples de variables, les équipes de vérification devraient contrôler les principales combinaisons. Lorsque des défauts de cohérence sont mis en évidence, des procédures de correction doivent avoir été mises au point.

2. TRAITEMENT DES DONNÉES INCONNUES

63. L'équipe de vérification doit décider à un stade précoce de la planification du recensement comment traiter les

données « non spécifiées » ou inconnues. Tel qu'indiqué plus haut, les colonnes ou les lignes de données inconnues dans les tableaux étant dépourvues aussi bien de valeur informative que d'utilité, dans la plupart des pays, les planificateurs du recensement préfèrent imputer les données en question. En l'absence de traitement, de nombreux utilisateurs les répartissent dans les tableaux de résultats dans une proportion identique à celle des données connues, procédant ainsi à une imputation après coup. L'équipe de vérification doit donc décider du traitement systématique des données inconnues.

3. MODIFICATIONS PARASITES

64. D'ordinaire, les bureaux nationaux de recensement/statistique n'utilisent pas de modèles lorsqu'ils définissent leurs règles de vérification. Les équipes de vérification devraient définir des règles adaptées aux caractéristiques réelles de la population ou de l'habitat. Toutes les données recueillies doivent satisfaire à ces règles. Par exemple, il est possible de spécifier par une série de règles, que l'enfant d'un chef de ménage doit avoir au moins 15 ans de moins que lui. Toutefois, il peut s'agir d'un enfant « par alliance » et non biologique. Cet enfant peut être l'enfant biologique du conjoint, mais non du chef de ménage, auquel cas, la différence d'âge peut être inférieure à 15 ans. Puisque les planificateurs dans la plupart des pays ne réservent pas un traitement distinct pour les enfants et les enfants par alliance, si dans les circonstances ci-dessus les règles de vérification ont pour effet de corriger l'âge de l'enfant, il risque d'apparaître des défauts de cohérence en matière de niveau de formation, de participation à la population active et dans d'autres domaines. Aussi, convient-il de tester ce principe afin d'examiner les résultats avant de l'appliquer pleinement.

4. DÉTERMINATION DE TOLÉRANCES

65. L'équipe de vérification doit définir des niveaux de tolérance pour chaque donnée et parfois pour des combinaisons de données. Les niveaux de tolérance indiquent le nombre des réponses non valides et incohérentes admis avant que les équipes de vérification adoptent des mesures correctives. Ainsi, pour la plupart des rubriques d'un recensement, un petit pourcentage des personnes interrogées ne fourniront pas de réponses acceptables, pour une raison ou une autre. Pour certains types de données, par exemple l'âge et le sexe, généralement combinées à d'autres types de données à des fins de planification, le niveau de tolérance peut être très bas. Lorsque le pourcentage de réponses manquantes ou incohérentes est faible (inférieur à 1 ou 2 %), aucune règle judicieuse de vérification ne risque d'affecter l'utilisation des données. En revanche, lorsque le pourcentage est élevé (5 à 10 % ou davantage selon la situation), une imputation simple ou même complexe risque d'entraîner une distorsion dans les résultats du recensement.

66. Pour réduire au minimum la proportion des réponses manquantes, les bureaux nationaux de recensement/statistique devraient veiller à ce que les recenseurs fassent tout leur possible pour obtenir les données sur le terrain. Si un pays donné décide qu'il n'a pas besoin d'une précision très élevée pour certains types de données, telles que la capacité à lire ou

à écrire, ou les caractéristiques d'invalidité, le niveau de tolérance relatif à ces données peut être nettement plus élevé. Parfois, les équipes de vérification peuvent corriger des données comportant des erreurs trop nombreuses, en renvoyant les agents recenseurs sur le terrain, en effectuant de nouvelles entrevues téléphoniques ou en mettant à profit leur connaissance d'un secteur de recensement particulier. Souvent toutefois, il s'avère trop coûteux de retourner sur le terrain ou de réaliser d'autres opérations de suivi, et le bureau national de recensement/statistique peut convenir soit de ne pas utiliser les données en question, soit de les utiliser seulement en leur associant des notes de mise en garde.

67. La question se pose de savoir qui doit déterminer le niveau de tolérance pour une information. L'équipe de vérification, notamment les démographes et les informaticiens, est parfois tenue de prendre une décision à cet égard. Les spécialistes doivent utiliser ces informations de façon durable et le fait de veiller à obtenir les données de la meilleure qualité possible est un enjeu professionnel. Quant aux spécialistes du traitement des données, ils peuvent estimer ne pas être effectivement en mesure de mettre au point des programmes de vérification appropriés afin de ramener le niveau de tolérance à des valeurs acceptables ou encore que les données proprement dites n'autorisent pas l'utilisation efficace d'un programme quelconque, avec un niveau de tolérance donné.

5. ENSEIGNEMENTS DU PROCESSUS DE VÉRIFICATION

68. Une fois terminée la vérification des données, il faut garder une trace des analyses détaillées des effets positifs et négatifs obtenus, afin d'améliorer la qualité tant de l'enquête et du recensement actuel que des enquêtes ou des recensements futurs. L'équipe de vérification doit s'employer constamment à déterminer ce qui fonctionne correctement et ce qui ne fonctionne pas. Elle doit en outre établir si les aspects du processus qui fonctionnent correctement peuvent être améliorés et simplifiés, pour que les utilisateurs disposent plus facilement des résultats. Les bureaux nationaux de recensement/statistique seront d'autant plus en mesure de corriger les erreurs qu'ils les auront détectées plus tôt au cours du recensement.

6. ASSURANCE QUALITÉ

69. L'assurance qualité joue un rôle important dans tous les recensements. Par conséquent, des mécanismes formels d'assurance qualité doivent incontestablement être mis en place pour surveiller le bon déroulement de la phase de vérification et d'imputation automatiques. Les analyses rétrospectives, les mesures de performance et les statistiques de diagnostic sont des outils essentiels pour l'analyse de la qualité des vérifications, comme pour la rapidité du dépouillement (Granquist et Kovar 1997; Statistique Canada, 1998).

7. COÛT DES OPÉRATIONS DE VÉRIFICATION

70. Le présent manuel peut aider les pays à réduire les coûts élevés en termes de temps et de moyens matériels à engager pour mener à bien la vérification et l'imputation des

données des recensements ou des enquêtes. Comme le font observer Granquist et Kovar (1997; p. 418), même « dans les années 90, les opérations de vérification s'avèrent pratiquement aussi coûteuses qu'elles l'étaient dans les années 70, bien qu'elles aient été dans une large mesure rationalisées grâce à la mise à profit permanente de l'évolution technologique ». Pour la plupart des pays, les opérations de vérification absorbent un temps et des moyens financiers disproportionnés : aussi chacun d'eux doit-il déterminer quelle doit être la rentabilité de son investissement. D'après Granquist et Kovar (1997) le coût de la vérification des enquêtes sur les ménages, représentait quelque 20 % du budget total des recensements réalisés dans le monde au début des années 90.

71. Les vérifications excessives risquent de retarder la publication des résultats. Bien que les agents nationaux chargés des recensements ou des enquêtes aient peut-être uniquement des confirmations anecdotiques de ce type d'expérience en matière de recensement, il ressort d'une étude de Pullum, Harpham et Ozsever (1986) que la vérification automatisée de l'enquête mondiale sur la fécondité (*World Fertility Survey*) a contribué à retarder la publication des résultats d'environ un an. Les bureaux nationaux de recensement/statistique auraient peut-être intérêt à consacrer leurs moyens financiers en priorité à l'amélioration de la qualité de la phase de dénombrement du recensement ou de l'enquête.

8. IMPUTATION

72. L'imputation désigne le processus consistant à résoudre les problèmes concernant les réponses manquantes, non valides ou incohérentes détectées par les vérifications. L'imputation a pour effet de modifier une ou plusieurs des réponses ou informations manquantes d'un ou plusieurs enregistrements en cours de vérification, afin d'assurer l'obtention de résultats plausibles, dotés d'une cohérence interne. La consultation de la personne interrogée ou l'examen manuel du questionnaire permet d'éliminer certains des problèmes rencontrés à un stade moins avancé du processus. Toutefois, il est généralement impossible de résoudre tous les problèmes dans la phase initiale en raison de préoccupations touchant au fardeau de réponse, au coût et au respect des délais. Les opérations d'imputation traitent ensuite des autres anomalies décelées par la vérification, puisqu'il est souhaitable d'établir un fichier complet et cohérent conte-

nant des données imputées. Le meilleur travail d'imputation est réalisé par les membres de l'équipe qui ont pleinement accès aux microdonnées et qui détiennent de bonnes informations complémentaires :

a) Après imputation, le fichier corrigé ne doit pas différer considérablement du fichier vérifié où la présence d'erreurs a été constatée. Aussi, l'imputation d'un nombre minimal de variables est généralement la meilleure méthode, puisqu'elle protège autant que possible les données fournies par les personnes interrogées. Cela suppose implicitement (ce qui n'est pas toujours vrai dans la pratique) qu'une personne interrogée est plus susceptible de faire seulement une ou deux erreurs qu'un nombre plus important;

b) Après imputation, le fichier obtenu doit satisfaire aux exigences de toutes les vérifications;

c) Les équipes de vérification devraient identifier comme telles les valeurs imputées, tandis que les méthodes et les sources d'imputation devraient être clairement identifiées. Afin de pouvoir évaluer l'importance des opérations d'imputation et leurs répercussions, l'équipe de vérification devrait conserver les valeurs imputées et non imputées des champs correspondants des enregistrements de données.

9. ARCHIVAGE

73. Une partie des opérations d'assurance qualité relatives aux recensements ou aux enquêtes consistent à documenter toutes les tâches, puis à archiver les informations ainsi réunies. Les bureaux nationaux de recensement/statistique doivent conserver les fichiers de données vérifiées et non vérifiées à des fins d'analyse ultérieure. Certaines procédures, comme la numérisation, conservent automatiquement l'image saisie initialement. De manière analogue, immédiatement après la saisie de lots de données, il convient de concaténer les données ainsi obtenues et de les conserver en vue d'analyses ultérieures éventuelles.

74. La documentation établie doit être suffisamment complète pour permettre aux organisateurs des recensements ou des enquêtes de reconstituer les mêmes opérations à une date ultérieure afin de garantir la compatibilité avec l'enquête ou le recensement considéré. Les opérations effectuées et les résultats obtenus doivent nécessairement être reproductibles. Enfin, les données non vérifiées, comme les données vérifiées doivent être enregistrées en plusieurs endroits, et des mesures appropriées doivent être prises pour garantir leur disponibilité permanente au fil des ans.

II. — MISE EN ŒUVRE DES OPÉRATIONS DE VÉRIFICATION

75. Le chapitre II donne une vue d'ensemble globale des applications du processus de vérification et d'imputation. Il présente certains exemples destinés à illustrer la nature des problèmes posés le cas échéant aux utilisateurs par des données non vérifiées et les raisons pour lesquelles il est plus intéressant de disposer de données dûment vérifiées. Il examine les problèmes de saisie et de codage dans le cadre des opérations préliminaires de vérification. Le chapitre présente en outre les principaux aspects des opérations de vérification automatisée et les principes généraux en matière, par exemple, de contrôle de validité et de cohérence. Deux types généraux de vérification automatisée, les techniques d'imputation statique (*cold deck*) et d'imputation dynamique (*hot deck*) font l'objet d'un examen détaillé.

76. La vérification des recensements et des enquêtes a pour objet de mettre en évidence les omissions et les défauts de cohérence des enregistrements de données; quant aux opérations d'imputation, elles visent à introduire les corrections nécessaires. La vérification définit les procédures spécifiques adaptées au traitement des omissions et des données inacceptables. Les opérations d'imputation modifient les données non valides et résolvent les incohérences détectées dans l'ensemble des informations introduites. Il en résulte un fichier de microdonnées vérifiées en vue des divers types d'exploitation, contenant des informations concernant toutes les données élémentaires pertinentes relatives à chaque unité de logement et à chaque personne recensée.

77. Il importe néanmoins de signaler à nouveau que la vérification, si poussée soit-elle, ne saurait remplacer un bon travail de dénombrement. Le processus de vérification est efficace lorsque les imputations s'appliquent aux omissions et aux incohérences aléatoires. Toutefois, si des erreurs systématiques affectent la collecte de données, la vérification ne peut améliorer la qualité des données quelle que soit la complexité des procédures mises en œuvre. Par ailleurs, le choix des sujets étudiés est d'une importance décisive pour la qualité des données recueillies. Les personnes interrogées doivent être désireuses et capables de fournir les informations adéquates. Il est donc parfois nécessaire d'éviter des sujets susceptibles d'éveiller des craintes, des préjugés locaux ou des superstitions, ou encore de poser des questions trop complexes et trop difficiles pour permettre à la personne interrogée « moyenne » de répondre facilement dans le contexte d'un recensement de la population. La formulation précise de chaque question nécessaire à l'obtention de la réponse la plus fiable possible sera obligatoirement fonction du contexte national particulier et devrait être soigneusement testée au préalable. Pour obtenir des données de meilleure qualité, il importe donc au plus haut point que les bureaux nationaux de recensement/statistique engagent des ressources suffisantes consacrées au recensement.

78. Afin de mettre en place la phase de vérification automatisée du processus, l'équipe de vérification établit à cet effet des instructions ou des spécifications rédigées par écrit, des tables de décision, des organigrammes et des pseudo-codes². Les organigrammes permettent aux spécialistes du sujet de mieux comprendre les différentes relations entre les variables et facilitent l'élaboration des instructions de vérification. L'annexe IV présente des exemples d'organigrammes. Les spécialistes du sujet rédigent les instructions de vérification en collaboration avec les informaticiens, en décrivant l'opération à effectuer pour chaque élément d'information; ces instructions doivent être claires, concises et dépourvues d'ambiguïté puisqu'elles constituent le point de départ du progiciel de vérification.

79. Toute l'équipe de vérification du recensement, tant les démographes que les informaticiens, doit être parfaitement familiarisée avec le traitement et l'analyse des données démographiques. En effet, un personnel non qualifié risque d'introduire dans le recensement des erreurs et des biais supplémentaires.

A. — Vérification manuelle et vérification automatisée

80. La vérification manuelle d'un recensement peut prendre des mois ou des années, et donner lieu ainsi à de nombreuses possibilités d'erreur humaine. De fait, il s'agit d'une alternative qui laisse à désirer par rapport à la vérification automatisée, notamment en raison de l'impossibilité de créer ou de reconstituer une analyse rétrospective du processus de correction manuelle. Les vérifications informatisées ou automatisées réduisent les délais et limitent les risques d'erreur humaine. Les vérifications tant informatisées que manuelles ont pour effet de vérifier la validité d'une valeur introduite par comparaison à une valeur acceptable, tandis que les programmes informatisés contrôlent également la cohérence de la valeur introduite par rapport aux valeurs connexes. Enfin, et avant tout, les vérifications automatisées permettent d'établir une analyse rétrospective et sont donc reproductibles, contrairement aux vérifications manuelles.

81. Lorsque les recensements et les enquêtes recueillent de grandes quantités de données, il n'est pas toujours possible de se référer aux documents originaux pour corriger les erreurs. Même si l'on dispose des questionnaires d'origine, les données consignées sur ces derniers risquent parfois d'être erronées ou incohérentes. Or, un système informatisé de vérification et d'imputation permet de corriger ou de modifier immédiatement des données erronées, et d'établir des rapports pour toutes les erreurs décelées et pour toutes les

² Pseudo-code est un ensemble d'instructions écrites ou de spécifications tel qu'indiqué à la figure 7.

modifications effectuées. Il convient de planifier soigneusement les vérifications informatisées afin de préserver le temps dont dispose le personnel pour les autres activités de traitement des données. Le passage d'importantes quantités de données dans un système informatique peut être long, mais il prend moins de temps que les corrections manuelles.

82. Les corrections manuelles sont réalisées sous différentes formes. Considérons l'exemple simple d'une erreur concernant le sexe d'une personne recensée : un contrôleur vérifie le travail d'un agent recenseur et constate une erreur évidente, telle que l'enregistrement de la mention « homme » dans le cas d'une personne prénommée « Marie ». S'il corrige le sexe et inscrit la mention « femme », le contrôleur effectue une correction manuelle. Si le contrôleur ne corrige pas le questionnaire, mais l'envoie au contraire au bureau de recensement local, le problème en question peut y être constaté et corrigé manuellement. Au bureau central, le défaut de concordance entre le prénom et le sexe peut être décelé au cours du codage et corrigé manuellement par le codeur; sinon, le personnel chargé d'introduire les données relatives au questionnaire peut observer le défaut de concordance entre le prénom et le sexe et apporter les corrections nécessaires avant la saisie.

83. Toutefois, si l'erreur n'est pas décelée, et si l'opérateur de saisie introduit le code « homme », plusieurs procédures peuvent alors être exécutées à ce stade. Pour les éléments d'information liés au sexe, comme ceux recueillis par le bloc de questions sur la fécondité, le programme de vérification peut indiquer de fait qu'il s'agit d'un homme auquel est associée une information sur la fécondité et produire un message à cet effet au moment de la saisie des données. L'opérateur pourrait ensuite examiner le questionnaire, établir qu'il s'agit effectivement d'une femme, et effectuer manuellement la correction. Sinon, au cas où le bureau national de recensement/statistique utilise un programme de vérification indépendant de la saisie, le programme informatique pourrait indiquer cette personne comme étant un homme auquel sont associées des données sur la fécondité. Les employés du bureau peuvent retrouver le questionnaire original dans les boîtes, l'examiner, et établir que la personne interrogée prénommée « Marie » a été enregistrée par erreur en tant qu'« homme ». A ce stade, le personnel du bureau peut remonter cette information à l'opérateur, lequel peut reprendre l'enregistrement et le corriger manuellement.

84. Cet exemple fait apparaître conjointement les avantages et les inconvénients de la vérification manuelle. A chacune des étapes décrites ci-dessus, un employé du recensement a pu constater l'erreur, le défaut de concordance entre le prénom et le sexe, et faire la correction nécessaire. Toutefois, il existe dans les bureaux nationaux de recensement/statistique qui utilisent le système de vérification manuel des agents qui vérifient à chaque étape cette concordance. Il faut consacrer à cette tâche une somme de travail considérable, pour des résultats vraisemblablement peu différents (en particulier les résultats d'ensemble), de ceux obtenus avec la consigne de n'effectuer aucune vérification manuelle.

85. Jusqu'à une date récente, la seule façon de corriger un ensemble de données consistait à introduire manuellement les modifications. Au demeurant, de nombreux pays jugent encore peu pratique d'utiliser un système de correction automatique, et font appel au système manuel, à l'une des étapes évoquées ci-dessus. Si le jeu de données étudiées est limité ou si le facteur temps n'est pas décisif, ou encore si l'on dispose d'une main-d'œuvre importante, alors le choix des corrections manuelles sera viable dans de nombreux cas. Il a ainsi pour avantage, si l'information recueillie sur le questionnaire est complète et exacte, et si l'examen de ce dernier permet vraiment de résoudre les incohérences, d'entraîner probablement une amélioration limitée de la qualité du recensement ou de l'enquête (l'équipe de vérification est alors tenue de supposer par exemple que le prénom « Marie » n'est pas « Gary »; qu'en présence de données sur la fécondité, l'information effectivement recueillie correspond à cette personne; et que celle-ci n'a pas été recueillie de façon erronée). En fait, les procédures de vérification et d'imputation améliorent rarement la qualité de la collecte des données. Elles modifient seulement certains éléments.

86. Parfois, l'examen d'un questionnaire en vue d'une correction manuelle s'avère infructueux, l'information recherchée étant absente, pour une raison ou une autre. Parfois, une personne ne souhaite pas indiquer son âge et laisse en blanc la case correspondante, auquel cas, l'étude du questionnaire ne résoudra pas la question. Ensuite, l'équipe de vérification doit prendre une décision quant à la façon de résoudre le problème. Pour procéder à une correction manuelle, le bureau national de recensement/statistique doit soit attribuer la valeur « inconnu », soit utiliser un ensemble de valeurs susceptibles d'être attribuées à la rubrique « âge ».

87. Lorsque la personne interrogée est contactée, la correction manuelle altère inévitablement la qualité et la cohérence des réponses; elle prend plus de temps et coûte davantage. Or, les ordinateurs ne sont pas exposés à la fatigue et sont donc plus rapides; en outre, ils ne souffrent pas de problèmes personnels qui peuvent s'avérer préjudiciables au maintien de la qualité ou de la cohérence des données; et, dans la plupart des cas, ils permettent de réduire le coût du traitement. Aussi, la plupart des pays utilisent-ils à présent une forme ou une autre de correction automatique.

88. Les omissions et les incohérences observées dans les réponses réduisent la qualité des données et rendent problématique la présentation de résultats de recensement aisément compréhensibles. Certains utilisateurs préfèrent mettre en tableau les réponses manquantes et incohérentes regroupées sous la rubrique « non indiqué », tandis que d'autres préfèrent répartir ces cas proportionnellement aux données cohérentes indiquées. D'autres encore préconisent des règles d'imputation de réponses probables aux réponses manquantes ou incohérentes. Grâce aux ordinateurs, l'imputation de réponses à partir des autres informations figurant dans le questionnaire, ou sur la base d'informations indiquées pour une personne ou une unité de logement présentant des caractéristiques similaires, est une opération parfaitement réalisable et ne présentant pas de difficulté.

89. Puisque l'ordinateur offre la possibilité d'examiner de nombreuses caractéristiques, le processus de vérification doit mettre à profit cette capacité. Les procédures de vérification impliquant un grand nombre de caractéristiques connexes peuvent donc conduire à imputer davantage de réponses vraisemblables que ne le permettrait une vérification simple. En revanche, une opération de vérification mal conçue risque d'aboutir à la production de données de recensement de qualité médiocre. L'équipe de vérification devrait donc être constituée de spécialistes expérimentés des différentes disciplines concernées, ainsi que d'informaticiens. Les membres de cette équipe devraient procéder soigneusement au choix des variables à examiner dans les tests de cohérence afin de définir les spécifications de vérification et d'imputation. De plus, parmi les résultats du programme devrait figurer le pourcentage de réponses modifiées ou imputées : les analystes seront ensuite en meilleure position pour juger de la qualité des données; par exemple, un fort pourcentage d'imputations constituerait un avertissement, invitant à une utilisation prudente des résultats.

90. L'analyse rétrospective des vérifications fait apparaître les modifications apportées à chaque variable. Elle permet de tracer l'historique des réponses depuis la réception des données et tout au long du processus de vérification et d'imputation.

B. — Principes de correction des données

91. Manuel ou automatique, le processus de vérification devrait s'employer à rendre les données aussi représentatives que possible de la situation réelle en éliminant les omissions, les réponses non valides et en modifiant les réponses incohérentes.

92. Considérons le diagramme suivant (figure 1) relatif à un ménage particulier. Il décrit un ménage et fait apparaître

ENCADRÉ 2. Principes importants de correction des données

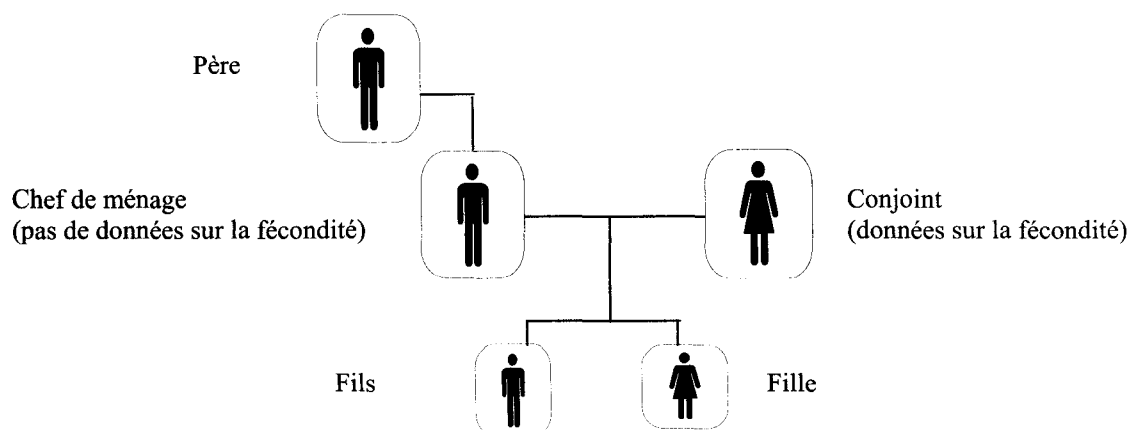
Lors de la mise au point des procédures de vérification, il peut être utile de ne pas perdre de vue les suggestions suivantes en matière de correction des données :

- 1) Effectuer le plus petit nombre possible de modifications des données initialement;
- 2) Eliminer les incohérences évidentes des réponses;
- 3) Etablir des valeurs correspondant aux éléments inexacts ou manquants sur la base d'autres éléments d'information concernant le logement, la personne ou les autres personnes du ménage ou d'un groupe comparable, conformément aux procédures spécifiées. Dans certains cas, il convient d'utiliser pour certains éléments d'information la catégorie de réponse « non indiquée ».

des relations et des données de sexe cohérentes. Le chef du ménage est un homme et ne fait pas l'objet de données sur la fécondité. Le conjoint est de sexe féminin et les informations appropriées de fécondité lui sont associées.

93. Dans de nombreux cas toutefois, les informations sont incohérentes. Il convient alors de répondre aux questions suivantes : quel devrait être le processus de vérification dans le cas d'un ménage faisant l'objet de réponses incohérentes ? Comment doit procéder l'équipe de vérification si le chef de ménage et le conjoint sont signalés comme étant de sexe masculin, comme sur la figure 2 ? Autrefois, le principe de vérification généralement adopté aurait consisté à supposer que la première personne dans un couple est un homme, en particulier s'il s'agit du chef de ménage, et que la deuxième personne, c'est-à-dire le conjoint, est une femme.

FIGURE 1. Exemple de ménage fictif, avec indication du lien avec le chef de ménage, du sexe et de la fécondité de ses membres



94. Si en l'occurrence le chef de ménage s'avère être la femme et non le mari, alors la règle de vérification adoptée serait erronée et le bureau national de recensement/statistique enregistrerait en définitive quatre erreurs :

- a) Erreur concernant le sexe du chef de ménage;
- b) Erreur concernant le sexe du conjoint;
- c) Perte des données sur la fécondité relatives au chef de ménage;
- d) Attribution erronée de données sur la fécondité au conjoint masculin.

Il s'agit évidemment d'une procédure de vérification incorrecte.

95. En revanche, après avoir constaté l'identité du sexe du chef de ménage et du conjoint, une procédure de vérification correcte contrôle ensuite les données sur la fécondité relatives à ces deux personnes. Puisque seul le chef de ménage comporte des données sur la fécondité, il s'agit de la personne interrogée de sexe féminin. Les règles de vérification concernant ces rubriques sont alors correctement appliquées.

96. L'exemple de la figure 3 illustre également ce problème. Pour la plupart des pays, l'âge de procréation est compris entre 15 et 49 ans. Considérons le cas d'une femme déclarant avoir un enfant à l'âge de 52 ans, d'après des données directes déduites du numéro de ligne indiqué pour la mère de l'enfant, ou d'après la différence d'âge calculée (la différence d'âge entre la mère et l'enfant ne peut dépasser 50 ans). L'équipe de vérification doit décider si la différence d'âge est acceptable ou si elle doit être modifiée, par le changement de l'un des deux âges en question. Si l'on augmente la différence d'âge acceptable et si d'autres femmes déclarent avoir des enfants à des âges plus avancés, alors des anomalies plus nombreuses risquent d'être introduites dans le

jeu de données en cas d'indications erronées concernant l'âge proprement dit. Là encore, l'équipe de vérification doit décider du caractère approprié des âges déclarés pour différentes variables.

97. La figure 4 illustre un autre scénario possible. Supposons que les vérifications détectent le cas d'une femme de 70 ans avec des enfants de 8 et 10 ans tel qu'indiqué à la figure 4. Cette situation est possible, le mari pouvant avoir eu les enfants avec une précédente épouse. Dans ces conditions, les enfants sont rattachés au chef de ménage et non à l'épouse en tant que telle, bien qu'il soit sans doute plus probable que les opérateurs aient fait une erreur en tapant le chiffre 7, croyant taper le chiffre 4 correspondant à 40. Pour une raison ou une autre, supposons que les spécialistes du sujet demandent aux informaticiens de modifier l'âge de la mère ou de l'enfant lorsque leur différence d'âge dépasse 50 ans. Or, une demande de ce type entraîne des vérifications encore plus compliquées. Puisque l'âge de la femme est de 70 ans et celui du premier enfant de 10 ans, l'équipe de vérification doit choisir la personne dont l'âge doit être modifié; elle pourrait ainsi décider de porter à 20 ans l'âge du premier enfant, ce qui résoudrait le problème dans ce cas particulier, ou encore de modifier l'âge du conjoint. Il reste néanmoins à résoudre le problème posé par l'âge du deuxième enfant, qui doit aussi faire l'objet d'une vérification.

98. Si l'on considère uniquement les âges de la mère et d'un enfant, une imputation attribuerait l'âge de façon aléatoire et donnerait des résultats exacts dans environ 50 % des cas. Toutefois, si l'on considère également l'âge du mari, l'équipe de vérification aurait davantage tendance à modifier l'âge de l'épouse compte tenu de cette information complémentaire. Cette modification aurait pour effet de renforcer la compatibilité des âges des différents membres de la famille.

FIGURE 2. Exemple de ménage : chef de ménage et conjoint de même sexe

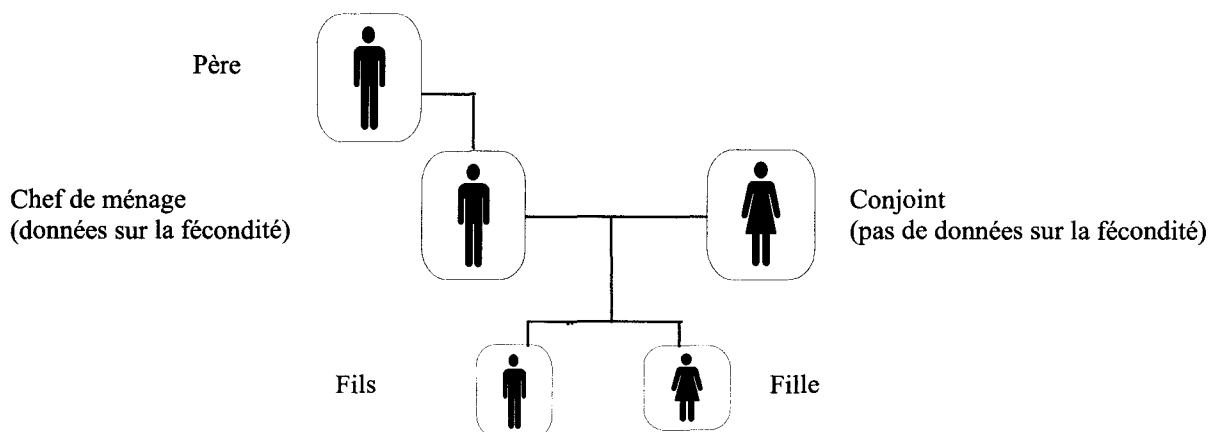


FIGURE 3. Exemple de ménage avec indication des âges de certains membres

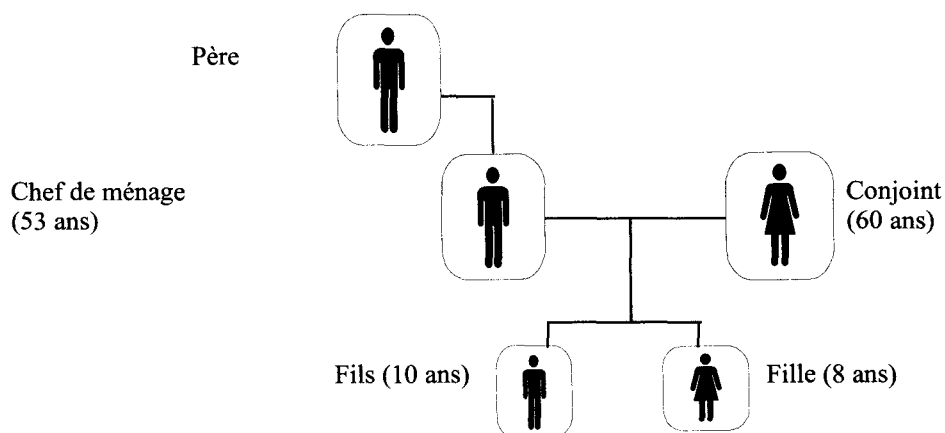
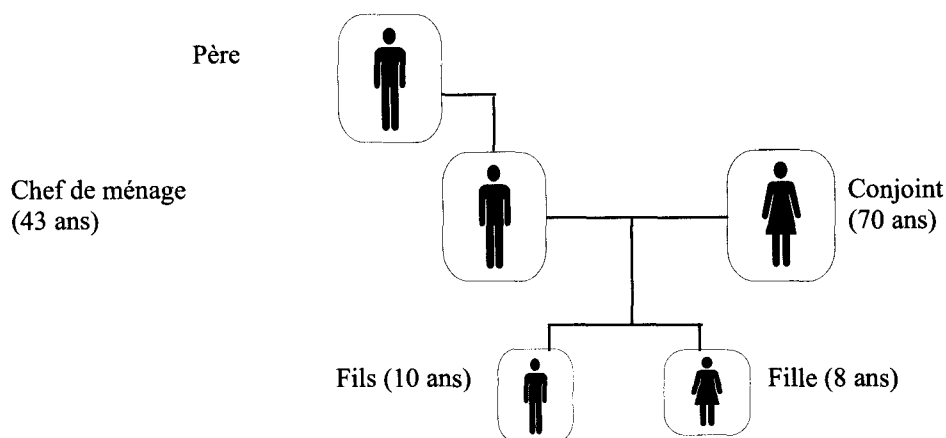


FIGURE 4. Exemple de ménage présentant des incohérences potentielles des âges déclarés



C. — Contrôle de validité et de cohérence

99. Parmi les principales exigences des opérations de vérification, figure l'élimination des valeurs non valides. En outre, il faut assurer la cohérence des réponses fournies pour chacune des rubriques connexes, à l'intérieur de chaque enregistrement et d'un enregistrement à l'autre.

100. Les opérations d'imputation devraient prendre en compte toutes les informations disponibles concernant les variables connexes au même moment, dans toute la mesure possible, et pas nécessairement de façon séquentielle, par rapport aux variables connexes. Dans certains cas cependant, la vérification peut procéder à un contrôle de cohérence avant de déterminer la validité d'une donnée. Lorsque l'imputation attribue une valeur sur la base d'un contrôle de cohérence, elle doit comparer la valeur à la donnée d'origine

pour déterminer s'il y a eu effectivement modification; sinon la donnée initiale reste inchangée.

101. Par exemple, pour la vérification de la situation de famille, le contrôle porte initialement sur le lien avec le chef de ménage pour vérifier si la donnée enregistrée est « conjoint »; dans cette hypothèse et si le conjoint n'est pas signalé comme étant marié, la valeur « marié » est attribuée à la variable « état matrimonial ». Avant d'attribuer le code « marié », le programme vérifie la nature de la réponse d'origine. Si le code correspondant à la valeur « marié » est déjà présent, alors le programme ne modifie pas la donnée enregistrée et aucune erreur n'est constatée.

1. MÉTHODE DE VÉRIFICATION DESCENDANTE

102. Cette procédure commence par la première rubrique à vérifier (le « haut »), généralement la première va-

riable du questionnaire, puis passe successivement aux différentes rubriques, jusqu'à la dernière. Bien que la méthode descendante ne tienne pas compte des relations entre les éléments d'information, il s'agit toutefois d'une procédure adéquate pour mener à bien la vérification.

103. Au cours de la vérification, certaines opérations modifient la valeur d'une rubrique à plusieurs reprises. Cette procédure risque d'introduire une ou plusieurs erreurs dans le jeu de données. Une valeur imputée peut en effet ne pas être cohérente par rapport aux autres données. Même lorsque les variables sont traitées de façon séquentielle, une variable particulière doit être vérifiée en tenant compte, si possible simultanément, de toutes les autres. Par exemple, s'il est imputé en fonction de l'âge de la mère, l'âge d'un enfant risque d'être incompatible avec le nombre d'années de scolarité déclaré ou d'années passées dans la circonscription. Dans ce cas, l'âge fera l'objet d'une nouvelle imputation, jusqu'à ce qu'il soit compatible avec les autres données. Un âge imputé constitue une variable intermédiaire jusqu'à ce qu'une valeur finale soit attribuée. Lors de la création de données vérifiées, les variables intermédiaires imputées ne doivent pas être enregistrées en tant que modifications, tant qu'elles n'ont pas été définitivement attribuées.

104. Bien que le programme de vérification puisse accepter la présence d'un « blanc », ou la valeur dite « non indiquée » pour certaines rubriques et dans certaines conditions, les informations connexes peuvent fournir des valeurs pour la plupart des rubriques laissées en blanc ou auxquelles des valeurs erronées ont été attribuées. Les valeurs obtenues de cette façon peuvent être individuellement correctes ou non. Cependant, la puissance et la vitesse de l'ordinateur pour comparer différentes valeurs enregistrées permettent de déterminer les valeurs de remplacement qui rendent compte de la situation de façon satisfaisante. Les tableaux de valeurs obtenus alors, dans la plupart des cas, présenteront une cohérence parfois supérieure par rapport aux tableaux établis à partir des enregistrements non vérifiés ou des enregistrements dans lesquels le processus d'imputation transforme toutes les réponses inacceptables en leur attribuant le code « non indiquée ».

105. Le programme de vérification doit également effectuer des contrôles structurels (voir chapitre III). Les contrôles doivent porter sur les données de population (chap. IV), comme sur les données d'habitation (voir chap. V). En outre, les procédures de vérification devraient vraisemblablement créer une ou plusieurs variables recodées dans l'enregistrement particulier nécessaire à la mise en tableau, tel qu'indiqué à l'annexe I.

106. Il est extrêmement important d'éviter les vérifications circulaires, qui consistent à modifier une ou plusieurs rubriques puis, à un stade ultérieur, à les rétablir à leur valeur initiale. Le présent manuel fait observer par la suite qu'il incombe au personnel d'exécuter à plusieurs reprises le programme de vérification pour être sûr d'avoir entièrement contrôlé toutes les rubriques. Il est possible de définir des règles de vérification qui modifient les données lors d'un premier passage, mais qui, lorsqu'elles sont appliquées aux données modifiées lors d'un second passage, rétablissent la

configuration initiale. Cette procédure risque de se répéter à l'occasion de traitements multiples. L'équipe de vérification doit éviter d'introduire des règles de ce type dans le processus opératoire.

2. MÉTHODE DE VÉRIFICATION PORTANT SUR PLUSIEURS VARIABLES

107. La méthode « descendante » de vérification des recensements et des enquêtes, présentée au chapitre I ci-dessus, ne donne pas toujours les meilleurs résultats (qui se rapprochent le plus de la distribution réelle des variables). Tel qu'indiqué, la méthode descendante, appliquée sans prendre les précautions qui s'imposent, pose souvent des problèmes de vérification.

108. Une autre méthode à envisager est celle des vérifications portant sur plusieurs variables, qui est fondée sur le système de Fellegi-Holt. Cette méthode exige certes davantage de compétences en matière de traitement de données, et une puissance de calcul plus importante, mais elle permet d'obtenir des résultats sans doute plus proches de la « réalité ». L'annexe V intitulée « Méthodes d'imputation » présente différents types de vérifications portant sur plusieurs variables. Ce système comporte nécessairement l'établissement d'une série de propositions « vraies » et la définition du lien entre les variables. Ensuite, le programme applique chaque proposition aux données concernant le ménage afin de vérifier si elles sont toutes exactes. Pour toute proposition erronée, le programme conservera la trace, pour chaque rubrique, des valeurs non valides ou des incohérences. Une fois tous les tests effectués, le système de vérification et d'imputation doit déterminer la meilleure façon de modifier l'enregistrement, pour qu'il subisse avec succès toutes les vérifications. Les équipes de vérification optent généralement pour une stratégie de modification minimale et modifient le plus petit nombre possible de variables afin d'obtenir un enregistrement acceptable.

109. Les onze propositions de la figure 5 donnent un exemple de règles susceptibles d'être appliquées dans le cadre d'une vérification à plusieurs variables de certaines caractéristiques démographiques. Dans cet exemple, le chef de ménage doit avoir 15 ans au moins. Pour des programmes de vérification généraux, il serait plus indiqué de faire figurer un âge minimum de x années, x étant la valeur minimale déterminée pour le pays considéré. Les propositions définies dans cet exemple sont axées sur différentes variables primaires importantes, telles que le lien avec le chef de ménage, le sexe, l'âge, l'état matrimonial et la fécondité. Puisqu'il s'agit de variables étroitement liées, les équipes de vérification auraient intérêt à les examiner conjointement, afin de définir la méthode la plus efficace de contrôle des données. Il convient de signaler en l'occurrence que toutes les variables sont certes importantes, mais que certaines jouent un rôle plus décisif du point de vue de la présentation des données.

110. Dans l'exemple de la figure 6, les deux conjoints appartiennent à la même population que ceux de la figure 5. Ils sont signalés l'un et l'autre comme étant de sexe masculin. En l'occurrence la procédure de vérification est simple

et évidente. La variable pour laquelle les erreurs les plus nombreuses ont été dénombrées est celle qui sera contrôlée en premier. A la figure 6, le programme de vérification met en place la procédure d'imputation relative à la variable « sexe » puisque, d'après les données de la figure 5, cette variable comporte le plus grand nombre d'erreurs eu égard : 1) au lien avec le chef de ménage et au sexe; et 2) à la fécondité et au sexe. Lorsque le programme de vérification contrôle les données sur la fécondité, et constate qu'une information sur la fécondité est associée au chef de ménage et non au conjoint, la procédure d'imputation attribue le sexe féminin au chef de ménage. Enfin, lorsque l'équipe de vé-

fication contrôle à nouveau toutes les erreurs dénombrées, et si toutes les propositions irréfutables sont exactes, aucune vérification supplémentaire n'est nécessaire.

111. Les prescriptions techniques concernant cette vérification peuvent être formulées tel qu'indiqué à la figure 7. Si les données sur la fécondité sont précisées dans les deux cas, le programme de vérification fonctionnera. Toutefois, la vérification n'est évidemment pas complète puisqu'elle considère uniquement les cas dans lesquels les données sur la fécondité sont indiquées et exactes, tant pour le chef de ménage que pour le conjoint.

FIGURE 5. Exemple de règles de vérification à variables multiples de certaines caractéristiques démographiques

N°	Règle	Lien avec le chef de ménage	Sexe	Age	Etat matrimonial	Fécondité
1	Le chef de ménage doit avoir 15 ans au moins					
2	Le conjoint doit avoir 15 ans au moins					
3	Un conjoint doit être marié					
4	En présence d'un conjoint, le chef de ménage doit être marié					
5	En présence d'un conjoint, le chef de ménage et le conjoint ne doivent pas être du même sexe	1	1			
6	Les individus de moins de 15 ans ne sont jamais mariés					
7	Aucune donnée sur la fécondité n'est associée aux personnes de sexe masculin		1			1
8	Aucune donnée sur la fécondité n'est associée aux personnes de sexe féminin de moins de 15 ans					
9	Les données sur la fécondité doivent toujours être indiquées dans le cas des personnes de sexe féminin de 15 ans ou plus					
10	Un enfant doit toujours être plus jeune que le chef de ménage					
11	Un parent doit toujours être plus vieux que le chef de ménage					
TOTAUX		1	2			1

FIGURE 6. Exemple comportant un chef de ménage et un conjoint de même sexe dans un jeu de données non vérifiées

Personne	Lien avec le chef de ménage	Sexe	Nombre d'enfants
Données non vérifiées			
1	Chef de ménage	Masculin	03
2	Conjoint	Masculin	BLANC
Données après vérification portant sur la variable sexe			
1	Chef de ménage	Féminin	03
2	Conjoint	Masculin	BLANC

Figure 7. Exemple de spécification de vérification visant à corriger les valeurs indiquées pour la variable sexe (en langage symbolique)

```

If SEXE du CHEF DE MÉNAGE = SEXE du CONJOINT
If FÉCONDITÉ du CHEF DE MÉNAGE n'est pas en blanc
If FÉCONDITÉ du CONJOINT est en blanc
(sic la variable SEXE du chef de ménage n'est pas déjà mise à la valeur « féminin ») attribuer la valeur
SEXE = féminin
(sic la variable SEXE du conjoint n'est pas déjà mise à la valeur « masculin ») attribuer la valeur SEXE = masculin
Else Une autre correction s'impose... car les deux conjoints ont le même sexe et comportent l'un et l'autre
des données sur la fécondité !
End-if
Else Tel est le cas quand les données sur la fécondité du chef de ménage sont laissées en blanc
If FÉCONDITÉ du CONJOINT n'est pas en blanc
(sic la variable SEXE du chef de ménage n'est pas déjà mise à la valeur « masculin ») attribuer la valeur
SEXE = masculin
(sic la variable SEXE du conjoint n'est pas déjà mise à la valeur « féminin ») attribuer la valeur SEXE = féminin
Else Une autre correction s'impose parce qu'alors AUCUN DES DEUX conjoints ne comporte de données
sur la fécondité !
End-if
End-if
    
```

FIGURE 8. Exemple d'analyses de vérification à plusieurs variables, dans le cas d'une veuve très jeune avec trois enfants

N°	Règle	Lien avec le chef de ménage	Sexe	Age	Etat matrimonial	Fécondité
1	Le chef de ménage doit avoir 15 ans au moins	1		1		
2	Le conjoint doit avoir 15 ans au moins					
3	Un conjoint doit être marié					
4	En présence d'un conjoint, le chef de ménage doit être marié				1	
5	En présence d'un conjoint, le chef de ménage et le conjoint ne doivent pas être du même sexe	1				
6	Les individus de moins de 15 ans ne sont jamais mariés			1	1	
7	Aucune donnée sur la fécondité n'est associée aux personnes de sexe masculin					1
8	Aucune donnée sur la fécondité n'est associée aux personnes de sexe féminin de moins de 15 ans		1	1		
9	Les données sur la fécondité doivent toujours être indiquées dans le cas des personnes de sexe féminin de 15 ans ou plus					
10	Un enfant doit toujours être plus jeune que le chef de ménage					
11	Un parent doit toujours être plus vieux que le chef de ménage	2	1	3	2	1

112. La figure ci-dessus (figure 8) correspond à un exemple de procédure de vérification appliquée à un chef de ménage de sexe féminin âgé de 13 ans, qui est veuve, mais a 3 enfants, d'après les données saisies. L'application des règles de vérification contenues dans le programme donne les résultats suivants :

113. D'après la série de propositions vraies, le plus grand nombre d'erreurs concerne la variable âge et celle-ci doit faire l'objet de modifications en premier lieu. Après modification de la variable âge, la vérification sera terminée si les corrections introduites résolvent toutes les incohérences. Sinon, le programme procède à la vérification de la va-

riable qui se classe au second rang pour le nombre d'incohérences détectées.

D. — Problèmes de codage

114. Les bureaux nationaux de recensement/statistique élaborent une liste des codes à utiliser dans les programmes de vérification et dans les traitements ultérieurs; il est donc vraisemblable qu'ils soient désireux de définir les mêmes codes pour certaines rubriques. Par exemple, dans beaucoup de pays, les codes de lieu (lieu de naissance, lieu de naissance des parents, lieu de résidence antérieur, lieu de travail), de langue, de groupe ethnique, de race et de citoyenneté sont très voisins. De fait, un système de codage commun du « lieu » pourrait être établi, fondé sur l'utilisation de codes à 3 chiffres, le premier correspondant au continent, le deuxième à la région, le troisième au pays particulier. Les bureaux nationaux de recensement/statistique peuvent également utiliser les codes de pays numériques établis par des organismes internationaux tels que la Division de statistique de l'Organisation des Nations Unies (Nations Unies, 1999). L'emploi d'un ensemble de codes communs pour des variables étroitement liées, permet de limiter les erreurs de codage et facilite la tâche des informaticiens pendant le processus de vérification. Si nécessaire, cela permet en outre aux informaticiens, de transférer une valeur codée d'une rubrique à une autre.

115. La structure des codes peut faciliter le processus du codage proprement dit, ainsi que les traitements ultérieurs au cours des opérations de vérification, de mise en tableaux et d'analyse. Dans le cas de grands pays comportant un nombre élevé d'immigrants ou de groupes ethniques, l'utilisation de codes fondés sur l'identification du continent, de la région, et du pays, par des codes ou des chiffres différents pour chacun d'eux, serait préférable à celle d'une simple liste.

116. La figure 9 donne des exemples de codes communs pour des rubriques telles que le lieu de naissance, la citoyenneté, la langue et le groupe ethnique. Aux Philippines, les codes des locuteurs d'ilokano et de tagalog sont diffé-

rents du code général correspondant aux langues pratiquées aux Philippines. En fonction de la situation propre à chaque pays, ces codes pourraient également varier d'un pays à l'autre. Ainsi, alors que la langue anglaise est associée à un seul code, elle est parlée par plusieurs groupes; au Canada et aux Etats-Unis, le lieu de naissance, la citoyenneté et le groupe ethnique correspondent donc à des codes légèrement différents. Pour les personnes nées en France, de citoyenneté française, parlant français et appartenant au groupe ethnique français, le même code est utilisé. En cas d'omission de l'une de ces données élémentaires et si l'équipe de vérification décide que cette solution convient, le traitement de données peut ainsi affecter le code correspondant de l'un des autres éléments d'information.

117. Lorsqu'une série de rubriques d'un questionnaire ne sont pas indépendantes, les agents chargés du recensement/de l'enquête devraient sans doute s'abstenir de solliciter une réponse pour chacune d'elles. Il incombe à l'équipe de vérification de décider au cas par cas, quand il faut utiliser directement d'autres rubriques données pour attribuer une valeur et quand il faut utiliser les autres variables disponibles.

118. Les différences de définitions d'un recensement à l'autre ou entre un recensement et une enquête concernant des variables telles que l'emploi ou le groupe ethnique soulèvent un autre problème. Le bureau national de recensement/statistique doit alors décider des modalités de prise en compte de ces changements, tant pour les données en cours de vérification que pour les jeux de données du précédent recensement, afin de dégager des tendances. Lorsque les données originales non vérifiées sont disponibles, les informaticiens peuvent modifier en conséquence les programmes de vérification et les exécuter à nouveau.

119. Par exemple, un pays européen peut utiliser un seul code de pays d'origine pour tous les pays d'Asie du Sud, lorsque ces données concernent uniquement un petit nombre d'individus. Toutefois, en raison de l'évolution de la structure des flux migratoires, l'enquête ou le recensement suivant est susceptible d'exiger pour tout le traitement, l'utili-

FIGURE 9. Exemples de codes communs attribués à certaines rubriques

<i>Groupe</i>	<i>Lieu de naissance</i>	<i>Nationalité</i>	<i>Langue</i>	<i>Groupe ethnique</i>
France/Français	10	10	10	10
Espagne/Espagnol	20	20	20	20
Amérique latine	25	25	20	25
Philippines/ Philippin	30	30	30	
Ilokano			32	
Tagalog			32	
Angleterre/Anglais	40	40	40	40
Canada	50	50	40	50
Etats-Unis	52	52	40	52

sation de codes distincts pour l'Inde, le Bangladesh, le Pakistan, le Sri Lanka et les autres pays d'Asie du Sud.

E. — Méthodes de correction et d'imputation des données

120. Tel qu'indiqué plus haut, les données recueillies dans tous les recensements et toutes les enquêtes comportent des blancs, qui correspondent aux mentions « non indiqué » ou « inconnu » ou aux informations manquantes pour une autre raison. La présence de données non valides est due par ailleurs aux erreurs de la personne interrogée, de l'agent de recensement ou de l'opérateur de saisie. Les méthodes utilisées pour introduire les corrections nécessaires dépendent du type de données considéré. Dans la plupart des cas, il est possible d'attribuer aux données en cause des codes valides, tout en étant raisonnablement assuré de leur exactitude en utilisant les réponses fournies sous d'autres rubriques et contenues dans l'enregistrement du ménage ou de la personne ou dans ceux d'autres ménages ou d'autres personnes.

121. Ce manuel présente deux techniques informatisées de correction des données erronées. La première est celle de l'imputation statique (méthode *cold deck*, c'est-à-dire à donne fixe) utilisée principalement pour les données manquantes ou inconnues. La seconde est celle de l'imputation dynamique ou méthode *hot deck* (c'est-à-dire à donne variable³), applicable aussi bien aux données manquantes, qu'aux données incohérentes ou non valides.

1. TECHNIQUE DE L'IMPUTATION STATIQUE OU *COLD DECK*

122. Selon la technique d'imputation statique ou *cold deck*, le programme de vérification attribue une réponse déterminée à un élément d'information manquant à partir d'un ensemble de valeurs prédéterminées, ou impute la réponse suivant une règle de proportionnalité à partir d'une distribution de réponses valides. Dans la méthode dite *cold deck*, le programme n'actualise pas le jeu initial de variables. Ainsi, les valeurs utilisées ne diffèrent pas de celles de la matrice statique initiale, une fois terminé le traitement des enregistrements concernant la première, la deuxième, la dixième ou toute autre personne. Les valeurs initiales permettent d'effectuer les imputations nécessaires pour toute donnée manquante.

123. L'imputation statique est une méthode stochastique, ainsi que l'imputation dynamique, mais les valeurs utilisées restent inchangées dans le temps. Cette approche est décrite à l'annexe V.

124. Parfois, l'imputation statique repose sur l'utilisation d'un rapport, et attribue les réponses selon des proportions prédéterminées. Afin d'illustrer ce type de répartition proportionnelle des réponses, considérons un tableau de données valides, c'est-à-dire de données correspondant à des réponses effectivement fournies et non manquantes,

concernant la durée du travail hebdomadaire des hommes de 33 ans travaillant dans le secteur de l'agriculture : cette durée était de 50 heures par semaine pour 25 % d'entre eux, de 60 heures pour 40 % et de 70 heures pour 35 %. Dans ce cas, les valeurs manquantes ou non valides de la durée de travail hebdomadaire des hommes de 33 ans travaillant dans le secteur de l'agriculture serait remplacée dans 25 % des cas par 50 heures, dans 40 % par 60 heures et dans 35 %, par 70 heures. Toutefois, à moins de disposer de données fiables provenant de recensements, d'enquêtes ou d'une autre source, l'application de cette technique exige au préalable l'exploitation des réponses valides fournies au recensement en cours, solution qui n'est pas nécessairement viable du point de vue économique ou pratique.

2. TECHNIQUE D'IMPUTATION DYNAMIQUE OU *HOT DECK*

125. Une autre méthode utilisable pour éliminer du jeu de données les éléments d'information réputés « inconnus » est celle de l'imputation dynamique ou technique dite de la donne variable (ou *hot deck*) qui consiste à attribuer des valeurs à des données non disponibles, inconnues, incorrectes ou incohérentes. Le United States Bureau of the Census a initialement mis au point la méthode, mais d'autres administrations ont depuis lors ajouté des perfectionnements. La technique de l'imputation dynamique utilise une ou plusieurs variables pour évaluer la réponse probable lorsque le jeu de données comporte une réponse non spécifiée (ou parfois plusieurs valeurs non spécifiées). Cette technique est maintenant de plus en plus répandue pour les vérifications des recensements en raison de sa simplicité et du fait de la reproductibilité et de l'élimination des erreurs des résultats. En outre, grâce à l'élimination des valeurs non spécifiées, il est plus facile de mettre en évidence les tendances entre des enquêtes et des recensements successifs, puisque l'analyste n'est plus tenu de traiter au cas par cas le problème des valeurs inconnues.

126. Dans le cas de l'imputation dynamique, les données connues concernant des individus présentant des caractéristiques similaires déterminent les informations les plus pertinentes à utiliser lorsque un (ou plusieurs) élément d'information concernant un autre individu n'est pas connu. Parmi ces caractéristiques figurent le sexe, l'âge, le lien avec le chef de ménage, la situation économique et le niveau d'instruction. La matrice d'imputation proprement dite est constituée d'un ensemble de valeurs que l'on peut comparer à la donne d'un jeu de cartes. Ces matrices permettent d'enregistrer, puis de fournir les informations utilisées en présence de valeurs inconnues. Cette donne est modifiée en permanence par des opérations de mise à jour et/ou « en battant les cartes suivant une procédure logique de façon à ce que les imputations évoluent au cours du traitement des données, d'où le terme de *hot deck* (donne variable).

127. A titre d'exemple simple, supposons qu'une valeur unique soit enregistrée en tant que donne. Par exemple, si le sexe d'une personne est non valide pour une raison ou une autre, une valeur initiale arbitraire (de sexe masculin ou féminin) détermine une valeur initiale. La valeur de départ ainsi obtenue est attribuée à la première personne dont la va-

³ D'après les résultats des questionnaires reçus de quelque 130 pays et régions par la Division de statistique des Nations Unies, les appellations *hot deck* et *cold deck* ne sont généralement pas connues. Toutefois, certains pays ont appliqué ces méthodes à la vérification de leurs données.

FIGURE 10. Exemple de ménage utilisé pour fournir les données d'entrée d'une imputation dynamique

Numéro d'identification	Lien avec le chef de ménage	Sexe	Age
1	1	1	39
2	2	2	35
3	3	1	13
4	3	9	10
5	4	2	40
6	4	1	99
7	4	2	13
8	5	9	99
9	5	1	44
10	5	2	36

Note : 9, 99 = information manquante.

riable « sexe » n'est pas spécifiée. En revanche, en cas de validité de la valeur enregistrée pour la variable sexe de la première personne, cette même valeur est substituée à la valeur de départ. Si ensuite la valeur de la variable sexe de la deuxième personne n'est pas connue, alors la matrice d'imputation lui attribue la valeur enregistrée. Dans ce cas, le sexe imputé est donc celui de la première personne. En principe, lorsque le programme de vérification établit le caractère acceptable de la valeur d'une rubrique, il place cette valeur dans la matrice d'imputation. Par contre, lorsqu'il établit le caractère inacceptable d'une valeur, le processus d'imputation lui substitue la valeur valide contenue dans la matrice d'imputation.

128. La méthode de l'imputation dynamique (*hot deck*) décrite ci-dessus pose notamment le problème suivant : si deux rubriques différentes ont une valeur inconnue, l'attribution de réponses valides risque de ne pas utiliser le même individu donneur. Chaque valeur peut provenir d'une personne « réelle », mais il peut s'agir de personnes différentes. Une meilleure façon de procéder consisterait à attribuer simultanément une valeur aux deux variables, à partir de la même personne. La programmation de ces matrices complexes risque toutefois de présenter certaines difficultés.

129. Les données ci-dessus (figure 10) illustrent le cas d'un ménage regroupant un ensemble de 10 individus. Les

valeurs 9 et 99 attribuées au sexe et à l'âge désignent des informations manquantes. Bien que d'autres variables soient utilisables pour effectuer les imputations, par exemple le niveau d'instruction et la profession, elles ne sont pas prises en compte dans cet exemple succinct.

130. Si la valeur initiale utilisée dans la matrice d'imputation SEXARRAY est « sexe masculin » (code = 1) la matrice d'imputation deviendra : SEX = 1.

131. Après traitement des données de la personne n° 1, la valeur reste égale à 1. Toutefois, après traitement des données concernant la deuxième personne, il lui est substitué la valeur 2, puisque la personne considérée est de sexe féminin. La variable sera alors égale à : SEXARRAY = 2.

132. Dès qu'une valeur valide est saisie en ce qui concerne le sexe d'une personne dont les données sont traitées, le code correspondant au sexe de cette personne remplace la valeur stockée dans la matrice d'imputation. Lorsque les données de la troisième personne sont traitées, l'imputation rétablit la valeur 1 (sexe masculin).

133. Dans le cas de la quatrième personne, la valeur de la variable « sexe » est inconnue : le programme de vérification examine la valeur contenue dans la matrice d'imputation qui correspond en l'occurrence au sexe masculin, et remplace la valeur inconnue par la valeur contenue dans la matrice d'imputation. La personne n° 5 est de sexe féminin,

FIGURE 11. Matrice initiale d'imputation statique de l'âge d'après le sexe et le lien avec le chef de ménage

	Lien avec le chef de ménage				
	Chef de ménage (1)	Conjoint (2)	Fils/fille (3)	Autre parent (4)	Personne non apparentée (5)
Homme (1)	35	35	12	40	40
Femme (2)	32	32	12	37	37

et la valeur correspondante remplace la valeur précédente contenue dans la matrice d'imputation provenant de la personne n° 3 (sexe masculin). Ce processus se poursuit jusqu'à la personne n° 8.

134. Le programme de vérification utilise à nouveau la procédure d'imputation et la personne n° 8 se voit attribuer la valeur « sexe féminin », puisqu'il s'agit de la valeur correspondante de la matrice d'imputation provenant de la personne n° 7. Le programme de vérification a utilisé à deux reprises la matrice d'imputation pour déterminer des valeurs : une première fois pour obtenir une valeur « sexe masculin » et une deuxième fois pour obtenir une valeur « sexe féminin ». Puisque les fréquences respectives des sexes sont sensiblement identiques, à long terme le processus d'imputation utilise chaque valeur de cette variable, sensiblement une fois sur deux. Après traitement des données concernant chacune des 10 personnes, la variable prendra la valeur : SEX ARRAY = 2.

135. Il existe par ailleurs d'autres modalités plus complexes d'utilisation de cette procédure d'attribution d'une valeur à la variable sexe au moyen d'une matrice d'imputation. Par exemple, la vérification peut utiliser le lien avec le chef de ménage et le sexe pour faciliter la détermination de l'âge d'un individu. Considérons la liste partielle suivante des codes attribués au lien avec le chef de ménage :

- 1 = Chef de ménage
- 2 = Conjoint
- 3 = Enfant
- 4 = Autre parent
- 5 = Personne non apparentée

136. Le traitement peut créer pour la variable âge des valeurs initiales pouvant rendre compte de la situation réelle des liens, pour chaque sexe. Ces valeurs ne seront cependant pas très importantes, puisque le programme de vérification les remplacera très certainement avant de les utiliser. De plus, le programme comporte nécessairement l'imputation de nombreuses valeurs, de telle sorte que peu de valeurs initiales ont une incidence sur les tableaux obtenus en défini-

tive. Ces valeurs pourraient se présenter tel qu'indiqué à la figure 11.

137. Examinons le cas des dix personnes du ménage de la figure 10. Puisque la première personne de notre échantillon figure dans la liste en tant que chef de ménage (code = 1) et qu'il s'agit d'un individu de sexe masculin (code = 1), son âge (39) se substitue au premier élément d'information (coordonnées 1,1) au cours de la procédure d'imputation. La « donne » contient alors les valeurs indiquées à la figure 12.

138. La deuxième personne de l'échantillon est le conjoint (code = 2), qui est de sexe féminin (code = 2); aussi, son âge (35) remplace-t-il la valeur figurant à la deuxième ligne et à la deuxième colonne, se substituant aux valeurs de la donne. Les âges des autres personnes du ménage sont également substitués aux valeurs contenues dans la matrice d'imputation, jusqu'à la personne n° 5.

139. Comme on peut le constater, la précédente procédure d'imputation de la variable sexe a attribué la valeur 1 à la personne n° 4. Puisque le programme de vérification exige l'imputation d'une valeur à la variable sexe, la matrice d'imputation n'est pas actualisée au moyen de l'âge de cette personne. Le programme de vérification procédera à une mise à jour uniquement avec les valeurs tirées des enregistrements dans lesquels le sexe et le lien avec le chef de ménage ont des valeurs initialement correctes dans les deux cas. Pour le traitement des données de la personne n° 6, le programme constate toutefois que la valeur de l'âge n'est pas spécifiée. Il s'agit d'une personne de sexe masculin, figurant parmi les « autres parents » du chef de ménage. Le programme de vérification utilise par conséquent l'élément de la matrice d'imputation relatif aux individus de sexe masculin de la catégorie « autre parent » définie quant au lien avec le sexe de ménage (quatrième colonne de la première ligne), et attribue la valeur de l'âge pour cette catégorie (« autre parent de sexe masculin », dans ce cas, 40).

140. Aucune donnée de sexe, ni d'âge n'est mentionnée dans le cas de la personne n° 8. Le programme de vérification impute à la variable sexe la valeur sexe féminin, puis at-

FIGURE 12. Exemple de matrice d'imputation dynamique à la suite d'une modification

	<i>Lien avec le chef de ménage</i>				
	<i>Chef de ménage (1)</i>	<i>Conjoint (2)</i>	<i>Fils/fille (3)</i>	<i>Autre parent (4)</i>	<i>Personne non apparentée (5)</i>
Homme (1)	39	35	12	40	40
Femme (2)	32	32	12	37	37

FIGURE 13. Exemple de matrice d'imputation dynamique à la suite de plusieurs modifications

	<i>Lien avec le chef de ménage</i>				
	<i>Chef de ménage (1)</i>	<i>Conjoint (2)</i>	<i>Fils/fille (3)</i>	<i>Autre parent (4)</i>	<i>Personne non apparentée (5)</i>
Homme (1)	39	35	13	40	44
Femme (2)	32	35	12	13	36

tribue l'âge en fonction du sexe et du lien avec le chef de ménage. (5) Dans ce cas, l'âge est de 37 ans.

141. Bien que le programme de vérification ait imputé la valeur de l'âge à partir du lien connu avec le chef de ménage, il a utilisé pour les autres variables une valeur préalablement attribuée au sexe. En l'occurrence, l'usage des valeurs déjà attribuées pour de nouvelles imputations est un exemple de procédure de vérification sujette à caution (voir section 3, *d* ci-dessous). Il serait en effet préférable de chercher d'autres données connues, par exemple l'état matrimonial, pour procéder aux imputations.

142. Après la personne n° 10, les valeurs de la matrice d'imputation sont indiquées à la figure 13. Dans cet exemple, les deux imputations ont utilisé la matrice statique initiale. Les opérations d'imputation n'utiliseront qu'un petit nombre de valeurs initiales, sinon aucune. Dans la majorité des cas, les valeurs attribuées utilisées proviendront de la population recensée.

3. PROBLÈMES POSÉS PAR L'IMPUTATION DYNAMIQUE (MÉTHODE *HOT DECK*)

a) *Considérations géographiques*

143. Si le programme de vérification attribue les valeurs manquantes selon la méthode de l'imputation dynamique, il doit alors chercher à utiliser des données classées en fonction de la zone géographique de moindre étendue. Cette procédure doit en principe augmenter la probabilité d'obtention d'une réponse exacte, en raison de l'homogénéité relative normalement observée, parmi les gens qui vivent dans la même unité géographique restreinte, eu égard à leurs différentes caractéristiques, notamment démographiques. Sans homogénéité de la population, il n'y aura pas de corrélation et l'équipe de vérification devra examiner les variables au cas par cas.

b) *Utilisation des rubriques connexes*

144. Avant de faire appel à la méthode d'imputation dynamique pour obtenir les valeurs manquantes, il convient de chercher à utiliser les rubriques connexes pour attribuer une valeur susceptible d'être correcte. Par exemple, en cas d'omission concernant l'état matrimonial d'une personne, le programme de vérification établira si la personne en question est associée à un conjoint au sein du ménage. Si tel est le cas, le programme attribuera le code relatif aux personnes mariées, sans utiliser de matrice d'imputation. Toutefois, faute de données utilisables de ce type, le programme devra sans doute faire appel à une matrice d'imputation.

c) *Effet de l'ordre des variables sur les matrices*

145. Les bureaux nationaux de recensement/statistique qui utilisent les matrices d'imputation, devraient réfléchir au choix des variables dont ils ont besoin lorsqu'ils établissent l'ordre dans lequel s'effectuent les vérifications. Pour les variables démographiques, les bureaux de recensement souhaiteront procéder d'emblée à la vérification des données de sexe et d'âge, pour pouvoir utiliser ces valeurs dans les autres matrices d'imputation. La vérification générale ne de-

vrait pas utiliser de variables non vérifiées dans les matrices d'imputation, bien que la plupart des progiciels acceptent des colonnes intitulées « valeurs inconnues ». Les matrices d'imputation ultérieures peuvent utiliser les données dûment vérifiées. Toutefois, dans la mesure du possible, les bureaux de statistique devraient envisager d'exclure les données vérifiées de la matrice d'imputation.

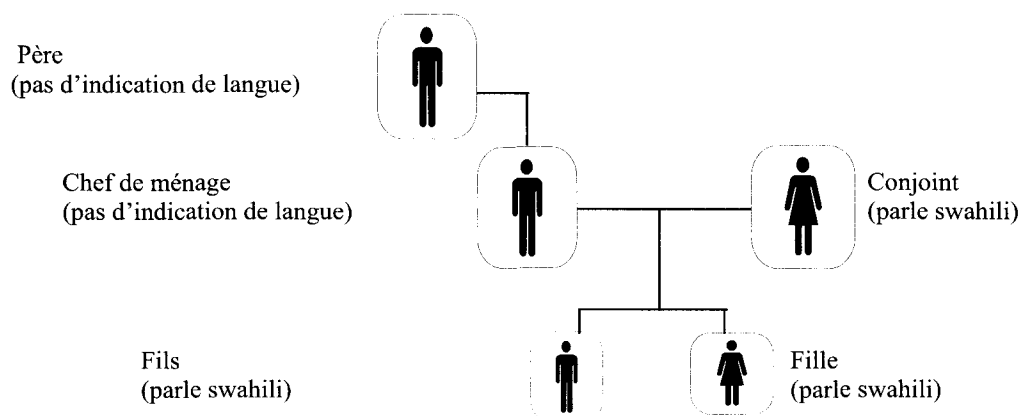
146. Par exemple, si le programme de vérification impute l'âge en fonction du sexe et du lien avec le chef de ménage, les éléments de la rangée correspondante de cette matrice d'imputation (sexe et lien avec le chef de ménage) ne devraient pas être mis à jour en cas d'imputation de l'une ou l'autre des variables sexe et lien avec le chef de ménage. En règle générale, le progiciel de vérification devrait introduire une valeur de l'âge correspondant aux valeurs appropriées des variables sexe et lien avec le chef de ménage uniquement lorsque les données d'âge, de sexe et de lien avec le chef de ménage sont toutes valides et cohérentes. Cependant, l'usage de données corrigées est parfois inévitable pour d'autres raisons.

d) *Complexité des matrices d'imputation*

147. Le bureau national de recensement/statistique augmente la probabilité d'obtention d'une valeur « correcte » et cohérente au moyen de la matrice d'imputation, en augmentant le niveau de détail de cette dernière. Ainsi, le programme pourrait imputer l'état matrimonial à partir du seul lien avec le chef de ménage. Toutefois, la probabilité de veuvage ou de divorce augmentant avec l'âge, l'imputation de l'état matrimonial sur la base de l'âge et du lien avec le chef de ménage est donc parfaitement fondée. Partant de l'âge et du lien avec le chef de ménage de la personne considérée, le programme de vérification utilise l'« état matrimonial » contenu dans la matrice d'imputation et provenant d'une personne dotée des mêmes caractéristiques, figurant dans l'enregistrement valide qui précède immédiatement.

148. Or, la procédure décrite ci-dessus risque de créer de nouveaux problèmes. Le bureau national de recensement/statistique vérifie généralement les rubriques du questionnaire dans un ordre fixe, l'âge à la suite de l'état matrimonial, selon une approche descendante. Si tel est le cas, lorsque l'état matrimonial et l'âge ont tous les deux été omis, il est impossible de prendre la valeur de la variable « état matrimonial » provenant du fichier qui précède immédiatement et comportant des valeurs identiques de l'âge et du lien avec le chef de ménage. De ce fait, le programme risque de ne pas pouvoir déterminer la catégorie d'âge relative à cet enregistrement. Une autre solution consisterait à prévoir dans la matrice d'imputation une rangée ou une colonne correspondant aux rubriques « non indiqué ». Or, deux arguments vont à l'encontre de ce choix. D'une part, les cases « non indiquées » correspondant à la même combinaison sont si peu nombreuses qu'il serait difficile de mettre à jour la matrice pour la rubrique manquante. D'autre part, il est pratiquement impossible d'obtenir des valeurs d'imputation statique (*cold deck*) appropriées pour ces combinaisons de valeurs « inconnues » puisqu'elles n'existent pas dans le monde « réel ».

Figure 14. Exemple de défauts d'indication de la langue du chef de ménage et du père du chef de ménage



149. La solution au problème présenté ci-dessus alourdit le travail de traitement des données, mais aboutit à un résultat moins entaché d'erreurs ou plus « pur » du point de vue statistique. Le programme de vérification contrôle tout d'abord la validité des codes attribués aux rubriques. Si l'enregistrement relatif à la personne considérée ne comporte pas un code valide attribué à la rubrique en question, alors la matrice d'imputation n'utilise pas ladite rubrique pour cet enregistrement. Les personnes chargées du traitement des données peuvent faciliter cette opération en mettant au point une matrice d'imputation simplifiée. Considérant à nouveau l'exemple précédent, au cas où le programme de vérification doit imputer une valeur à la variable « état matrimonial » (en raison de l'absence de cette donnée), la matrice d'imputation aura normalement les deux dimensions suivantes : âge et lien avec le chef de ménage. Si, après vérification, le programme établit qu'aucun code valide n'est attribué à la variable âge, il impute alors l'état matrimonial en fonction du seul lien avec le chef de ménage. Puisque la vérification concernant le lien avec le chef de ménage précède celle qui concerne l'état matrimonial, le code de lien avec le chef de ménage sera valide. Le programme utilise ces mêmes principes pour toutes les procédures d'imputation dynamique.

e) *Etablissement de la matrice d'imputation*

150. En collaboration avec les informaticiens, les spécialistes du sujet ont également la possibilité d'établir les matrices d'imputation appropriées (certaines équipes de vérification en utilisent plusieurs). Seules les réponses valides servent à la mise à jour des matrices, les équipes de vérification s'abstenant par ailleurs d'utiliser les valeurs déjà attribuées ou imputées. Il incombe aux spécialistes du sujet comme aux informaticiens de vérifier la cohérence et le caractère exhaustif des spécifications de vérification et des valeurs d'imputation dynamique (*hot deck*).

151. Il faut consacrer tout le temps nécessaire et réfléchir soigneusement à la mise au point d'une matrice d'imputation; il convient notamment d'étudier les possibilités

d'utilisation des enregistrements administratifs et des résultats des enquêtes ou recensements précédents. Même au terme de ce travail, les équipes de vérification ont intérêt à ne pas utiliser au hasard les matrices d'imputation. En l'absence de cohérence interne individuelle, l'harmonisation des matrices d'imputation exige un travail considérable. Par ailleurs, lorsque ces mêmes matrices n'utilisent pas des conventions normalisées, le personnel chargé de la vérification doit les examiner une à une.

152. Bien que, pour les exemples cités dans le présent manuel, chaque case des matrices d'imputation contienne une valeur et une seule, certaines équipes de vérification conservent pour chacune plusieurs valeurs possibles, ce qui a pour effet d'introduire une dimension supplémentaire : si les âges de chacun des enfants d'une famille sont inconnus, par exemple dans une famille comptant quatre enfants de sexe masculin, l'ordinateur n'attribuera pas la même valeur quatre fois et s'abstiendra donc de « créer » des quadruplés; il y aura en revanche attribution de quatre âges différents. Toutefois, même dans ce cas, la même valeur peut être attribuée à plusieurs reprises, selon le contenu des matrices d'imputation.

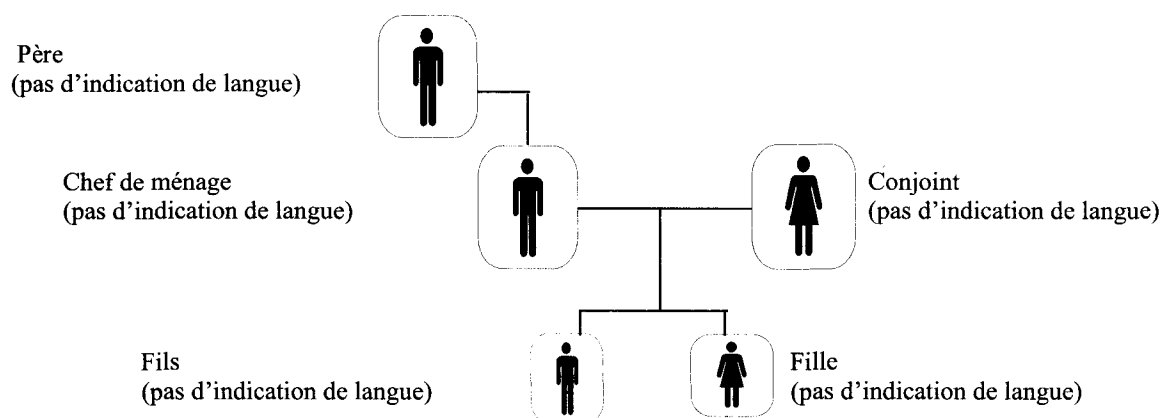
f) *Matrices d'imputation normalisées*

153. L'utilisation de matrices d'imputation normalisées permet de simplifier le processus de vérification. Il est possible de mettre à l'essai et d'appliquer rapidement des matrices d'imputation de dimensions normalisées et se reportant à différentes variables sociales et économiques, telles que le groupe d'âge et le sexe.

154. Par exemple, le bureau national de recensement/statistique peut souhaiter créer une matrice d'imputation permettant d'attribuer le code correspondant à la langue lorsque aucun code n'est indiqué. Le programme de vérification devra, de façon pratiquement certaine, chercher en premier lieu, au sein du ménage considéré, la présence d'une autre personne signalée en tant que locuteur d'une langue

FIGURE 15. Valeurs initiales d'une matrice d'imputation dynamique de la langue

Sexe	Age					
	Moins de 15 ans	15-29 ans	30-44 ans	45-59 ans	60-74 ans	75 ans et plus
Homme	Langue 1	Langue 1	Langue 1	Langue 1	Langue 1	Langue 2
Femme	Langue 1	Langue 1	Langue 1	Langue 1	Langue 1	Langue 2

Figure 16. Exemple de membres d'un ménage dans lequel aucune langue n'a été attribuée


donnée. Sinon, le programme peut choisir la langue d'une personne dont les données ont été précédemment vérifiées, de même sexe et du même groupe d'âge (après mise à jour de la matrice d'imputation en cas de validité des trois rubriques). Cette procédure permettra d'obtenir une valeur probable du code langue, puisque les personnes qui parlent la même langue ou des langues voisines se trouvent généralement dans des zones géographiquement proches.

155. Dans le schéma de la figure 14, aucune valeur n'est attribuée à la variable « langue » concernant le chef de ménage. Pour une raison quelconque, le lecteur optique ou l'opérateur n'a sans doute pas saisi l'indication ou le code de la langue, à moins qu'il n'y ait eu un autre problème. Or, puisque le conjoint et les enfants parlent tous swahili, cette langue peut être imputée au chef de ménage, ainsi qu'au père du chef de ménage dont la langue n'a pas non plus été indiquée.

156. Lorsqu'il n'y a d'indication de langue pour aucune des personnes du ménage, le programme de vérification doit procéder autrement; il examine les autres variables susceptibles de fournir une estimation indirecte de la langue utilisée. Parfois, la race, le groupe ethnique ou le lieu de naissance peuvent donner une indication de la langue à imputer; en présence d'un indicateur de ce type, l'équipe de vérification peut alors décider de l'utiliser pour déterminer la langue du chef de ménage. Sinon, le programme peut se servir des don-

nées d'âge et de sexe. Dans ce cas, la matrice d'imputation pourrait se présenter tel qu'indiqué à la figure 15.

157. Si l'on décide d'imputer une valeur, le programme attribue au chef de ménage un code de langue en fonction du groupe d'âge et du sexe. Dans ce cas, les données dans la matrice d'imputation proviendront uniquement des précédents chefs de ménage puisque toutes les autres personnes d'un ménage donné se verront attribuer le même code de langue que le chef de ménage.

158. A ce stade, si à l'intérieur du ménage personne ne déclare parler une langue particulière, le programme de vérification utilise la matrice d'imputation pour attribuer une langue au chef de ménage en s'appuyant sur l'âge et le sexe de ce dernier. La langue ainsi attribuée est celle qui a été identifiée le plus récemment dans le fichier de données comme étant parlée par un autre chef de ménage du même âge et du même sexe. Puisque la matrice d'imputation est « mise à jour » de façon continue au fur et à mesure de l'identification de données acceptables, la langue attribuée est susceptible d'être une langue parlée par l'ensemble de la communauté considérée.

159. Dès le début du processus de vérification les règles fixées donneront lieu à des exceptions. Ainsi, il faut veiller à prendre soin des éventuelles différences de langue lorsqu'on passe d'une zone géographique à l'autre. Dans certains pays, il faut par ailleurs prêter attention aux combinaisons

locales de locuteurs. Toutefois, même dans ce cas, à moins que certaines langues ne soient sous-déclarées de façon sélective, le processus d'imputation devrait conduire à des pourcentages sensiblement identiques de valeurs attribuées et non attribuées.

160. Une autre vérification pourrait porter sur la religion. Là encore, l'imputation de réponses concernant la variable religion est possible à partir des données d'âge et de sexe. Le programme de vérification poursuivra la mise à jour lorsque toutes les informations requises seront disponibles et utilisera les réponses fournies par la matrice d'imputation pour les affecter aux données « inconnues ». Cette matrice d'imputation présentera un aspect voisin de celle obtenue pour la variable langue, les données concernant la religion étant substituées à celles concernant la langue dans les cases correspondantes.

161. Il est à noter que cette description suppose une approche séquentielle de type descendant. Avec des méthodes sophistiquées, comme celle de Fellegi-Holt et la nouvelle méthode d'imputation (NIM, *New Imputation Method*) [voir Annexe V], les équipes de vérification réalisent simultanément tous les contrôles connexes. La procédure décrite suppose par contre l'existence d'un ordre approprié de réalisation des vérifications.

162. Nombre de caractéristiques économiques telles que l'appartenance à la population active, le nombre d'heures de travail au cours de la semaine passée et le nombre de jours de travail au cours de l'année écoulée, peuvent être imputées sur la base de caractéristiques similaires. Grâce à l'utilisation de matrices d'imputation similaires, le programme de vérification peut contrôler rapidement la valeur des caractéristiques économiques utilisées et l'ensemble du processus devrait être accéléré.

163. Il y a lieu d'observer qu'il est parfois difficile d'obtenir des données dûment vérifiées pour les premières

matrices d'imputation. D'ordinaire, la prise en compte de données non vérifiées dans une matrice d'imputation est a priori écartée; ainsi, le programme ne doit pas utiliser de données de sexe ou d'âge pour les affecter à la matrice d'imputation avant de les avoir contrôlées. Aussi les toutes premières matrices d'imputation utiliseront-elles des variables différentes n'exigeant pas de vérification ou des variables dont la valeur ne peut être modifiée. Quant à la première matrice d'imputation de données démographiques, le processus pourrait utiliser le nombre de personnes de l'unité d'habitation, celui-ci étant égal à 0 dans le cas des unités inoccupées.

164. Pour les vérifications des données sur l'habitat en règle générale, il serait également possible d'utiliser le nombre de personnes présentes dans les unités d'habitation en tant que dimension initiale de la première matrice d'imputation; l'équipe de vérification pourrait cependant modifier les ajustements concernant les unités d'habitation de manière à tenir compte des unités inoccupées. Par exemple, si la première vérification des données sur l'habitation porte sur le « matériau de construction des murs extérieurs » ou sur le « type de murs », les valeurs initiales pourraient être choisies d'après le nombre de personnes présentes dans l'unité d'habitation, notamment une valeur particulière correspondant au cas d'une unité d'habitation inoccupée.

165. Lorsque l'unité d'habitation est inoccupée, mais que la donnée « type de murs » est valide, le programme de vérification met à jour la première case avec la valeur attribuée au type de murs extérieurs. Si le type de murs est connu, pour une unité d'habitation occupée, le programme met à jour la case qui correspond au nombre de personnes qui occupent ladite unité de logement. Si le matériau de construction des murs extérieurs est inconnu, la matrice d'imputation attribuera toutefois une valeur à la donnée « matériau de construction des murs extérieurs » en fonction du nombre de personnes de l'unité d'habitation.

FIGURE 17. Exemple de données concernant un chef de ménage et un enfant avec omission de l'âge de l'enfant

<i>Emploi</i>	<i>Lien avec le chef de ménage</i>	<i>Age</i>	<i>Niveau de scolarité</i>	<i>Emploi</i>	<i>Profession</i>	<i>Nombre d'enfants</i>
1	1	40	12	1	33	BLANC
3	3	X	7	BLANC	BLANC	BLANC

Note : X = Données incorrectes
BLANK = sans objet

FIGURE 18. Exemple de données concernant un chef de ménage et un enfant avec omission de l'âge et du niveau de scolarité de l'enfant

<i>Emploi</i>	<i>Lien avec le chef de ménage</i>	<i>Age</i>	<i>Niveau de scolarité</i>	<i>Emploi</i>	<i>Profession</i>	<i>Nombre d'enfants</i>
1	1	40	12	1	33	BLANC
3	3	X	X	BLANC	BLANC	BLANC

166. Après avoir utilisé une première fois cette matrice d'imputation, l'équipe de vérification peut ensuite souhaiter passer à une autre caractéristique de l'habitat, telle que le « type de toiture » ou les « modalités de jouissance ». Quelle que soit la caractéristique choisie, elle doit être propre à chaque unité d'habitation et comporter une diversité suffisante pour que le même attribut ne soit pas retenu systématiquement. En effet, la sélection récurrente du même attribut risque de conduire à un type d'imputation pratiquement statique et non à une imputation dynamique (méthode *hot deck*). L'utilisation de la méthode dite d'imputation dynamique, par exemple dans le cas d'un « logement collectif » constitué de casernes militaires, risque d'entraîner l'utilisation répétée de la même valeur, si l'âge et le sexe sont les seules caractéristiques retenues. En l'occurrence tous les résidents seraient vraisemblablement de sexe masculin et se situeraient pour la plupart dans une tranche d'âge bien définie. Par conséquent, l'emploi de cette matrice particulière risque d'altérer la qualité des résultats. Lorsque la variable « modalités de jouissance » présente une diversité suffisante, avec des pourcentages suffisants de propriétaires et de locataires, son utilisation peut s'avérer efficace. Sinon le choix pourrait porter sur la variable « type de toiture ».

167. En règle générale, nombre d'équipes de vérification estiment que l'utilisation de matrices d'imputation de dimensions comparables permet de réduire les contrôles, d'obtenir plus rapidement les résultats et sans doute d'améliorer leur précision.

g) *Cas de non-utilisation
de la méthode d'imputation dynamique*

168. Si l'équipe de vérification décide de ne pas du tout utiliser la méthode d'imputation dynamique, l'ordre de réalisation des vérifications conserve son importance. Par exemple, l'âge est lié à de nombreuses rubriques, notamment au lien avec le chef de ménage, au niveau d'instruction, à l'emploi et à la fécondité (pour les femmes). Considérons les membres du ménage décrits à la figure 17.

169. Les données concernant la personne n° 3 indiquent le code 3 de lien avec le chef de ménage (enfant), mais ne mentionnent pas d'âge. Afin de déterminer l'âge, le programme de vérification peut utiliser la différence d'âge entre le chef de ménage et l'enfant (valeur obtenue par imputation statique (méthode *cold deck*) ou d'après une unité précédemment vérifiée). Si cette différence est égale à 25 par exemple, l'âge de l'enfant est égal à 15 ans (âge du chef de ménage, soit 40 ans, moins différence d'âge de 25).

170. Le nombre d'années de scolarité est également connu et égal dans ce cas à 7 ans. L'âge de 15 ans est en effet parfaitement susceptible de correspondre à ce niveau d'instruction. Puisque la fourchette de valeurs appropriées du nombre d'année de scolarité correspondant à un âge donné est moins étendue que la fourchette des âges correspondant à une différence d'âge donnée entre chef de ménage et enfant, il est préférable de vérifier en premier si le nombre d'années de scolarité est adéquat. Si ce nombre d'années est mentionné, une différence d'âge obtenue par imputation statique (*cold deck*) ou dynamique (*hot deck*) peut servir à définir un

âge approprié. Si le nombre d'années n'est pas connu, alors la différence d'âge entre chef de ménage et enfant peut servir à attribuer l'âge.

171. Or, même la différence d'âge peut même figurer parmi les informations manquantes. De fait, dans la plupart des pays le niveau de formation risque davantage d'être omis que l'âge. L'exemple ci-dessous décrit les différentes étapes du processus que peut suivre l'équipe de vérification en cas d'omission de l'âge et du niveau d'instruction.

172. Les données de la figure 18 ne comportent aucune information d'âge ni de niveau d'instruction, mais contiennent par contre d'autres informations. La personne n° 3 n'est pas suffisamment âgée pour occuper un emploi, mais elle est trop jeune pour avoir eu des enfants (ou cette personne est de sexe masculin). D'après les indications disponibles sur l'emploi, une série de valeurs obtenues par imputation statique permet de déterminer un âge, mais celui-ci sera inférieur à l'âge minimum acceptable pour occuper un emploi. Sinon, l'équipe de vérification procède à une imputation dynamique, la valeur de la matrice d'imputation permettant d'obtenir une valeur de l'âge. La valeur retenue sera vraisemblablement obtenue d'après l'âge du chef de ménage, en tant que variable permettant de maintenir la cohérence des données. Par exemple, si le chef de ménage est âgé de 20 ans et non de 40, il ne conviendrait évidemment pas d'attribuer un âge de 14 ans à la personne n° 3. Lorsque l'âge est fixé, alors il est possible de déterminer également le niveau d'instruction, ce dernier devant être en effet cohérent avec l'âge et la situation du point de vue de l'emploi.

173. Si l'équipe de vérification décide d'imputer toutes les rubriques considérées ou presque, elle doit alors définir une stratégie d'élaboration logique du processus de vérification. En ce qui concerne les données démographiques, la vérification devrait commencer par tous les éléments comportant potentiellement des valeurs inconnues. Les équipes de vérification doivent mettre à profit toute autre information disponible pour faciliter la prise en compte de chaque élément d'information dans la toute première matrice d'imputation. La définition des caractéristiques détaillées des matrices d'imputation est certes particulièrement liée au pays considéré, mais tous les bureaux de recensement/statistique disposent vraisemblablement à cet effet de certaines informations.

174. Nombre de progiciels de vérification conservent au fur et à mesure de leur déroulement, la valeur du nombre de personnes de l'unité d'habitation. Une matrice d'imputation pour la détermination des valeurs inconnues de la variable « sexe » par exemple, permet d'imputer la valeur de cette rubrique d'après le nombre d'occupants de l'unité de logement. Par conséquent, la valeur à choisir initialement pour une personne de sexe non spécifié ou non valide dans le cas d'un ménage constitué d'une personne unique, pourrait être « sexe masculin ». Dans le cas d'un ménage constitué de deux personnes, la valeur initiale pourrait être « sexe féminin ». Enfin, pour un ménage de trois personnes la valeur pourrait être « sexe masculin », etc. La matrice d'imputation ne serait utilisée qu'en dernier ressort, une fois terminés tous les contrôles et les vérifications de cohérence, portant no-

tamment sur le sexe du chef de ménage, et du conjoint et la présence de données sur la fécondité.

h) *Quelle doit être la taille des matrices d'imputation ?*

175. La plupart des progiciels informatiques peuvent traiter des matrices d'imputation multidimensionnelles. Les observations suivantes doivent être prises en compte lors de la création des matrices d'imputation.

i) *Problèmes posés par une taille excessive de la matrice d'imputation*

176. Pour certains bureaux nationaux de recensement/statistique, l'excès de perfectionnisme des vérificateurs est l'un des principaux problèmes rencontrés dans le cadre de la collaboration entre spécialistes du sujet et informaticiens. Les concepteurs des progiciels de vérification se laissent volontiers emporter par leur inspiration de telle sorte que la programmation prend beaucoup plus de temps que nécessaire et ralentit le dépouillement du recensement et de l'enquête. L'équipe de vérification doit décider par exemple en vue de la détermination de l'âge, qu'il faut également tenir compte du « nombre d'enfants » dans le cas des individus de sexe féminin, outre les variables « sexe » « niveau d'instruction » et « appartenance à la population active ». Le fait d'ajouter la variable « nombre d'enfants » est susceptible d'améliorer légèrement l'estimation de l'âge, bien que cette amélioration ne justifie pas nécessairement la complexité

accrue de la programmation. Il incombe donc aux équipes de vérification de décider des dimensions de la matrice d'imputation qui fournira les meilleurs résultats, tant en termes de précision que d'efficacité. Les matrices d'imputation de taille excessive (comportant un trop grand nombre de cases) ne peuvent être parfaitement actualisées, auquel cas, des valeurs obtenues par imputation statique risquent d'être utilisées à tort à leur place.

ii) *Connaissance précise des tâches effectuées au moyen de la matrice d'imputation*

177. Outre la taille excessive des matrices d'imputation, les chemins de vérification risquent d'être compliqués. Il importe de veiller à ce que spécialistes du sujet et informaticiens soient en mesure de suivre l'exécution de toutes les séquences. Ils doivent conjointement s'assurer du fait que la matrice d'imputation exécute la tâche voulue. Là encore, la collaboration entre spécialistes et informaticiens est indispensable afin de vérifier la bonne mise en œuvre de chaque variable ou de chaque dimension de la matrice d'imputation. De plus, ils doivent vérifier que toutes les combinaisons fonctionnent correctement.

iii) *Problèmes posés par une matrice d'imputation de taille insuffisante*

178. La matrice d'imputation est trop petite, si les variables sont trop peu nombreuses ou si, en raison des regroupe-

FIGURE 19. Exemple de jeu de valeurs utilisées dans un tableau d'imputation statique et codes imputés correspondants

22	A01-AGE-FM-SEXRL (2,6)						
23	<i>Chef de ménage</i>	<i>Conjoint</i>	<i>Enfant</i>	<i>Autre parent</i>	<i>Parent</i>	<i>Non indiqué</i>	
24	40	40	10	20	65	20	Homme
25	40	40	10	20	65	20	Femme

40	if AGE = 0:98
41	Let A01-AGE-FM-SEXRL (SEX, RELATIONSHIP) = AGE
42	else
43	Message « Age is unknown, so imputed » AGE
44	Write « Age is unknown, so imputed, Age = » AGE
45	Impute AGE = A01 (SEX, RELATIONSHIP)
46	Message « AGE is now known » AGE
47	end-if

FIGURE 20. Exemple de rapport indiquant le nombre d'imputations par erreur

<i>Décompte</i>	<i>Numéro d'erreur</i>	<i>Message</i>	<i>Numéro de ligne</i>
-	14-1	Nombre excessif d'enfants par femme	2 629
-	14-2	Nombre excessif d'enfants par femme	2 645
2	14-3	Non-indication des jeunes gens présents	2 669
2	14-4	Non-indication des jeunes filles présentes	2 678
33	14-5	Non-indication du mois de la dernière naissance	2 723
7	15-6	Aucun nombre d'enfants; différence d'âge correcte entre la mère et l'enfant	2 892

Note : dans ce cas la mention « 14 » désigne simplement la rubrique n°14 d'une série de données; les erreurs sont numérotées séquentiellement.

ments pratiqués (groupes d'âge ou niveaux de formation en nombre insuffisant), la même valeur est imputée à plusieurs reprises avant d'être mise à jour. Par exemple, sans prise en compte de la variable sexe, dans un tableau d'âges, tous les enfants d'une famille sont davantage susceptibles de se voir attribuer le même âge, lorsque cette donnée est inconnue. Avec le concours des informaticiens, les spécialistes du sujet devraient s'employer à tester les matrices d'imputation pour toutes les combinaisons possibles et veiller à ce qu'aucune combinaison n'intervienne trop fréquemment.

iv) Rubriques difficiles à corriger au moyen de matrices d'imputation

179. Certaines rubriques, telles que « la profession » et le « secteur d'activité » se sont avérées manifestement difficiles à corriger. Tandis que l'utilisation de matrices d'imputation distinctes pour la profession et le secteur d'activité risquent d'aboutir à des résultats incohérents, le coût et la difficulté d'un travail systématique de vérification de tous les couples de profession et de secteur d'activité peuvent être considérables. Ainsi, la détection de coiffeurs qui travaillent dans des usines de transformation du poisson doit donner lieu à une forme ou une autre de vérification. De plus, la multiplicité des professions et des secteurs d'activité risque de compliquer énormément le processus d'imputation dynamique. De fait, pour certaines rubriques, l'équipe de vérification conviendra vraisemblablement de l'inefficacité des contrôles et choisira de leur substituer l'utilisation des réponses « non indiqué » ou « non signalé ». Sinon, le recours à une méthode d'imputation statique (*cold deck*) pourra suffire.

4. CONTRÔLE DES MATRICES D'IMPUTATION

180. La matrice d'imputation utilisée dans un progiciel de vérification devrait présenter une structure de base proche de celle de la figure 19. Les spécifications de vérification doivent déterminer les tableaux de valeurs utilisés pour les imputations et déterminer le jeu de valeurs initiales obtenues par imputation statique.

a) Constitution de la matrice statique initiale

181. La procédure décrite ci-dessous met à jour la matrice d'imputation chaque fois qu'elle trouve une personne à laquelle sont associées des valeurs valides pour chacune des trois rubriques considérées, en l'occurrence le « lien avec le chef de ménage », le « sexe » et « l'âge ». Toutefois, lorsque le programme de vérification trouve une valeur non valide (ou absente) de la variable « sexe », la matrice d'imputation choisit une valeur d'après les codes valides attribués au lien avec le chef de ménage et au sexe (variables préalablement vérifiées).

b) Messages relatifs aux erreurs

182. Les progiciels de vérification doivent proposer plusieurs moyens pour garantir l'exécution effective et correcte des vérifications et des imputations. Deux de ces moyens, les instructions de message et les instructions d'écriture, sont passés en revue ci-dessous.

183. L'affichage d'un message, tel qu'indiqué ci-dessus à la figure 19, est une source d'information. Cette instruction génère des messages et des décomptes récapitulatifs particuliers (nombre total d'occurrences du message) pour divers niveaux d'agrégation géographique (par exemple, secteur de dénombrement, circonscription administrative secondaire, circonscription administrative principale), comme pour chaque questionnaire. Pour l'ensemble des questionnaires, un rapport récapitulatif peut se présenter comme suit (voir figure 20) :

184. Un rapport établi pour chaque questionnaire particulier (figure 21) indiquera le cas échéant le numéro du questionnaire, ainsi que tous les codes géographiques spécifiés. Il pourrait également indiquer la liste des erreurs détectées par le programme, classées par rubrique (dans ce cas l'âge) et par numéro de ligne du logiciel, tel qu'indiqué ci-dessous dans la colonne de droite. Dans cet exemple, l'âge était laissé en blanc, mais la matrice d'imputation a permis de calculer un âge de 48 ans, d'après les valeurs du lien avec le chef de ménage et du sexe concernant cette personne. En l'occurrence, l'âge en question n'était pas spécifié, mais l'instruction d'édition d'un message pourrait également mentionner cette information si nécessaire :

c) *Etablissement à la demande de listes d'erreurs*

185. Le logiciel peut également proposer une autre instruction, permettant une analyse plus détaillée des spécifications de vérification et du déroulement des opérations correspondantes. Cette instruction peut servir à présenter les données avant modification, puis toutes les modifications effectuées; enfin, elle présente l'enregistrement ou les enregistrements considérés, avec les modifications introduites. De cette façon, l'analyste peut acquérir la certitude du bon déroulement des vérifications selon tous les chemins définis. Les résultats obtenus peuvent se présenter tel qu'indiqué à la figure 22. La première ligne indique les variables (par exemple province, lien avec le chef de ménage, sexe, âge). Ensuite, les données d'entrée sont indiquées, puis l'erreur détectée (en l'occurrence pas d'indication d'âge), et enfin, les données obtenues après modification.

186. Cette procédure aide l'équipe de vérification à déterminer si le programme suit les chemins appropriés.

187. Ce type de contrôle est une partie importante du travail de vérification des recensements et des enquêtes. La méthode décrite ci-après est un moyen de contrôle des procédures de vérification. La marche à suivre pourrait consister à demander aux spécialistes de commencer systématiquement par la définition d'un ménage « parfait ». On entend par « ménage parfait » un ménage complet — chef de ménage, conjoint, enfant, autres parents et non-parents — avec indication de toutes leurs caractéristiques. Le ménage parfait doit satisfaire à toutes les vérifications sans donner lieu à aucune erreur. La procédure se poursuit tel qu'indiqué ci-dessous :

a) Le personnel chargé de traiter les données introduit une erreur unique dans chaque ménage, de façon séquentielle, correspondant à la séquence des spécifications de vérification et au déroulement du programme de contrôle;

b) L'analyste contrôle ensuite tous les chemins au début du processus de vérification;

c) Lorsque la vérification suit correctement tous les chemins, le personnel chargé de traiter les données applique

le programme à un échantillon du jeu de données complet, y recherche les anomalies et effectue les modifications éventuellement nécessaires;

d) Enfin, le personnel chargé du traitement exécute le programme sur la totalité du jeu de données.

188. Après avoir constaté le fonctionnement satisfaisant des messages et avoir apporté les modifications appropriées, le personnel informaticien peut décider de les désactiver au niveau du questionnaire. Si les pays de taille importante devaient traiter la totalité de leurs jeux de données en maintenant la publication des messages correspondant à chaque questionnaire, il en résulterait une quantité prohibitive de lignes d'impression et de papier. Toutefois, le rapport récapitulatif concernant ces messages doit être conservé, puisqu'il donne des indications intéressantes aux différents niveaux d'agrégation géographique. L'état imprimé ressemblera à celui de la figure 22.

189. Les vérifications automatisées comprennent généralement une procédure de sauvegarde. L'analyse rétrospective fait apparaître toutes les modifications apportées aux données ainsi que des repères indiquant les dites modifications et les valeurs substituées. La consultation de l'analyse rétrospective déterminera si les changements apportés sont suffisamment peu nombreux pour autoriser l'acceptation du groupe d'enregistrements.

190. Si une rubrique particulière comporte un nombre d'erreurs excessif, il est possible qu'elle n'ait pas fait l'objet de tests préalables adéquats, soit isolément, soit en rapport avec les autres rubriques. Parfois, les agents de recensement sont induits en erreur, et recueillent les informations sur la fécondité uniquement auprès des adultes de sexe masculin et non auprès des femmes. Si les données sont collectées systématiquement de cette façon, l'équipe de vérification devra sans doute demander aux programmeurs de transférer les données sur la fécondité des hommes aux femmes, à l'intérieur d'un couple marié. Sinon, elle ne peut pas faire grand-chose à ce stade pour corriger cette erreur.

191. Généralement, le programme de vérification doit être appliqué à plusieurs fichiers différents pour couvrir tou-

FIGURE 21. Exemple de rapport d'erreurs contenues dans un questionnaire

<i>Questionnaire ID : 01 01 017</i>	<i>Numéro de ligne</i>
AGE (1) = Age inconnu, donc imputation	# 46
AGE (1) =48 Age maintenant connu	

FIGURE 22. Exemple de liste imprimée d'erreurs supplémentaires portant sur plusieurs variables

	<i>Province</i>	<i>District</i>	<i>Chef de ménage</i>	<i>Lien avec le chef de ménage</i>	<i>Sexe</i>	<i>Age</i>
Données d'entrée	01	01	17	1	1	
Erreur	Age inconnu, donc imputation âge = BLANC					
Données vérifiées	01	01	17	1	1	48

tes les situations. En outre, le personnel informaticien devra effectuer des changements en raison des erreurs de syntaxe ou de logique. Même les informaticiens les plus expérimentés peuvent taper accidentellement un signe « supérieur à » au lieu d'un signe « inférieur à » et l'erreur n'est ensuite constatée qu'après plusieurs passages, puisque ce type de problème n'est pas nécessairement immédiatement apparent. De manière analogue de petites erreurs dans le schéma logique de traitement risquent de ne pas apparaître d'emblée. Là encore, spécialistes du sujet et informaticiens doivent travailler de concert à la résolution de ces problèmes, si possible dès le début du processus de vérification.

d) *Nombre de passages du programme de vérification*

192. En règle générale, il y a tout intérêt à exécuter trois fois un programme de vérification, tel qu'indiqué ci-dessous :

193. Le premier passage affecte aux matrices d'imputation des valeurs réelles et non les valeurs créées dans la matrice statique initiale. De nombreux pays utilisent des données d'une autre provenance — enquêtes ou recensements antérieurs ou données administratives — pour déterminer les valeurs à faire figurer dans un tableau d'imputation statique. Le personnel informaticien vérifie le jeu complet de données ou une grande partie, afin d'obtenir les valeurs à faire figurer dans la matrice d'imputation. Les données d'imputation statique tirées du jeu de données réelles ont plus de chances d'être exactes et actualisées. Les vérifications utilisent seulement 2 % du contenu de cette matrice statique initiale : le reste est constitué de valeurs d'imputation dynamique.

194. Le deuxième passage du programme procède à la vérification proprement dite et comporte plusieurs passages répétés afin de couvrir toutes les situations. A ce stade, le personnel informaticien devra introduire des modifications pour corriger les erreurs dues aux défauts de syntaxe ou de logique. Or, l'informaticien le plus expérimenté étant faillible et tel problème particulier ne se manifestant pas forcément immédiatement, il faut souvent plusieurs passages avant de constater les erreurs de ce type. De manière analogue, les petits défauts du schéma logique de traitement ne sont pas toujours apparents d'emblée.

195. Le troisième passage établit de façon certaine : 1) que le jeu de données ne comporte plus d'erreurs; et 2) que le programme de vérification n'en a pas introduit de nouvelles. Pour ce dernier passage, les états imprimés obtenus ne doivent plus faire apparaître aucune erreur. Si des erreurs persistent, le schéma logique de traitement est probablement défectueux et le personnel informaticien doit alors le modifier. De plus, ce passage indique généralement si la vérification a introduit accidentellement de nouvelles erreurs imputables au schéma logique de traitement.

5. INDICATEURS D'IMPUTATION

196. L'utilisation d'indicateurs d'imputation permet de conserver l'information relative à des données non vérifiées. Tel qu'indiqué plus haut, nombre d'équipes de vérification

sont préoccupées par la perte d'information potentielle en cas de modification des réponses non vérifiées. Lorsqu'une valeur est changée en raison d'une incohérence, les équipes de vérification souhaitent parfois conserver la ou les valeurs initiales afin de procéder à des études démographiques ou des analyses d'erreurs plus poussées, une fois le recensement terminé. Ainsi, spécialistes du sujet et programmeurs souhaiteront analyser différents aspects des données omises, non valides ou incohérentes. Les membres de l'équipe de vérification devront s'assurer de la cohérence et de la compatibilité des distributions imputées et non imputées, afin de déceler l'apparition éventuelle d'une erreur systématique dans le programme de vérification et d'imputation. Par exemple, les informaticiens peuvent utiliser incidemment uniquement les valeurs d'imputation statique parce que le programme a omis de mettre à jour la matrice d'imputation. En cas de réalisation d'un essai précensitaire national, l'équipe de vérification doit parfois étudier les liens entre certaines variables à la suite de l'essai pour mettre au point la version définitive du questionnaire. Lors des recensements précédents, alors que les micro-ordinateurs équipés de disques durs de grande capacité n'étaient pas encore répandus, nombre de bureaux de statistique n'avaient pas assez de place sur leurs bandes magnétiques ou sur leurs supports de stockage pour conserver une quantité supplémentaire de données; en revanche de nos jours, et dans la plupart des pays, la conservation des informations relatives à des données non vérifiées n'est plus problématique.

197. Certains pays décident de tenir à jour une simple variable comptable binaire en tant qu'indicateur relatif à chaque rubrique. Cette méthode est simple et elle utilise un octet unique pour chaque variable. Par exemple, lors du recensement de 1990, le United States Bureau of the Census a placé des indicateurs d'imputation pour chaque variable à la fin de chaque enregistrement, aussi bien pour les données de population que pour les données d'habitat. Ainsi, pour chaque variable concernant l'habitation, l'indicateur avait initialement la valeur 0, mais passait à la valeur 1 en cas de modification quelconque de la donnée élémentaire d'origine. Le programme utilisé alors ne conservait pas la valeur d'origine, bien que certains bureaux de statistique procèdent à la compilation desdites valeurs, soit pour chaque enregistrement, soit au niveau de l'agrégat.

198. Il existe d'autres méthodes de sauvegarde des réponses non vérifiées. Si l'on considère l'exemple de la figure 23, le bureau national de recensement/statistique a modifié l'âge d'un conjoint en le faisant passer de 70 à 40 ans au moyen d'une matrice d'imputation. Il est alors facile de placer la valeur pré-imputation (en l'occurrence 70 ans) dans la zone réservée aux indicateurs d'imputation et de réserver la variable utilisée dans les tableaux destinés à la publication pour la valeur attribuée en définitive, en l'occurrence 40 ans. Afin d'examiner les changements introduits dans le jeu de données, le bureau de statistique peut établir des distributions de fréquence ou des tabulations croisées des valeurs allouées et non allouées. Si à la suite de cette analyse de l'incidence des vérifications sur le jeu de données, les tableaux obtenus au terme de la vérification sem-

blent comporter des valeurs douteuses ou anormales, les équipes de vérification peuvent envisager une modification totale ou partielle du schéma de vérification.

199. La figure 24 se rapporte à une femme âgée de 13 ans dont les données indiquent qu'elle a un enfant (nombre d'enfant = 1). Or, l'équipe de vérification a convenu que l'âge minimum à la première naissance était de 14 ans et que la mention d'enfants nés de femmes plus jeunes correspondait plus vraisemblablement à une erreur qu'à un fait. Comme toujours cela soulève la question de savoir si ce type de cas représente un bruit dans le jeu de données ou bien une valeur réelle.

200. Selon les règles de vérification, le processus d'imputation « efface » l'information concernant le nombre d'enfants. Signalons que l'indicateur relatif au nombre d'enfants est un peu plus complexe, puisqu'il doit tenir compte d'un vide (BLANC) qui a été imputé, ainsi que des valeurs numériques. Supposons que les démographes souhaitent étudier les effectifs et les caractéristiques des personnes de 13 ans dont les données indiquent qu'elles ont eu un enfant. Le personnel informaticien peut enregistrer l'information d'origine dans une zone de l'enregistrement réservée aux indicateurs, généralement à la fin. Ensuite, la série de tableaux publiés omettra l'information concernant le nombre d'enfants de cette femme, mais cette information restera disponible en vue de travaux de recherche ultérieurs. Par la suite, notamment lors de l'organisation d'une enquête de suivi, ou à l'occasion du recensement suivant, les équipes de vérification peuvent utiliser l'information concernant les enfants nés de femmes de 13 ans, pour décider s'il faut abaisser l'âge limite de prise en compte.

201. Un des problèmes posés par l'utilisation des indicateurs d'imputation est la place considérable occupée dans l'ordinateur par la procédure décrite ci-dessus. Si les indicateurs répètent chaque variable, le jeu de données vérifiées occupera deux fois plus de place que le jeu de données non vérifiées. Pour de nombreux pays cette exigence serait inac-

ceptable dans la perspective d'un stockage à long terme. Données d'origine et vérifications pourraient cependant être enregistrées en vue de leur reconstitution ultérieure.

202. Les pays dont les populations sont très importantes préféreraient utiliser des indicateurs d'imputation en procédant par échantillonnage, en vue de travaux de recherche. Par exemple, un pays peut souhaiter créer un jeu de données concernant une unité d'habitation sur 100. Le programme de vérification affectera alors des indicateurs sur ce jeu de données réduit, ce qui facilitera l'évaluation de l'incidence de la vérification sur la qualité de données, ainsi que la détermination des différences entre données vérifiées et non vérifiées.

F. — Autres systèmes de vérification

203. La plus grande partie du présent manuel s'emploie à décrire l'application des méthodes de type descendant à la vérification automatisée des recensements et des enquêtes. De fait, peu de pays mettent en œuvre à cet effet une procédure différente et plus compliquée connue sous le nom de vérification à plusieurs variables (voir section C.2 ci-dessus). Fellegi et Holt (1976) ont été les premiers à mettre au point ces méthodes couramment appliquées aux principales variables des recensements ou des enquêtes : âge, sexe, lien avec le chef de ménage et état matrimonial. Ces procédures sont néanmoins applicables à tout groupe de variables ou à l'ensemble des variables figurant sur le questionnaire d'un recensement ou d'une enquête. Selon cette méthode, le programme de vérification examine simultanément les réponses fournies à toutes ces rubriques pour une personne ou pour l'ensemble des personnes d'un ménage afin d'identifier les réponses manquantes ou incohérentes.

204. Statistique Canada a mis au point la méthode Fellegi-Holt et l'a utilisée pour les recensements au Canada de 1976 à 1991. Pour le recensement du Canada réalisé en 1996, cette approche a été perfectionnée et intitulée Nouvelle méthodologie d'imputation (NMI). Cette approche a

FIGURE 23. Exemple de données sur la population comportant des indicateurs associés aux valeurs imputées

Personne	Sexe	Age	Nombre d'enfants (NE)	Indicateur associé au sexe	Indicateur associé à l'âge	Indicateur associé au nombre d'enfants (NE)
1	1	40	BLANK	0	0	1
2	2	40	7	0	70	0

FIGURE 24. Exemple d'indicateur correspondant à une jeune femme dont l'enregistrement comporte des données sur la fécondité laissées en blanc, avec ajout d'un indicateur

Personne	Sexe	Age	Nombre d'enfants (NE)	Indicateur associé au sexe	Indicateur associé à l'âge	Indicateur associé au nombre d'enfants (NE)
Données sur la fécondité en blanc						
4	2	13	1	0	0	
Données sur la fécondité en blanc avec indicateur						
4	2	13	BLANC	0	0	1

permis pour la première fois de « résoudre des problèmes à grande échelle [de vérification et d'imputation], en procédant par imputation simultanée de variables numériques et qualitatives » (Bankier, Houle et Luc, n.d.).

205. Si le processus de vérification utilise la méthode traditionnelle d'imputation dynamique (*hot deck*), les informations d'imputation concernant une série de rubriques élémentaires du questionnaire peuvent provenir d'un grand nombre d'individus différents, selon l'information utilisée pour la mise à jour de la matrice d'imputation. Par exemple, si le sexe de la personne A, le lien avec le chef de ménage et l'état matrimonial sont corrects, ces valeurs serviront à la mise à jour des matrices appropriées. Par contre, si l'âge de la personne A est manquant ou non valide, il ne servira évidemment pas à mettre à jour les matrices d'imputation. En fait, les autres rubriques serviront à mettre à jour cette valeur. Si la valeur de la variable « sexe » de la personne suivante est incohérente, et en cas d'imputation du sexe, la personne A fera alors office de « donneur ». Si l'âge est également inconnu, le programme de vérification utilisera l'âge d'une autre personne.

206. La nouvelle méthodologie d'imputation utilise des donneurs pour les rubriques considérées, dans l'espoir que toutes les informations manquantes ou incohérentes pourront provenir d'un donneur unique ou d'un petit nombre de donneurs. L'obtention de la totalité ou de la plus grande partie de l'information auprès d'un donneur unique exige l'enregistrement des fichiers de données dans la mémoire de l'ordinateur. Alors, en cas de données non spécifiées ou non valides d'âge et de sexe, les mêmes variables enregistrées fournissent des valeurs pour ces deux rubriques.

207. La méthode automatisée d'imputation dynamique doit s'attacher aux objectifs suivants :

a) Le ménage dont les données ont été imputées doit ressembler étroitement au ménage dont la vérification a mis en évidence des erreurs.

b) Les données imputées concernant un ménage doivent provenir si possible d'un seul donneur, plutôt que de plusieurs. En outre, le ménage dont les données ont été imputées doit ressembler étroitement à celui du donneur unique;

c) Différentes imputations de qualité identique, fondées sur les donneurs disponibles, doivent avoir une probabilité voisine d'être retenues, afin d'éviter un accroissement disproportionné de la taille de groupes d'effectif réduit, dont le rôle est néanmoins important (Bankier, Houle et Luc, n.d.).

208. Ces objectifs sont atteints en identifiant tout d'abord les ménages dûment vérifiés qui sont aussi proches que possible de ceux qui ont été rejetés par la vérification. Autrement dit, les deux ménages doivent coïncider en ce qui concerne le plus grand nombre possible de variables qualitatives, et présenter seulement des différences limitées en ce qui concerne les variables numériques. Les ménages qui possèdent ces caractéristiques sont appelés « voisins les plus proches ». L'étape suivante consiste à identifier, pour chacun des voisins les plus proches, les plus petits sous-ensembles de variables ne coïncidant pas (tant numériques que qualitatives) qui, en cas d'imputation, permettent au ménage de passer avec succès la vérification. On procède ensuite à la sélection aléatoire de l'une des opérations d'imputation qui permettent de passer avec succès la vérification et de regrouper les ménages rejetés et les ménages acceptés (Bankier, Houle et Luc, n.d.).

III. — VÉRIFICATIONS STRUCTURELLES

209. Les vérifications structurelles contrôlent la couverture des enregistrements réalisés et déterminent comment les divers enregistrements s'ajustent entre eux. Ces vérifications structurelles doivent assurer que : *a)* les enregistrements relatifs à tous les ménages et à tous les logements collectifs à l'intérieur d'un secteur de dénombrement sont présents et figurent dans le bon ordre; *b)* que toutes les unités d'habitation occupées comportent des enregistrements relatifs à des personnes, contrairement aux unités inoccupées; *c)* que les ménages ne comportent aucun enregistrement personnel en double ou manquant; et *d)* que les secteurs de dénombrement ne comportent aucun enregistrement d'habitation en double ou manquant. Les vérifications structurelles s'emploient donc à garantir le caractère globalement exhaustif des questionnaires.

210. Les vérifications structurelles particulières réalisées pour un recensement ou une enquête sont susceptibles de devoir évoluer au cours du temps compte tenu de la rapidité de changement des techniques appliquées à la détermination et à la correction des erreurs de structure. Aussi le présent chapitre examine-t-il la question plus générale de la validité des données et des relations entre elles, à l'intérieur d'un enregistrement et d'un enregistrement à l'autre. Les chapitres IV et V traitent de rubriques particulières concernant la population et l'habitation.

A. — Vérifications géographiques

1. EMPLACEMENT DES LOCAUX D'HABITATION (LOCALITÉ)

211. Une localité, d'après la publication intitulée *Principes et recommandations concernant les recensements de la population et de l'habitat*, première révision (Nations Unies, 1998, p. 68) est définie comme « un groupement de population distincte... dont les membres occupent des locaux à usage d'habitation voisins et qui a un nom ou un statut localement reconnu ». Des informations supplémentaires concernant l'emplacement des locaux d'habitation figurent dans la définition des termes « localité » et « population urbaine et rurale » aux paragraphes 2.49-2.59 des *Principes et recommandations*. Il est essentiel que les personnes responsables de la réalisation des recensements de l'habitation étudient ces indications en raison de l'importance particulière des concepts géographiques servant à décrire l'emplacement des locaux d'habitation dans le cadre des recensements de l'habitat, tant pour l'exécution du recensement que pour l'exploitation des résultats (Nations Unies, 1998, par. 2.312).

212. Lors de la vérification des données d'emplacement, les codes géographiques doivent être parfaitement exacts. Ainsi, l'obtention de codes complets et exacts relatifs aux différents niveaux du découpage géographique en vue du traitement des données figure-t-elle parmi les tâches

les plus délicates de tout le recensement. En cas d'erreur de code géographique, les opérateurs de saisie risquent d'attri-

ENCADRÉ 3. *Principes directeurs concernant les vérifications structurelles*

Les vérifications structurelles doivent assurer l'exécution de tâches suivantes :

- 1) Vérifier que chaque lot de secteurs de dénombrement a reçu les codes géographiques appropriés (provinces, districts, secteurs de dénombrement, etc.) et que la désignation des lots suit la pratique courante;
- 2) Vérifier que toutes les unités d'habitation sont prises en compte et que tous les ménages d'un secteur de dénombrement sont inscrits;
- 3) Regrouper les ménages au sein de leurs secteurs de dénombrement respectifs et regrouper les secteurs de dénombrement au niveau d'agrégation géographique supérieur approprié;
- 4) Faciliter le choix entre les pages concernant les personnes et les pages concernant le ménage, comprises ou non dans le questionnaire ou en dehors, en fonction de l'effectif de la population et de l'organisation du questionnaire;
- 5) Attribuer chaque enregistrement individuel au type d'enregistrement valide approprié;
- 6) Traiter séparément les enregistrements concernant les logements ou locaux collectifs à usage d'habitation et ceux concernant les unités d'habitation;
- 7) Vérifier l'existence d'une correspondance entre les différents types d'enregistrement : par exemple, aucune personne n'habite les unités inoccupées; une personne au moins habite les unités dites occupées. Veiller à ce que le nombre d'enregistrements personnels concernant chaque ménage corresponde au nombre total de personnes du ménage indiqué dans l'enregistrement relatif à l'habitation. Veiller à la présence du nombre correct de questionnaires en cas d'utilisation de documents multiples pour un ménage donné et vérifier qu'ils sont liés entre eux de manière adéquate.
- 8) Éliminer les enregistrements en double, aussi bien à l'intérieur des ménages (personne figurant en double), que d'un ménage à l'autre (ménage inscrit en double ou partiellement inscrit en double), afin d'éviter les erreurs de surcouverture;
- 9) Classer les enregistrements vides à l'intérieur d'un type d'enregistrement particulier.
- 10) Traiter les unités d'habitation manquantes.

buer l'unité ou les unités d'habitation à une autre région du pays. Il est extrêmement difficile de corriger ce type d'erreur

2. RÉSIDENCES URBAINES ET RURALES

213. La distinction traditionnelle entre les zones urbaines et rurales d'un pays repose sur l'hypothèse selon laquelle dans les zones urbaines quelle que soit la définition que l'on en donne, le niveau de vie y est différent et en général plus élevé que dans les zones rurales. Dans bien des pays industrialisés, cette distinction n'est plus très nette, et la principale différence entre les conditions de vie des zones urbaines et des zones rurales tient avant tout au degré de concentration de la population. Des différences entre les modes de vie et les niveaux de vie urbains et ruraux dans les pays en développement sont encore sensibles mais ces pays ont désormais grand besoin eux aussi, en raison de l'urbanisation rapide qu'ils connaissent, de données relatives aux zones urbaines de différentes dimensions (Nations Unies, 1998, par. 2.53).

214. La plupart des pays définissent les zones géographiques dites « urbaines » et « rurales » préalablement au recensement et introduisent les ajustements nécessaires une fois les données de recensement recueillies. En cas d'attribution de codes correspondants à une résidence urbaine ou rurale (par exemple 1 pour la résidence urbaine et 2 pour la résidence rurale), ces codes peuvent être enregistrés lors de la saisie ou établis au cours des opérations de vérification, selon les critères définis par l'équipe de vérification. Lorsque l'équipe de vérification fournit une liste des unités géographiques urbaines et rurales, les personnes chargées de traiter les données peuvent facilement attribuer les codes appropriés aux enregistrements concernant l'habitation.

215. Il convient de s'employer à vérifier la cohérence générale des caractéristiques démographiques avec la nature du secteur de dénombrement. Par exemple, dans certains pays, à l'exception des médecins, des enseignants et des personnes exerçant des professions analogues, les zones rurales comptent peu de personnes ayant reçu une formation supérieure et les zones urbaines ne comptent guère de travailleurs agricoles. L'équipe de vérification devrait s'assurer de la classification correcte de la zone géographique.

B. — Contrôles de couverture

1. DÉNOMBREMENT DE FAIT ET DE DROIT

216. Les bureaux nationaux de recensement/statistique effectuent généralement des recensements de fait (là où les personnes se trouvent le soir du recensement) ou de droit (là où elles se trouvent habituellement). Le contrôle du lien entre les enregistrements concernant l'habitation, en particulier le décompte des personnes qui occupent des logements, et les enregistrements individuels doivent tenir compte du type de recensement. Parfois, les pays recueillent conjointement les informations de fait et de droit. Une rubrique définie pour chaque personne peut indiquer si celle-ci : 1) y réside en permanence; 2) se trouve temporairement en visite, mais habite généralement ailleurs; ou 3) y habite d'ordinaire, mais se trouve temporairement absente.

Les tableaux de résultats établis sur la base des données de fait recueillies utilisent uniquement les données de type 1 et 2 en présence de données des trois types; les tableaux de résultats établis à partir de données de droit utilisent uniquement les données de type 1 et 3 en présence des trois types de données⁴. Le programme de vérification devrait être conçu pour assurer que, en présence des trois types d'enregistrement, les correspondances sont appropriées. Si les enregistrements de fait comportent peu de réponses, cela peut indiquer qu'il s'agit en fait de résidents absents ou qu'un autre problème de dénombrement doit faire l'objet le cas échéant d'un traitement particulier.

2. HIÉRARCHIE DES MÉNAGES ET DES UNITÉS D'HABITATION

217. Le chapitre V étudie les relations entre ménages, unités d'habitation et locaux à usage d'habitation. Chaque organisme de recensement national/statistique utilise ces concepts d'une façon qui lui est propre. Toutefois, avant de procéder aux différents contrôles concernant l'habitation, l'équipe de vérification doit définir des méthodes opératoires assurant le respect de la hiérarchie au cours de la collecte et de la saisie des données.

3. FRAGMENTS DE QUESTIONNAIRES

218. Avant de procéder à une vérification rubrique par rubrique, le programme informatique doit contrôler dans le cadre de la vérification structurelle les enregistrements valides, les enregistrements manquants et les numéros de ligne en double. Il doit en outre déterminer si les enregistrements vérifiés se rapportent à des personnes qui vivent dans des collectivités. Les opérateurs de saisie peuvent faire une erreur d'introduction de données ou incidemment oublier de supprimer des informations fragmentaires (parties d'enregistrement). Les vérifications préliminaires ont notamment pour rôle d'examiner la présence d'enregistrements fragmentaires dans le fichier afin de les détruire. Le cas le plus répandu sera celui d'un enregistrement contenant des codes géographiques mais sans aucune donnée concernant la population ou l'habitation.

C. — Structure des enregistrements sur l'habitation

219. Le nombre de logements dans un immeuble figure parmi les thèmes susceptibles de faire l'objet d'une collecte

⁴ Les bureaux nationaux de recensement/statistique qui mettent en place ces catégories doivent les utiliser de façon particulièrement stricte lors de la collecte et du traitement de données, mais aussi au cours de leur analyse ultérieure. En présence de ces trois catégories, les utilisateurs doivent savoir de quelle population il s'agit, car l'analyse de la totalité du jeu de donnée se traduira par l'inclusion de certaines personnes en double. S'il faut considérer une population « de fait », la mise en tableau doit exclure les personnes de la catégorie 3 momentanément absentes; s'il faut considérer une population de type « de droit », la mise en tableau doit exclure celles de la catégorie 2. Au cours de l'exploitation initiale des résultats, en ce qui concerne les rapports imprimés et les sujets supplémentaires, l'équipe de vérification peut choisir de traiter seulement un sous-ensemble du jeu de donnée global. Pour les mises en tableau ultérieures la documentation de fichiers devrait spécifier explicitement le mode de traitement des différents tableaux possibles.

de données dans le cas de recensements nationaux sur l'habitat. Dans ce cas, l'unité de dénombrement est un immeuble et l'information recueillie se rapporte au nombre de logements classiques de base qui s'y trouvent (voir Nations Unies, 1998, par. 2.418).

220. Le terme « vérification générale » désigne l'opération consistant à assurer que le nombre d'unités d'habitation faisant partie de l'immeuble correspond au nombre total d'unités figurant dans l'enregistrement. S'il n'y a pas concordance, le nombre d'unités d'habitation inscrit comme caractéristique de l'immeuble doit être corrigé pour correspondre au nombre d'enregistrements d'unités d'habitation. Si le codage de l'immeuble en question indique qu'il comporte 5 unités d'habitation, alors que le décompte réel des unités d'habitation individuelles dans cet immeuble est de 4, alors l'équipe de vérification doit décider de l'ajustement à effectuer : a) modifier le premier chiffre, ce qui dans la plupart des cas devrait s'avérer plus indiqué; et b) introduire une valeur sur la base d'informations concernant les enregistrements existants (ce qu'il conviendrait d'éviter).

D. — Correspondance entre les enregistrements relatifs à l'habitation et à la population

221. Lorsque le recensement ou l'enquête comprend des enregistrements sur l'habitation et sur la population, une vérification de structure s'impose pour s'assurer que les deux types d'enregistrement concordent.

1. HABITATIONS INOCCUPÉES ET OCCUPÉES

222. A l'inverse d'une unité d'habitation occupée, une unité d'habitation inoccupée ne comporte aucun enregistrement concernant la population. En présence d'enregistrements relatifs à la population, mais lorsque l'unité d'habitation en question figure sur la liste comme étant inoccupée, il faut alors substituer l'état occupé à l'état inoccupé. Parfois, l'enregistrement est organisé de façon à regrouper l'état d'occupation et les « modalités de jouissance » sous la même rubrique, de telle sorte que cette même information doit également être prise en compte lors du choix. S'il existe une réponse concernant la valeur d'une unité occupée par le propriétaire ou le « loyer payé » pour une unité occupée par des locataires, alors les programmes de vérification utilisent ces renseignements pour déterminer la nature du logement; sinon, l'utilisation d'une matrice d'imputation peut s'avérer nécessaire.

223. Lorsqu'une unité d'habitation théoriquement occupée ne fait l'objet d'aucun enregistrement de type démographique, l'équipe de vérification doit décider s'il faut recenser l'unité comme étant inoccupée ou substituer la présence de personnes provenant d'une autre unité. Si l'unité est inoccupée, la modification par imputation de la variable « état d'occupation » ne pose aucune difficulté. Si l'unité est occupée, par contre l'équipe de vérification doit décider si et comment attribuer les personnes provenant d'une autre unité dans laquelle habite le même nombre de personnes et présentant si possible des caractéristiques similaires. Puisqu'il n'est pas possible de connaître les caractéristiques des personnes omises, cette méthode ne devrait

être utilisée, pour autant qu'elle puisse l'être, que si l'équipe de vérification convient qu'elle n'a pas d'autres choix. Trois cas possibles sont présentés ci-après:

a) *Décision de laisser une unité d'habitation inoccupée*

224. Dans ce cas, l'équipe de vérification décide que les unités d'habitation inoccupées relevées sur le terrain doivent rester considérées comme inoccupées, de telle sorte qu'aucune valeur n'est imputée. Les vérifications dans le cas des unités d'habitation inoccupées sont décrites au chapitre V.

b) *Réexamen des unités d'habitation à plusieurs reprises afin de compléter les questionnaires*

225. Le bureau national de recensement/statistique peut choisir d'appliquer des procédures demandant aux agents de recensement de continuer à réexaminer les données concernant les unités inoccupées jusqu'à ce qu'ils aient la certitude que ces mêmes unités sont effectivement occupées ou inoccupées et jusqu'à ce qu'ils aient recueilli au moins les caractéristiques de base correspondantes. Dans ce cas, l'équipe de vérification doit mettre au point des contrôles permettant de savoir si l'unité est ou non inoccupée ou présente des caractéristiques suffisantes pour être considérées comme « occupée ». Selon le choix de l'équipe de vérification quant aux indications minimales nécessaires, il est procédé soit à la vérification ordinaire décrite au chapitre IV, soit à l'obtention d'enregistrements « donneurs » pour les personnes « manquantes » tel qu'indiqué aux paragraphes 205-207.

c) *Substitution d'une autre unité de logement pour les personnes manquantes*

226. La description des procédures de substitution de ménages entiers ou de personnes manquantes figure ci-après dans le présent chapitre. Ces procédures impliquent l'hypothèse selon laquelle les personnes manquantes ont les mêmes caractéristiques que les personnes substituées. Ce qui, de façon presque certaine, n'est généralement pas le cas, et les procédures proprement dites sont particulièrement délicates. Toutefois, sans utiliser les procédures en question, les effectifs de personnes et les nombres de personnes par caractéristique relevée risquent de diminuer.

2. MÉNAGES ET UNITÉS D'HABITATION FIGURANT EN DOUBLE

227. Les unités d'habitation peuvent figurer en double pour différentes raisons. Parfois, un opérateur de saisie introduit à deux reprises la même unité d'habitation; parfois, des opérateurs de saisie différents réintroduisent accidentellement les mêmes unités de logement ou des secteurs de dénombrement entiers, faute d'avoir mis en place un système d'assurance qualité au niveau du bureau national de recensement/statistique; enfin, un agent de recensement peut très bien enregistrer de façon inadéquate le code géographique d'une unité d'habitation, créant ainsi une information en double, en lui attribuant la même identité géographique que celle d'une autre unité.

228. Aucun enregistrement ne figurera en double si les lots introduits sont contrôlés. Toutefois, il conviendrait de mettre au point un programme de vérification garantissant qu'aucun ménage ne figure en double, suite à la saisie en double par erreur du même ménage ou des mêmes ménages. Les pays devraient s'abstenir de trier leur données, tant que les vérifications structurelles ne sont pas terminées et tant que les problèmes posés par les enregistrements en double ne sont pas éliminés. Avant de procéder au tri, le personnel est en mesure de corriger manuellement les lots; après le tri, il risque par contre de ne pas pouvoir détecter le problème. Lorsque les données sont triées, la présence de ménages enregistrés en double peut être contrôlée au moyen d'une opération de vérification et les données figurant en double par la suite peuvent être éliminées par imputation.

3. MÉNAGES MANQUANTS ET UNITÉS D'HABITATION

229. De manière analogue, à la suite du tri, l'existence de ménages manquants peut être constatée. Par exemple, le programme de vérification anticipe l'existence d'une séquence de ménages à l'intérieur du niveau d'agrégation géographique le moins élevé, par exemple 1, 2, 3 et 4, mais constate uniquement la séquence 1, 2 et 4. Il faut alors décider soit de renuméroter les unités, soit de trouver une méthode acceptable quelconque de substituer une autre unité à l'unité 3 manquante.

4. CORRESPONDANCE ENTRE LE NOMBRE D'OCCUPANTS ET LA SOMME DES OCCUPANTS

230. Le nombre d'occupants inscrit dans l'enregistrement relatif à l'habitation doit être exactement égal au décompte des personnes du ménage. Le programme de vérification additionne donc le nombre de personnes, puis compare cette valeur au nombre d'occupants inscrit dans l'enregistrement relatif à l'habitation. Si ces deux valeurs diffèrent de la valeur du nombre d'occupants, il faut alors ajuster soit le nombre d'occupants de façon à ce qu'il soit identique au décompte des personnes, soit les données individuelles. Le chapitre V décrit de façon détaillée le processus de vérification des données d'habitation en ce qui concerne le nombre d'occupants.

a) *Quand le nombre d'occupants est supérieur au décompte des occupants*

231. Si la valeur attribuée au nombre d'occupants figurant dans l'enregistrement concernant l'habitation est supérieure au décompte des individus, l'équipe de vérification est alors confrontée à une véritable difficulté. Personne ne peut connaître les caractéristiques des personnes manquantes; les équipes de vérification qui choisissent d'imputer les personnes manquantes, caractéristique par caractéristique ou en les remplaçant par des personnes qui vivent dans des ménages similaires, risquent donc de se trouver face à un dilemme. Il convient en effet de ne pas remplacer les personnes manquantes. Toutefois, si le nombre d'occupants est accepté, la solution de rechange consiste à diminuer l'effectif de la population dénombrée. L'équipe de vérification doit

étudier la situation globale, puis décider d'un chemin de traitement approprié.

232. Il existe plusieurs moyens pour repérer et remplacer des enregistrements manquants, mais aucun d'eux n'est parfaitement satisfaisant. Il est possible de sauvegarder des enregistrements de ménages complets comportant des variations des principales caractéristiques. Lorsqu'on trouve un ménage comportant un certain nombre d'individus, mais pas la totalité, il est possible de rechercher dans le jeu de données un ménage dont la plus grande partie sinon la totalité des caractéristiques connues coïncident, ce qui permet ensuite d'ajuster les caractéristiques des personnes manquantes d'après celles des autres personnes du ménage donneur. Cependant, la programmation de cette opération est extrêmement complexe, et les bureaux nationaux de recensement/statistique qui utilisent cette méthode doivent préparer longtemps à l'avance sa mise en œuvre.

233. Une variante de cette procédure consiste à associer des indicateurs à tous les ménages comportant des enregistrements manquants puis à effectuer le reste des vérifications. Au terme du processus de vérification, après avoir corrigé toutes les données individuelles, l'équipe de vérification peut choisir de demander aux informaticiens d'effectuer sur le fichier les ajouts et les modifications possibles au moyen du jeu de données complet dûment vérifié; selon cette approche descendante, l'équipe est en mesure d'identifier des donneurs acceptables.

b) *Contrôle du nombre de personnes par sexe*

234. Parfois l'enregistrement concernant l'habitation porte la mention du nombre d'occupants, par sexe. Le programme de vérification doit alors additionner séparément le nombre de personnes de chaque sexe. Là encore, si la somme diffère des nombres respectivement indiqués, l'une des deux valeurs doit être ajustée dans chaque cas. D'ordinaire, on préfère ajuster les totaux inscrits dans les enregistrements relatifs au logement, plutôt que d'additionner des enregistrements manquants ou de supprimer des enregistrements comportant des informations utiles.

c) *Numérotation séquentielle*

235. Les enregistrements relatifs à la population devraient être numérotés dans un ordre séquentiel. Ces numéros devraient figurer sur le questionnaire comme une variable (telle qu'un numéro de ligne ou un numéro séquentiel). En outre, les numéros séquentiels devraient apparaître dans l'ordre numérique. Des erreurs peuvent néanmoins se produire; parfois les questionnaires ou les feuilles relatives aux personnes sont déclassés parce que les agents de recensement regroupent les informations dans le mauvais ordre, ou sautent des pages, laissant involontairement des pages vides dans le jeu de données. Bien qu'une lacune de numérotation séquentielle n'affecte ni les opérations de vérification ni celles de mise en tableau, nombre de bureaux nationaux de recensement/statistique choisissent de reclasser les individus dans l'ordre adéquat. Le programme de vérification doit donc pouvoir identifier les personnes qui ne sont pas dans le bon ordre et les renuméroter. Puisque la re-

numérotation affecte parfois le lien avec le chef de ménage, les spécifications de vérification doivent être réexaminées. Ainsi, la renumérotation affectera certainement des variables telles que le numéro de ligne de la mère ou du mari.

5. CORRESPONDANCE ENTRE LES OCCUPANTS
ET LE TYPE D'IMMEUBLE/DE MÉNAGE

236. Le type de lien entre les membres d'un ménage doit être compatible avec le type d'unité d'habitation. Parfois, les membres d'un ménage figurent dans une habitation déclarée en tant que logement collectif ou vice versa. En l'occurrence, le type de lien ou le type d'unité d'habitation doit nécessairement tenir compte de la taille du ménage et de différentes variables.

E. — Enregistrements en double

237. Les questionnaires à lecture optique ou autre ne sont guère susceptibles de présenter des numéros de ligne en double. Pour les formulaires dont le contenu doit être introduit au clavier, le bureau national de recensement/statistique peut convenir de vérifier la correspondance entre la liste des ménages et les numéros de ligne du ménage à introduire manuellement. Ce contrôle manuel peut améliorer la qualité des données saisies, en particulier si l'on rapproche : 1) les noms des personnes figurant sur une page comportant la liste de toutes les personnes du ménage; et 2) les données figurant dans les colonnes, rangées ou pages relatives aux personnes. Ainsi, deux personnes qui à première vue semble constituer le dédoublement d'une seule peuvent en fait être des jumeaux si l'on se réfère à leurs noms.

238. Les formulaires introduits au clavier ne doivent pas comporter de numéros de ligne en double si les écrans de données et les instructions de saut de champ sont correctement configurées. La plupart des progiciels actuels créent automatiquement des numéros séquentiels dans le cadre du processus de saisie. Or, une erreur peut être introduite en cas de saisie d'enregistrement en double pour une personne ou encore de création d'un enregistrement double dû à un numéro de ligne erroné. Au fur et à mesure du traitement de chaque enregistrement, le programme de vérification le compare aux précédentes données de population relatives à l'unité d'habitation. La vérification effectuée doit contrôler la saisie correcte de chaque numéro de ligne. Les numéros de ligne en double constituent des erreurs et doivent être modifiés.

239. Au lieu d'utiliser un progiciel du commerce les pays peuvent choisir de mettre au point leurs propres systèmes d'introduction des données. Dans ce cas, l'équipe de vérification doit convenir du niveau acceptable d'erreurs. Parmi les approches envisageables à cet effet, l'une d'elles pourrait consister à suivre les principes ci-dessous :

a) Si deux enregistrements différents correspondent au même numéro de ligne et si le nombre de caractéristiques différentes est inférieur ou égal à deux, le programme de vérification doit éliminer l'un des enregistrements puisqu'il s'agit vraisemblablement d'un double.

b) si le nombre de caractéristiques différentes est supérieur ou égal à 3, le numéro de ligne doit être modifié.

F. — Populations spéciales

1. PERSONNES VIVANT EN COLLECTIVITÉ

240. La vérification structurelle doit traiter les personnes vivant dans des collectivités telles que institutions, casernes ou maisons de soins différemment de celles qui vivent dans des unités d'habitation normales. Puisque dans les collectivités il n'existe généralement pas de chef de ménage, les pays doivent décider de la meilleure façon de distinguer les différents types d'unités. Une méthode consiste à définir un type d'enregistrement propre aux collectivités, une autre à attribuer au lien avec le chef de ménage un code particulier, correspondant aux logements en groupe ou collectifs.

a) *Cas dans lesquels les enregistrements collectifs constituent un type d'enregistrement différent*

241. Si le bureau national de recensement/statistique décide d'utiliser un type d'enregistrement distinct, l'équipe de vérification n'aura aucune difficulté à déterminer quels enregistrements sont de type collectif. Les tableaux correspondants sont aisément réalisables par simple référence directe aux enregistrements de ce type. Les variables propres aux enregistrements collectifs, (par exemple le type de collectivité) peuvent faire l'objet d'opérations distinctes de vérification et d'imputation. Quant aux variables exclues des enregistrements collectifs, il est facile de les vérifier en s'assurant que les données en question sont effectivement laissées en blanc. Il en résulte toutefois un fichier plus encombrant, puisque ces enregistrements sont généralement plus courts que les enregistrements normaux, mais ils occupent autant de place que dans un fichier de type rectangulaire. De plus, au cours de la vérification et de l'imputation, certains programmes peuvent devoir vérifier certaines rubriques tant des enregistrements de population que des enregistrements collectifs.

b) *Cas dans lesquels une variable distingue les enregistrements collectifs des autres enregistrements*

242. Lorsqu'on utilise une variable distincte, de préférence à un type d'enregistrement distinct, l'équipe de vérification peut avoir plus de difficultés pour identifier les enregistrements de type collectif. Dans ce genre de situation, les tableaux relatifs aux collectivités sont également faciles à établir en se référant à la variable proprement dite, qui indique quels enregistrements se rapportent à des personnes faisant partie de collectivités. Les variables propres aux collectivités (comme le type de collectivité) peuvent également faire l'objet de vérifications et d'imputations distinctes. Il est facile de vérifier les variables exclues des enregistrements collectifs, pour s'assurer qu'elles sont effectivement laissées en blanc en se référant au code des enregistrements collectifs. Il en résulte un fichier plus compact, puisque les enregistrements supplémentaires concernant les personnes des collectivités ne sont pas nécessaires, mais figurent simplement en tant qu'enregistrements de

données démographiques, dans lesquels la variable ménage/collectivité porte un code différent. Lors des opérations de vérification et d'imputation, le programme devra contrôler uniquement certaines rubriques des enregistrements de population, et non des deux types d'enregistrement (données de population et données sur les collectivités).

c) *Cas dans lesquels le code « type de collectivité » est omis*

243. Le code indiquant le type de collectivité est parfois manquant ou non valide, ou encore il peut y avoir un défaut de concordance entre le code de collectivité et les codes de lien avec le chef de ménage. En l'absence du code de type de collectivité, mais si les codes de lien avec le chef de ménage indiquent qu'il s'agit bien d'une collectivité, la solution proposée consiste à modifier en conséquence le codage du type de collectivité. Si ce code est présent, mais si le code du lien avec le chef de ménage est absent, ce dernier peut être établi en fonction du type de collectivité.

d) *Cas dans lesquels le code « type de collectivité » est présent, mais où il existe un lien entre toutes les personnes*

244. Si l'enregistrement comporte un code correspondant à une collectivité, mais si les codes « liens avec le chef de ménage » indiquent l'existence d'un lien entre toutes les personnes de l'unité d'habitation, il convient alors de modifier le type de collectivité pour indiquer qu'il s'agit d'une unité d'habitation. En revanche, si l'unité est codée en tant que ménage, mais qu'il n'existe aucun lien entre deux personnes quelconques de cette unité, il est parfois nécessaire de lui attribuer le code d'habitat de groupe ou collectif. Par ailleurs, un ménage peut compter 5 ou 6 personnes entre lesquelles il n'existe pas de lien, sans toutefois constituer un habitat collectif. Tel qu'indiqué plus haut, les membres de l'équipe de vérification devront parfois se consulter pour résoudre des cas particuliers inhabituels.

e) *Distinction entre différents types de collectivité*

245. La plupart des pays distinguent différents types de collectivité. Ils ventilent souvent l'information en différents types particuliers de logement collectif. Ces informations peuvent soit être codées séparément en tant que type de logement collectif, soit figurer en tant que choix de valeurs des codes de lien avec le chef de ménage.

2. POPULATIONS SANS HABITAT

a) *Migrations saisonnières*

246. Dans certains pays où l'on observe des migrations saisonnières, l'intervieweur devra savoir si l'unité est inoccupée ou occupée en fonction de la référence temporelle. Dès lors, même si les données concernant le ménage sont complètes, ce dernier peut également être compté (dénombré) dans un autre lieu. Evidemment le contraire peut également se produire : un ménage possédant deux logements, situés en des lieux distincts, risque également d'être omis si l'on n'y prend garde.

247. Parfois, des ménages entiers vivent très régulièrement une partie de l'année à un endroit et le reste du temps ailleurs. Le bureau national de recensement/statistique ainsi que l'équipe de vérification doivent décider du mode de traitement de divers types de situations. Par exemple, certaines personnes passent régulièrement une partie de l'année dans une autre maison : les gens qui vivent dans la partie la plus fraîche d'un pays pendant la saison chaude et dans la partie au climat plus clément pendant la saison froide. Un autre cas est celui des itinérants qui voyagent une partie de l'année, mais sont sédentaires le reste du temps; ce qui peut coïncider avec l'époque à laquelle le pays décide de procéder au recensement.

b) *Personnes sans abri*

248. Par définition, l'enregistrement d'une personne sans abri ne contiendra aucune information concernant l'habitation. Toutefois, la création d'un enregistrement « fictif » (nouvel enregistrement qui comprend initialement des valeurs laissées en blanc pour certaines variables) facilitera les vérifications structurelles et permettra d'obtenir un enregistrement cohérent avec la structure des autres unités d'habitation. L'équipe de vérification devra décider éventuellement de créer cet enregistrement fictif d'habitation pour faciliter le traitement et la mise en tableau des données.

c) *Réfugiés*

249. De manière analogue, des réfugiés peuvent occuper des logements temporaires; il faut alors enregistrer une indication de cette situation en affectant une valeur à une variable particulière. L'équipe de vérification devra définir et mettre en place les procédures adéquates.

G. — Détermination du chef de ménage et du conjoint

1. DÉTERMINATION DE LA VARIABLE CHEF DE MÉNAGE

250. Pour identifier les membres d'un ménage, il est d'usage d'identifier d'abord le chef de ménage ou le membre de référence, puis les autres membres, en fonction de leur lien avec lui. Le chef de ménage est la personne reconnue comme telle par les autres membres. Les pays peuvent utiliser le terme qu'ils jugent le plus approprié pour l'identifier (chef de ménage, membre de référence du ménage, etc.), cette personne servant uniquement à déterminer les liens entre les membres du ménage. Il est souhaitable que chaque pays présente dans les rapports qu'il publie les notions et les définitions utilisées (Nations Unies, 1998, par. 2.67).

a) *Ordre des liens avec le chef de ménage*

251. L'ordre des liens avec le chef de ménage au sein de l'unité d'habitation a une incidence sur les vérifications; en effet, celles-ci supposent souvent dans nombre de cas que le chef de ménage est la première personne et vérifient en premier lieu les données correspondantes. Par exemple, des variables telles que la langue, le groupe ethnique et la religion, sont contrôlées en premier lors de la vérification concernant

le chef de ménage. Si pour l'une ou l'autre de ces variables les informations le concernant sont valides, celles-ci sont imputées à toute autre personne du ménage, en cas d'omission, d'erreur de codage ou d'erreur d'introduction au clavier (voir chapitre IV). Il faut donc vérifier en priorité les données du chef de ménage, ses caractéristiques servant ensuite à attribuer ou à imputer des valeurs aux autres membres du ménage.

b) *Cas dans lesquels le chef de ménage n'est pas la première personne*

252. Les décisions prises sur le terrain par les agents de recensement, en fonction des différentes situations rencontrées quant à l'identification du chef de ménage, affectent le processus de vérification. Pour mieux comprendre le problème, examinons tout d'abord le cas du ménage présenté à la figure 25.

253. Ce ménage correspond aux cas types rencontrés sur le terrain : un chef de ménage et un conjoint, leurs enfants, ainsi que le père du chef de ménage. Si l'agent de recensement recueille des informations de cette façon, une vérification fondée sur l'inscription du chef de ménage en tête de liste du ménage ne posera pas de problème.

254. Toutefois, si le recensement se déroule de telle sorte que le grand-père est désigné comme étant le chef de ménage, les liens au sein du ménage sont modifiés, selon le schéma de la figure 26. Tel serait le cas si un agent de recensement se présentait dans une maison, y trouvait une famille nucléaire constituée d'un mari, d'une femme et de deux enfants, et si, pendant l'interview, le père du chef de ménage entrait dans la pièce et déclarait être le chef de ménage. Moyennant l'accord du chef de ménage présumé, la personne n° 5 deviendrait alors le chef de ménage et la personne 1 serait le fils, la personne 2 la belle-fille, etc.

255. Comme le montre évidemment le cas de ces deux ménages, la désignation de tel ou tel chef de ménage conduit

à des chemins de vérification différents. La détermination du chef de ménage effectif pour le reste des opérations de vérification et de mise en tableau donne lieu à trois possibilités distinctes : a) un pointeur (indicateur) peut signaler la personne considérée comme le chef de ménage, ce pointeur pouvant servir tout au long des opérations de vérification et de mise en tableau; b) si le chef de ménage ne figure pas en première position sur la liste, il est possible de le reclasser en première position et de modifier en conséquence la position des différentes personnes qui le précédaient; ou c) les codes attribués à la variable lien avec le chef de ménage peuvent être changés de façon à ce que la personne placée en premier soit le chef de ménage, indépendamment des autres liens.

i) *Attribution d'un pointeur à l'enregistrement du chef de ménage*

256. Dans le cadre des procédures de vérification concernant le chef de ménage, le numéro de ligne du chef de ménage au sein de l'unité d'habitation est déterminé au moyen d'un pointeur. Si le chef de ménage conserve le numéro de classement relevé, un pointeur peut être associé à ce numéro, pour pouvoir identifier facilement le chef de ménage, chaque fois qu'une vérification ou une mise en tableau particulière l'exigera. Une variable « pointeur du chef de ménage » peut être mise à une valeur égale au numéro de ligne du chef de ménage et utilisée au cours de la vérification pour procéder à l'imputation des caractéristiques manquantes ou non valides concernant les autres personnes de l'unité d'habitation. Si le chef de ménage est la personne portant le n° 1, la valeur de la variable « pointeur du chef de ménage » est égale à 1.

ii) *Désignation du chef de ménage en tant que personne n°1*

257. L'équipe de vérification peut décider de classer le chef de ménage en tant que personne n°1. Toutefois, la programmation de cette opération s'avère légèrement plus

FIGURE 25. Exemple de ménage dont le chef figure en premier sur la liste

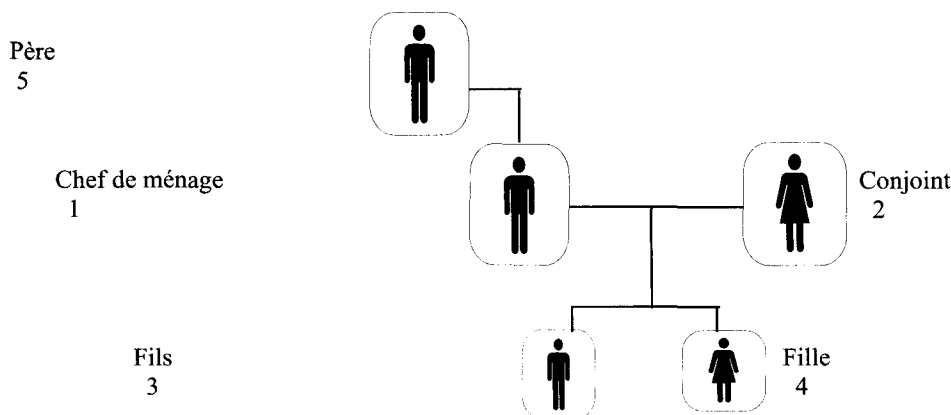
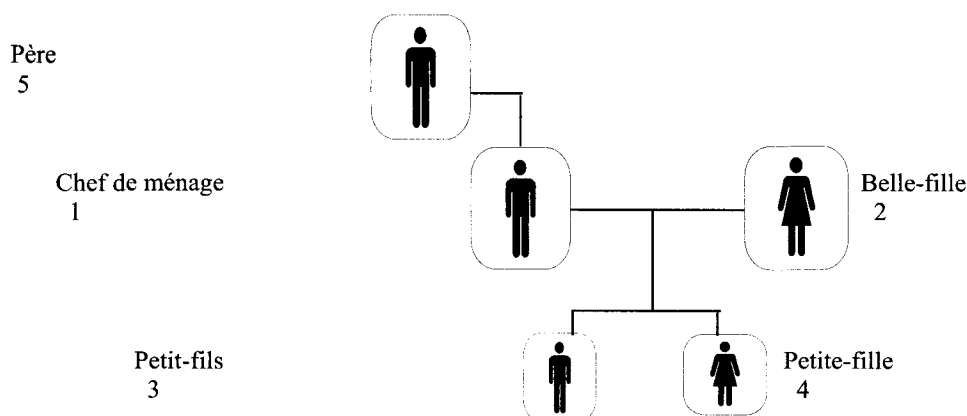


FIGURE 26. Exemple de ménage dont le chef est classé en cinquième position



complexe, par comparaison à l'opération requise dans le cas i). Les informaticiens doivent mettre au point un programme qui amène le chef de ménage en première position sur la liste des membres, où il précède désormais la personne qui s'y trouvait au préalable, puis la personne qui était en deuxième position, jusqu'à la personne qui précédait immédiatement le chef de ménage. Donc, le classement des différents membres passe de 1, 2, 3, 4, 5 à 5, 1, 2, 3, 4. Une fois la modification faite, le chef de ménage occupe la position numéro 1, ce qui facilite les vérifications suivantes, puisque le chef de ménage conservera ce numéro. Cette opération entraîne toutefois une certaine « détérioration » de l'intégrité du jeu de données. En effet, l'ordre de classement des personnes ayant été modifié, les analystes risquent d'avoir des difficultés pour établir l'ordre réel dans lesquels les réponses ont été recueillies sur le terrain et l'incidence potentielle de cet ordre sur l'interprétation des résultats.

iii) Réattribution de codes de lien avec le chef de ménage de façon à placer celui-ci en première position

258. Si l'équipe de vérification décide que la première personne de la liste doit être le chef de ménage, il faut alors intégrer les tâches *a* et *b* aux procédures de contrôle:

a) Le code affecté au chef de ménage est attribué à la première personne;

b) Un programme réattribue les valeurs de codage aux autres personnes du ménage afin de modifier le classement. Par exemple, à la figure 26 le parent commence en tant que chef de ménage. Si la qualité de chef de ménage est attribuée à la personne n° 1, il faudra attribuer la qualité de conjoint à la personne n° 2, puis celle de « enfant » aux personnes n° 3 et n° 4 et réattribuer la qualité de « parent » à la personne n° 5 (tel qu'indiqué à la figure 25). Le sous-programme contiendra alors une matrice regroupant les valeurs initiales et les valeurs modifiées.

259. Cette procédure affecte encore davantage l'intégrité du jeu de données. Le classement des personnes n'est

pas modifié comme il l'était dans l'exemple précédent, et les analystes n'auront aucun mal à déterminer l'ordre des personnes dans lequel les informations ont été effectivement recueillies sur le terrain. Les liens avec le chef de ménage seront alors tous modifiés et les analystes ignoreront qui était choisi initialement comme chef de ménage. En revanche, l'exploitation des données sera formellement facilitée avec le chef de ménage en première position. Contrairement à l'exemple précédent, cette procédure n'exige pas des programmeurs qu'ils permutent les enregistrements.

c) Plusieurs chefs de ménage

260. En présence de plusieurs chefs de ménage, il incombe à l'équipe de vérification de déterminer qui doit être considéré comme tel. La vérification doit s'appuyer sur les caractéristiques définies par les spécialistes du sujet et d'autre part sur les schémas de vérification adoptés. Le programme de vérification doit ensuite réattribuer la valeur de la variable lien avec le chef de ménage de la personne (ou des personnes) précédemment identifiée (s) en tant que chef de ménage.

d) Absence de chef de ménage

261. De manière analogue en l'absence de chef de ménage, il incombe à l'équipe de vérification de déterminer qui nommer en tant que tel. Dans ce cas, les relations entre les autres personnes du ménage devront vraisemblablement être modifiées au cours de la vérification.

2. VÉRIFICATION CONCERNANT LE CONJOINT

a) Cas d'un conjoint unique dans des sociétés monogames

262. Si l'enregistrement comporte un conjoint et un seul, la variable « pointeur de conjoint » enregistre le numéro de ligne du conjoint en vue des vérifications ultérieures. Ces vérifications pourraient consister notamment à s'as-

surer que le chef de ménage ou le conjoint sont de sexes opposés, qu'ils présentent une différence d'âge appropriée ou d'autres caractéristiques pertinentes.

b) *Cas de plusieurs conjoints dans des sociétés monogames*

263. Dans une société monogame, lorsque le jeu de données indique l'existence de plusieurs conjoints, une vérification est nécessaire pour déterminer qui est le conjoint et réattribuer les valeurs « lien avec le chef de ménage » des autres personnes identifiées en tant que conjoints. Là encore, les spécialistes du sujet doivent déterminer les caractéristiques et le schéma de traitement des vérifications.

c) *Conjoints dans des sociétés polygames*

264. Dans une société polygame, si le jeu de données fait apparaître plusieurs conjoints, l'équipe de vérification peut décider de laisser l'information inchangée ou d'effectuer différents contrôles de cohérence. Il conviendrait au minimum de vérifier si le chef de ménage et chacun des conjoints sont de sexes opposés. Si cette opération révèle l'existence de conjoints du même sexe, il convient alors d'effectuer la vérification précédente applicable au conjoint du même sexe.

H. — Age et date de naissance

1. INDICATION DE LA DATE DE NAISSANCE,
MAIS OMISSION DE L'ÂGE

265. Lorsque la date de naissance est indiquée, mais non l'âge, il est possible d'obtenir cette dernière information en soustrayant la date de naissance de la date de recensement ou de l'enquête. Certains bureaux nationaux de recensement/statistique conviennent de définir l'âge en fonction uniquement de l'année de recensement et de l'année de la naissance, ce qui donne une valeur pouvant s'écarter de la valeur réelle. En cas d'utilisation de l'année et du mois, l'âge sera plus précis, la précision optimale étant obtenue grâce à l'utilisation du jour, du mois et de l'année.

2. DÉFAUT DE CONCORDANCE DE L'ÂGE
ET DE LA DATE DE NAISSANCE

266. Lorsque l'âge et la date de naissance sont relevés dans le cadre du recensement ou de l'enquête, on obtient un âge calculé en soustrayant la date de naissance de la date de référence. Si cette valeur diffère de plus d'un an de l'âge déclaré, l'équipe de vérification peut décider d'introduire une correction. Normalement, la date de naissance prévaut sur l'âge déclaré et l'âge calculé est alors substitué à l'âge déclaré.

I. — Décompte des données non valides

267. Certaines équipes de vérification décident parfois de mettre en œuvre des procédures de comptage du nombre de données non valides et incohérentes concernant les principales variables (ou la totalité des variables), telles que l'âge et le sexe, avant de passer aux opérations de vérification proprement dites. L'équipe dispose parfois de plusieurs matrices d'imputation dynamique lorsqu'elle a effectué un travail de préparation préalable, ou lorsqu'elle réalise des enquêtes périodiques portant sur les mêmes rubriques. Lorsque le pourcentage de données non valides ou incohérentes est très faible, elle peut décider d'utiliser seulement un petit nombre de variables pour l'imputation. Par contre, si le pourcentage d'erreurs est plus élevé, l'utilisation d'un plus grand nombre de variables peut s'avérer nécessaire en fonction du plus grand nombre d'imputations requis.

268. Les matrices d'imputation de taille plus réduite sont généralement préférables puisqu'elles sont plus simples à contrôler au fur et à mesure de la mise au point des vérifications et des imputations et sont également d'un usage plus facile au cours de la vérification proprement dite. Si les valeurs sont employées plusieurs fois, une matrice d'imputation plus grande et plus adaptée s'avérera néanmoins indispensable.

IV. — VÉRIFICATIONS CONCERNANT LES DONNÉES SUR LA POPULATION

269. Le chapitre IV est consacré aux vérifications des données sur la population, notamment des données liées aux caractéristiques démographiques, migratoires, sociales et économiques. Les spécifications correspondantes prennent en considération la validité de chaque donnée et la cohérence d'une part, entre les données sur la population et d'autre part entre les données sur la population et sur l'habitation. Les connaissances disponibles quant aux liens entre les éléments d'information permettent de concevoir des vérifications de cohérence propres à garantir une meilleure qualité des données à mettre en tableaux. Par exemple, il ne devrait pas y avoir dans les enregistrements de données sur la population de femmes âgées de 15 ans et mère de 10 enfants, ou d'enfants de 7 ans inscrits à l'université.

270. Lors de l'attribution de valeurs aux données manquantes ou non valides concernant la population, l'équipe de vérification doit décider soit de leur attribuer la valeur « non indiqué », soit de procéder par imputation statique (*cold deck*) des valeurs inconnues ou autres ou encore par imputation dynamique (*hot deck*), d'après les caractéristiques d'autres personnes ou d'autres unités d'habitation. En maintes circonstances, la solution de l'imputation dynamique est choisie de préférence puisqu'elle supprime les vérifications au stade de l'exploitation, lorsque les décisions à prendre concernant les valeurs inconnues ne peuvent s'appuyer que sur le contenu des tableaux proprement dits. Les matrices d'imputation fournissent des données permettant de compléter les omissions, de remplacer les données non valides ou de résoudre les incohérences lorsqu'on ne dispose d'aucune rubrique connexe comportant des réponses valides. Dans certains cas, les caractéristiques démographiques présentent une diversité notable au niveau national, mais beaucoup plus réduite au niveau de la plupart des localités. Dans d'autres pays, les variations peuvent être considérables d'une localité à l'autre, en particulier en matière de résidence urbaine et rurale. Il faut prendre en compte cette variabilité lors de la mise au point des matrices d'imputation et surtout lors de la détermination des valeurs initiales d'imputation dynamique. L'équipe de vérification doit spécifier les conditions dans lesquelles une valeur déterminée doit être affectée à une information laissée en blanc. Cette valeur devrait provenir d'une unité d'habitation précédente, dotée de caractéristiques voisines.

271. Tous les enregistrements concernant la population devraient comporter un numéro de série pour faciliter le traitement informatique. Les vérifications structurelles décrites au chapitre III vérifient la correspondance entre le numéro séquentiel et l'ordre des numéros de série.

272. L'équipe de vérification devrait contrôler uniquement les rubriques appropriées des enregistrements concernant la population. Or, le choix de ces rubriques peut différer

en fonction de différentes caractéristiques (notamment milieu urbain/rural, climat, etc.). Bien qu'il y ait intérêt à effectuer des vérifications sélectives, en fonction de ces caractéristiques, peu de pays disposent du temps ou des compétences nécessaires pour définir et utiliser plusieurs tableaux de valeurs permettant de remplacer les données manquantes ou incohérentes. Les pays effectivement en mesure de suivre cette procédure supplémentaire sont encore moins nombreux.

273. L'information recueillie sur le questionnaire dépend par ailleurs de certaines caractéristiques de la population considérée. Par exemple, les questions concernant la fécondité sont posées uniquement aux femmes et celles concernant l'activité économique, aux adultes.

274. Parfois, l'équipe de vérification doit autoriser pour certaines rubriques la mention « non indiqué ». De plus, elle ne dispose pas nécessairement de données suffisamment probantes pour imputer des réponses concernant certaines caractéristiques. La décision de laisser la mention « non indiqué » doit être soigneusement pesée, compte tenu de la nécessité d'obtenir des données tabulaires appropriées à des fins de planification et pour les besoins de l'action des pouvoirs publics. Dans la mesure où les données « non indiquées » présentent une distribution analogue à celle des données indiquées, la répartition des mentions « non indiqué », lorsque les planificateurs ont besoin de certaines informations, ne devrait pas poser de problèmes. En revanche, si la distribution des réponses « non indiquées » comportent un certain biais, alors l'application des procédures d'imputation après compilation risque d'être problématique, en particulier dans des zones peu étendues ou pour des caractéristiques bien précises. Par exemple, si des adolescentes interrogées refusent de répondre aux questions concernant leur fécondité, et si aucune donnée n'est ainsi recueillie, le processus de vérification ne pourra aucunement faciliter l'obtention de cette information.

275. Les vérifications concernant les données sur la population sont d'ordinaire plus complexes que celles concernant l'habitation, en raison du caractère généralement beaucoup plus complexe des distributions croisées. La plupart des pays établissent chaque caractéristique d'habitation uniquement selon différents niveaux géographiques, alors qu'ils disposent parfois d'un grand nombre de strates de distribution croisée pour les données sur la population. Tel qu'indiqué plus haut, les pays qui décident de ne pas utiliser la méthode d'imputation dynamique devraient définir un identificateur pour les mentions « inconnu », à utiliser en présence de réponses non valides incohérentes.

276. Pour les pays qui appliquent une méthode d'imputation dynamique, les équipes de vérification devraient établir des matrices d'imputation simples dont les dimensions

permettent de distinguer différentes caractéristiques démographiques. Dans la plupart des cas, le groupe d'âge et le sexe sont les variables primaires les plus indiquées pour utiliser la méthode d'imputation dynamique et elles devraient faire l'objet de la vérification initiale. Les bureaux nationaux de statistique/recensement qui effectuent des vérifications à plusieurs variables, devraient prendre en compte simultanément l'âge, le sexe et différentes variables telles que le lien avec le chef de ménage et l'état matrimonial. Parmi les autres rubriques susceptibles de faciliter l'application d'une méthode d'imputation dynamique figurent le niveau de formation et la situation dans l'emploi.

277. Les équipes de vérification doivent veiller tout particulièrement à ne pas introduire de distorsions dans les données au cours du processus d'imputation. Ainsi, il n'y a pas lieu de supposer que les données imputées et non imputées ont nécessairement des distributions identiques. Souvent, les données inconnues comportent une distorsion intrinsèque. Par exemple, les personnes relativement âgées sont moins disposées à déclarer leur âge que les personnes plus jeunes.

A. — Caractéristiques démographiques

278. Dans tout recensement, les données individuelles recueillies sur le lien avec le chef de ménage, le sexe, l'âge et l'état matrimonial sont fondamentales et devraient sans doute être vérifiées conjointement. Les répartitions par âge et par sexe des populations ou des sous-populations sont indispensables à pratiquement toutes les activités de planification fondées des recensements. Ces rubriques sont en outre nécessaires à l'obtention de tableaux de données significatives, puisque pratiquement toutes les autres analyses reposent sur des distributions d'autres variables croisées par âge et par sexe.

279. La méthode dite à variables multiples (Fellegi-Holt) de vérification des données concernant la population et l'habitation a été présentée au chapitre II de ce manuel. Puisque les variables démographiques font partie intégrante de l'organisation de tous les recensements, cette méthode devrait être utilisée, dans la mesure où les délais et les compétences disponibles le permettent. De fait, la qualité de l'ensemble du jeu de données bénéficiera très certainement d'une vérification prioritaire axée sur les variables d'âge et de sexe et sur différentes variables particulières, afin de déterminer les erreurs ou les incohérences. Les rubriques les plus fréquemment entachées d'erreurs sont vérifiées en premier, puis celles pour lesquelles les erreurs ou les incohérences sont moins fréquentes.

1. LIEN AVEC LE CHEF DE MÉNAGE (P2A)⁵

280. La rubrique « lien avec le chef de ménage » facilite la détermination de la structure du ménage et de la famille.

⁵ Les chiffres entre parenthèses figurant aux chapitres IV et V se rapportent aux rubriques mentionnées dans la publication *Principes et recommandations concernant les recensements de la population et de l'habitat, première révision*, études statistiques, série M, n° 67/Rev.2 (Publication des Nations Unies, numéro de vente E.98.XVII.8). Donc, « P2A », corres-

Elle figure pratiquement au début de la plupart des questionnaires de recensement et d'enquête et permet d'assurer le décompte de chaque individu de l'unité d'habitation. Recenseur et recensé utilisent les informations concernant les liens des différents membres du ménage, avec le chef de ménage pour s'assurer que personne n'est oublié. La question de la détermination d'un chef de ménage unique et d'un conjoint unique (dans les sociétés non polygames) est traitée aux paragraphes 250-264.

a) Vérifications du lien avec le chef de ménage

281. Vu l'importance croissante des statistiques de lien avec le chef de ménage, il convient de définir soigneusement les vérifications autorisant ultérieurement différents types d'exploitation des données en procédant à des regroupements familiaux et sous-familiaux. Le fait de définir d'emblée un codage approprié de la variable « lien avec le chef de ménage » facilitera évidemment cette tâche (voir paragraphes 607-610 pour les enregistrements de type familial).

282. Lorsqu'il est impossible d'attribuer une valeur à la variable lien avec le chef de ménage et faute d'utiliser la méthode d'imputation dynamique, il faut alors attribuer la valeur « inconnu » aux réponses non valides ou incohérentes. Si par contre, la méthode d'imputation dynamique est utilisée, il est possible d'attribuer une valeur à la variable lien avec le chef de ménage d'après une matrice d'imputation définie en fonction de l'âge, du sexe ou d'autres caractéristiques appropriées. Les matrices d'imputation ne doivent pas imputer des liens avec le chef de ménage qui seraient incompatibles avec les liens déjà définis. Ainsi, il n'y a pas lieu d'imputer des deuxièmes et troisièmes conjoints, même dans des ménages polygames, sauf décision spécifiquement prise à cet effet par le groupe de vérification.

b) Cas dans lesquels le chef de ménage doit figurer en premier

283. Lorsque le chef de ménage ne figure pas en tant que première personne de la liste, les vérifications structurales décrites au chapitre III signalent la possibilité d'utiliser un pointeur pour repérer dans la liste la position du chef de ménage. Si l'équipe de vérification souhaite que le chef de ménage soit la première personne de la liste, il est possible de le placer en première position, soit en modifiant le classement des personnes, soit en conservant la position du chef de ménage, mais en modifiant les liens avec celui-ci, tel qu'indiqué dans le chapitre consacré aux vérifications structurales. La première méthode exige des compétences éprouvées en matière de programmation, tandis que la se-

pond à la donnée de population 2a de la « liste des sujets des recensements de la population », partie II, chap. V de la « liste des sujets de recensement de la population », partie II, chap. V de cette publication, p. 59 et 60. Au regard des différentes rubriques concernant l'habitation, sont mentionnés les codes proposés tirés des sujets fondamentaux et supplémentaires des recensements indiqués dans la partie II au chapitre VI des *Principes et recommandations*, par. 2.293 et 2.416.

conde risque d'altérer la qualité du jeu de données si cette opération n'est pas effectuée avec le plus grand soin.

c) *Cas dans lesquels les liens avec le chef de ménage sont codés « à l'envers »*

284. Les agents de recensement notent parfois « à l'envers » la variable « lien avec le chef de ménage » : au lieu d'enregistrer le lien de chaque personne avec le chef de ménage, ils enregistrent le lien du chef de ménage avec chacune des personnes qui le constituent. Ainsi, le lien de la troisième personne avec le chef de ménage prend la valeur « parent » et non « enfant ». On peut obtenir en définitive un ménage comportant quatre parents et non quatre enfants. Lorsque l'équipe de vérification se trouve confrontée à un problème systématique de ce type, il lui faut trouver une solution qui n'altère pas trop la qualité des données concernant le ménage.

d) *Présence de plusieurs conjoints (polygamie)*

285. Effectuées conformément aux indications du chapitre III, les vérifications structurelles auront déjà vérifié que les ménages monogames comportent un chef de ménage et un seul, ainsi qu'un conjoint et un seul. Pour les ménages polygames, l'équipe de vérification doit décider si la prise en compte de ce type de lien avec le chef de ménage est autorisée ou non. Parfois, les ménages comportant apparemment des liens de polygamie avec le chef de ménage, correspondent en réalité à des erreurs. Par exemple, un chef de ménage et un conjoint peuvent être identifiés à l'intérieur d'un ménage, mais les personnes formant un autre couple se sont déclarées mutuellement en tant que conjoints, portant ainsi à trois le nombre total de conjoints. Le programme de vérification doit vérifier que le deuxième couple n'est pas formé en réalité du père et de la mère ou du fils et de la belle-fille, ou encore de la sœur et du beau-frère, ou d'une autre combinaison. Parfois, les relations de ce type peuvent être déterminées avec quelque certitude et parfois non. Sous réserve du codage des « liens avec le chef de ménage » présentés ci-dessus de façon détaillée, l'équipe de vérification peut escompter des imputations appropriées. Lorsque les conjoints supplémentaires sont effectivement des conjoints (dans des ménages polygames), alors la vérification doit contrôler le sexe et le cas échéant, l'âge des personnes en question.

e) *Présence de plusieurs parents*

286. Les ménages ne doivent pas compter plus de deux « parents » déclarés et ces derniers ne doivent pas être du même sexe. Lorsque les parents déclarés sont plus de deux, la valeur « autre personne apparentée » devrait être sans doute attribuée aux parents supplémentaires. Les recensements ou les enquêtes ont parfois défini un code correspondant aux parents ou aux parents par alliance (beaux-parents), qui permettrait de dénombrer jusqu'à quatre « parents » et non deux, à raison de deux parents ou plus par sexe.

f) *Cas dans lesquels les recensements recueillent les données de lien avec le chef de ménage propres à un sexe*

287. Certains recensements ou certaines enquêtes recueillent des informations particulières à un sexe en matière de lien avec le chef de ménage : ainsi, la distinction de la qualité de « mari » et de « femme » au lieu de « conjoint » ; de « fils » et de « fille » au lieu d'« enfant ». Si ces réponses ne font pas l'objet d'une vérification, les tableaux de résultats risquent de contenir des données comportant des « filles » de « sexe masculin » ou des maris de « sexe féminin ». L'équipe de vérification doit donc fixer l'ordre de priorité des contrôles, à savoir si le lien avec le chef de ménage ou le sexe doit prévaloir. Dans certains cas, par exemple la distinction des qualités de mari et de femme, la vérification s'avère plus importante que dans d'autres, par exemple dans le cas d'un jeune enfant.

g) *Défaut de concordance entre le lien avec le chef de ménage et l'état matrimonial*

288. Le lien avec le chef de ménage et l'état matrimonial doivent concorder lorsqu'ils font double emploi : les personnes qui déclarent le lien de « conjoint » doivent être « mariées » selon la rubrique « état matrimonial ». L'équipe de vérification choisit les variables à modifier lorsqu'il n'y a pas concordance. Parfois, compte tenu de l'ambiguïté des liens avec le chef de ménage, il faut définir avec le plus grand soin les spécifications de vérification. Par exemple, dans de nombreux pays, un beau-frère est aussi bien le frère d'un conjoint (et n'est pas alors nécessairement marié) que le conjoint d'un frère ou d'une sœur (et devrait alors être marié).

289. Les valeurs déclarées pour la variable « lien avec le chef de ménage » font actuellement apparaître plusieurs autres problèmes plus contemporains. Lorsque deux personnes non mariées, de sexes opposés, cohabitent sans être mariées, la variable « lien avec le chef de ménage » pourrait être mise à la valeur « partenaire non marié » ou « conjoint ». Si le recensement ou l'enquête a prévu un code pour les partenaires non mariés, alors l'état matrimonial ne devrait pas être mis à la valeur « marié », sauf si la personne est mariée à une personne différente de celle avec laquelle elle vit.

290. De manière analogue, lorsque, à notre époque, des personnes du même sexe cohabitent, soit elles sont liées par une relation sentimentale, soit elles ne le sont pas : dans le deuxième cas de figure elles peuvent être codées en tant que « colocataires » ou « camarades de chambre » ou « personnes non apparentées » ; dans le premier la catégorie « partenaire non marié » est sans doute appropriée dans certains pays. Ensuite, l'équipe de vérification doit également décider de l'état matrimonial correspondant approprié.

2. SEXE (P3A)

291. L'information sur le sexe est sans doute l'une des caractéristiques les plus faciles à recueillir, mais sa vérification mérite réflexion. Elle constitue l'une des variables les plus importantes, puisque la plupart des caractéristiques dé-

mographiques sont analysées par sexe. L'imputation d'une valeur à la variable « sexe » doit donner lieu à différentes comparaisons avec d'autres variables. Dans certains cas, l'attribution du sexe doit reposer sur des différences entre les sexes de personnes apparentées, généralement le chef de ménage et le conjoint, mais aussi entre parents et parents du conjoint. Les réponses « non valide » ou « inconnu » ne doivent généralement pas être laissées pour la variable « sexe » compte tenu de son importance particulière. Aussi, convient-il de réfléchir soigneusement à la meilleure façon d'obtenir des résultats qui reflètent effectivement la situation réelle d'un pays.

292. Lorsqu'une personne n'est ni le chef de ménage, ni son conjoint, il n'existe aucune autre personne de référence; il y a lieu alors de vérifier d'autres rubriques de son enregistrement : s'il contient des données sur la fécondité, il y a lieu d'attribuer alors le code sexe féminin. Toutefois, si la réponse concernant le sexe de cette personne est manquante ou non valide, par exemple, mais s'il existe un conjoint dont le sexe est indiqué, alors le programme de vérification peut imputer à cette personne le sexe opposé.

a) *Cas dans lequel le code du sexe est valide, mais où le chef de ménage et le conjoint sont du même sexe*

293. Lorsque les données font apparaître des indications fortement contradictoires, il y a lieu de modifier le code attribué au sexe, même si celui-ci est valide; l'enregistrement considéré mentionne par exemple la présence d'un deuxième couple marié, alors que le ménage compte déjà un chef de ménage et un conjoint. Si les deux personnes du deuxième couple déclarent le même sexe, il est alors possible d'utiliser les données sur la fécondité, notamment pour déterminer le sexe de chacune. Ensuite, l'enregistrement erroné peut être modifié.

b) *Cas dans lequel une donnée sur la fécondité est associée à un homme ou ne figure pas en ce qui concerne une femme adulte*

294. Le programme de vérification peut détecter des données sur la fécondité et/ou une indication d'enfants à la maison concernant un homme, erreur susceptible d'être liée au numéro de classement de la mère sur la liste ou à une variable similaire. En l'absence de conjoint, il est possible d'attribuer à la variable sexe la valeur sexe féminin, au lieu de supprimer les données sur la fécondité. De manière analogue, une personne adulte de sexe féminin, à laquelle aucune information de fécondité ni aucun enfant n'est associé, peut dans certaines circonstances définies par l'équipe de vérification, faire l'objet d'une modification de la variable sexe, passant à la valeur sexe masculin.

c) *Code de sexe non valide et présence d'un conjoint*

295. Si la valeur de la variable sexe est laissée en blanc ou n'est pas valide, le programme de vérification doit utiliser les réponses recueillies pour le lien avec le chef de ménage et pour le sexe du conjoint, si le sexe indiqué est valide, pour déterminer le code correct à attribuer. Si la variable lien avec le chef de ménage est mise à la valeur « chef de mé-

nage » le programme vérifie ensuite la présence éventuelle d'un conjoint (en vérifiant s'il existe à l'intérieur du ménage une personne dont la variable « lien avec le chef de ménage » a la valeur « conjoint »). Une fois déterminé le code attribué au sexe du conjoint, le code correspondant au sexe opposé est alors attribué au chef de ménage.

d) *Non-validité du code de sexe du conjoint*

296. Si la personne considérée est liée au chef de ménage en temps que « conjoint », et si le sexe du chef de ménage est indiqué, le programme attribue alors à cette personne le sexe opposé à celui du chef de ménage.

e) *Code de sexe non valide et présence d'indication quant au sexe féminin*

297. Il existe dans le questionnaire de nombreux indices permettant d'établir si la personne interrogée est une femme. Si le programme n'a pas encore déterminé le sexe de la personne en question et en présence de tout indice permettant d'établir qu'il s'agit d'une femme, alors il convient d'attribuer le sexe féminin à l'enregistrement concernant cette même personne. Par exemple, une modification de ce type est possible lorsque la vérification détecte une ou plusieurs données sur la fécondité. Parmi les données sur la fécondité figurent le nombre d'enfants, le nombre d'enfants vivant dans ce ménage, le nombre d'enfants vivant dans un autre lieu, le nombre d'enfants décédés et le nombre d'enfants nés vivants au cours des 12 derniers mois. Par ailleurs, il pourrait également s'agir de la mère d'une autre personne du ménage, de telle sorte que le numéro de ligne de cette personne doit être égal au numéro de ligne de la mère d'un autre membre du ménage.

f) *Code de sexe non valide, lorsqu'il s'agit du mari du conjoint*

298. Si la personne considérée est le mari d'un autre membre du ménage (d'après une rubrique mentionnant le numéro de ligne du mari) il convient d'attribuer à la variable sexe la valeur sexe masculin.

g) *Code du sexe non valide en l'absence d'indications suffisantes pour déterminer le sexe*

299. Si l'équipe de vérification n'utilise pas la méthode d'imputation dynamique, il faut alors attribuer un code « sexe inconnu ». Malheureusement, cela signifie que tous les tableaux devraient comporter une colonne supplémentaire ou une rangée supplémentaire, ou encore une série de colonnes ou de rangées relatives aux personnes de sexe inconnu. Or, puisque le sexe est une variable binaire, il est possible d'attribuer alternativement l'une ou l'autre valeur, en commençant par l'une des deux, puis en affectant la valeur du sexe opposé à la deuxième donnée non valide et en continuant de la même façon.

h) *Remarques concernant les rapports des valeurs imputées de la variable sexe*

300. Le sexe féminin est susceptible d'être attribué plus souvent si l'on procède par imputation statique. Les femmes adultes étant les seules personnes faisant l'objet de données sur la fécondité, leur choix s'écarte légèrement d'une sélection aléatoire; c'est pourquoi, faute d'informations suffisantes, une personne pour laquelle on ne dispose d'aucune information est plus vraisemblablement de sexe masculin que de sexe féminin. Il est donc important d'envisager l'établissement de matrices d'imputation qui tiennent compte des proportions globalement observées entre les sexes.

3. DATE DE NAISSANCE ET ÂGE (P3B)

301. L'âge est l'une des caractéristiques dont l'obtention et la vérification sont particulièrement délicates. Toutefois, cette variable est sans doute la plus importante parce que pratiquement toutes les caractéristiques démographiques sont étudiées en fonction de l'âge. Sa vérification exige des comparaisons approfondies avec les autres variables. Dans la plupart des cas, l'âge imputé doit être déduit des différences enregistrées entre les âges des personnes apparentées. Si ce type d'imputation n'est pas possible, il convient alors d'utiliser d'autres caractéristiques figurant dans l'enregistrement concernant la personne considérée. Il est probable que la vérification exige une série de matrices d'imputation, notamment de l'âge selon le sexe, l'état matrimonial, le lien avec le chef de ménage et la fréquentation scolaire; selon la différence d'âge entre mère et enfant; selon la différence d'âge entre mari et femme; et selon la différence d'âge entre chef de ménage et conjoint.

a) *Age et date de naissance*

302. La vérification structurelle calcule l'âge à partir de la date de naissance. En premier lieu, il y a néanmoins intérêt à examiner la différence entre l'âge et la date de naissance. Tel qu'indiqué dans la publication *Principes et recommandations* (Nations Unies, 1998, par. 2.88), les renseignements sur l'âge peuvent être obtenus soit en demandant la date de naissance (année, mois et jour), soit en demandant directement l'âge au dernier anniversaire.

303. L'utilisation de la date de naissance fournit des renseignements plus précis et il convient de l'utiliser chaque fois qu'on le peut. Lorsque le jour ou même le mois exacts de la naissance ne sont pas connus, on peut indiquer la saison de l'année. La question relative à la date de naissance est appropriée dans tous les cas où les recensés connaissent cette date, qu'ils utilisent le calendrier solaire ou lunaire, et que les années portent un chiffre ou soient identifiées selon la tradition populaire par une série cyclique de noms. Toutefois, il est très important que le recenseur et le recensé soient bien d'accord sur le type de calendrier auquel correspond la date de naissance. Si certains recensés risquent de donner une réponse correspondant à un calendrier différent de celui des autres recensés, il faut prévoir d'indiquer sur le questionnaire quel est le calendrier utilisé. Il n'est pas souhaitable que le recenseur tente de convertir la date d'un calen-

drier à un autre. Le mieux est que cette conversion soit effectuée au moment de la conversion informatisée des données (Nations Unies, 1998, par. 2.89).

304. La question directe relative à l'âge risque de livrer des renseignements moins précis pour plusieurs raisons. Même si toutes les réponses sont basées sur la même méthode de calcul de l'âge, le recensé peut ne pas bien comprendre s'il lui est demandé d'indiquer son âge au dernier anniversaire, au prochain anniversaire ou à l'anniversaire le plus proche. En outre, d'autres problèmes peuvent se poser : l'âge peut être arrondi au chiffre le plus proche se terminant par zéro ou par cinq; le caractère estimatif de l'âge indiqué n'est pas nécessairement spécifié comme tel et des déclarations d'âge inexacts peuvent très bien être faites délibérément. (Nations Unies, 1998, par. 2.90).

305. Nombre de bureaux de recensement/statistique recueillent soit la date de naissance, soit l'âge, mais ne demandent pas conjointement les deux informations. Tel qu'indiqué dans la publication *Principes et recommandations* (Nations Unies 1998), l'âge exprimé en nombre d'années révolues revêt une très grande importance. Les vérifications l'utilisent fréquemment et l'âge constitue souvent une dimension des matrices d'imputation. Fait par ailleurs encore plus notable, nombre de politiques nationales reposent sur l'âge, de telle sorte qu'il faut s'employer à obtenir des données de qualité optimale. Toutefois, même dans des conditions idéales, l'âge ne sera pas toujours indiqué. Il faut donc veiller à ce qu'il soit calculé correctement et de façon compatible avec les autres réponses de chacun des membres du ménage.

b) *Relation entre âge et date de naissance*

306. Lors de la vérification structurelle, l'âge doit être calculé s'il n'a pas été recueilli parallèlement à la date de naissance. La vérification de l'âge dans le cadre des vérifications individuelles sera un bon test de cohérence interne des enregistrements et d'un enregistrement à l'autre; toutefois, une première étape consiste à calculer l'âge à partir de la date de naissance et de la date du recensement. Il importe de contrôler l'âge calculé d'après la date de naissance pour vérifier qu'il coïncide avec les valeurs limites associées à la date du recensement. L'âge des enfants nés au cours de l'année du recensement, mais après la tenue de ce dernier sera calculé comme égal à -1 et doit donc être corrigé. Il conviendra probablement d'omettre les enfants dont la date de naissance est postérieure à la date du recensement. Si toutefois, après examen, la date de naissance s'avère erronée, en raison du dénombrement ou du traitement des données, d'autres variables devraient être utilisées pour améliorer l'estimation de l'âge.

c) *Cas dans lequel l'âge calculé dépasse la valeur limite supérieure fixée*

307. Pour les recensements effectués en l'an 2000 et après, la plupart des pays choisiront d'enregistrer les quatre chiffres de l'année de naissance. Pour celle-ci le domaine de variation acceptable ira de la fin des années 1800 (peut-être à partir de 1880), jusqu'à l'année du recensement. Bien que

l'utilisation de trois chiffres soit suffisante pour les traitements informatiques, ce choix risque de dérouter aussi bien les agents de recensement que le personnel du bureau. Parfois, l'âge calculé dépasse la limite supérieure définie par le recensement et doit donc être ajusté. Si le recensement a lieu en l'an 2000 et si une personne déclare être née en 1850, l'âge calculé égal à 150 ans sera vraisemblablement en dehors du domaine de variation acceptable et devra donc être modifié.

d) *Vérification de l'âge*

308. Le programme de vérification doit s'assurer de la cohérence de l'âge déclaré de la personne avec l'âge déclaré de la mère, du père, ou de l'enfant de la personne en question. La vérification doit établir l'existence d'une différence d'âge minimum entre l'âge du père ou de la mère et l'âge de l'enfant. En cas d'imputation de l'âge, les contrôles de cohérence doivent utiliser des données telles que le nombre d'années passées dans la circonscription administrative, la durée de séjour (de résidence) et la dernière année d'études accomplies (niveau de formation). Toutes ces vérifications doivent être faites avant de modifier l'âge ou d'effectuer une imputation.

ENCADRÉ 4. *Vérification et imputation de l'âge*

La vérification et l'imputation de l'âge doivent se dérouler comme suit :

- a) Attribuer une valeur à l'âge lorsque l'âge est laissé en blanc;
- b) Contrôler l'âge minimum des personnes mariées;
- c) Contrôler l'âge minimum du chef de ménage;
- d) Contrôler l'âge minimum des parents; et
- e) Effectuer tout contrôle propre au pays.

309. La vérification doit commencer par un contrôle de validité. Si l'âge est valide les spécialistes peuvent souhaiter vérifier la cohérence de l'âge de la personne en question avec celui de sa mère (si la mère fait partie des membres du ménage) et avec l'âge des enfants (s'il s'agit d'une femme et si ses enfants font partie du ménage). En cas d'incohérence des âges, l'âge de la personne doit être noté et modifié ultérieurement.

e) *Vérification de l'âge en présence du chef de ménage et du conjoint*

310. L'étape suivante vise à déterminer la présence éventuelle d'un conjoint. Si tel est le cas, la validité de l'âge du conjoint doit être vérifiée (âge d'au moins x années, selon l'âge minimum à la date du mariage défini par le pays considéré). Si l'âge n'est pas cohérent avec cette donnée, et en cas d'utilisation de la méthode d'imputation dynamique, le programme attribuera alors une valeur d'imputation spéciale, calculée à partir de la différence d'âge du mari et de la femme. Les fluctuations des différences d'âge étant infé-

rieures à celles des âges proprement dits, la matrice d'imputation contenue dans le programme enregistrera la différence d'âge (d'après les enregistrements précédents) d'un mari et d'une femme. Cette valeur est ajoutée ou soustraite à l'âge du conjoint de la personne considérée pour obtenir un âge calculé.

311. Pour assurer la cohérence de cet âge calculé avec d'autres caractéristiques, la matrice d'imputation doit également utiliser des variables telles que l'état matrimonial, la durée de résidence et la dernière année d'études accomplies. Lorsque ces variables ne sont pas prises en compte il peut en résulter un âge calculé inférieur au nombre d'années vécues dans la localité par la personne ou inférieur au nombre d'années de scolarité impliqué. Par exemple, la matrice d'imputation peut indiquer un âge de 8 ans, alors que la personne peut avoir déclaré avoir déjà passé 10 ans à l'adresse mentionnée. Sans utilisation des autres variables, lorsque le programme de vérification contrôle la durée de séjour dans la localité, une autre matrice d'imputation aura pour effet de substituer une valeur erronée à une valeur exacte de la durée de séjour dans la circonscription administrative.

f) *Vérification de l'âge du chef de ménage en l'absence de conjoint, mais en présence d'un enfant*

312. Lorsque la détermination de l'âge du chef de ménage ne peut procéder par comparaison avec l'âge du conjoint, le programme peut ensuite examiner le lien avec le chef de ménage. Si la variable « lien avec le chef de ménage » est mise à la valeur « chef de ménage », le programme de vérification peut rechercher dans les autres enregistrements relatifs au ménage considéré la présence éventuelle d'un fils ou d'une fille dont l'âge est réputé exact. Le programme vérifie l'âge du fils ou de la fille et détermine un âge calculé par une imputation dynamique fonction de la « différence d'âge », selon une technique semblable à celle décrite ci-dessus pour des époux. De manière analogue, l'âge calculé tient compte de la durée de séjour et du niveau de formation atteint. L'âge ainsi obtenu sera cohérent avec ces variables et la prise en compte dans la matrice d'imputation de la durée du séjour et du niveau de formation évitera des erreurs évidentes.

g) *Vérification de l'âge du chef de ménage en présence d'un parent de ce dernier*

313. Lorsqu'une personne ne peut être classée dans l'une des catégories ci-dessus, le programme peut vérifier la présence d'un parent parmi les membres du ménage. Dans cette hypothèse, il est alors possible de calculer un âge au moyen d'une matrice d'imputation utilisant la différence d'âge. La différence d'âge entre un enfant et un parent peut être beaucoup plus importante qu'entre époux. Pour cette raison, cette vérification intervient seulement en cas d'échec de la méthode utilisant la différence d'âge mari/femme. L'âge calculé doit tenir compte des caractéristiques de formation et de la dernière année d'étude accomplies, de la durée de résidence dans la circonscription administrative, de l'état matrimonial, de la fécondité et de l'activité économique. Le programme doit supposer qu'une personne a at-

teint l'âge minimum acceptable si il/elle a déjà été marié(e), a des enfants ou déclare une activité économique quelconque.

h) Vérification de l'âge du chef de ménage, en présence d'un petit-enfant de ce dernier

314. Lorsqu'une personne ne peut être classée dans l'une des catégories ci-dessus, le programme peut vérifier la présence d'un petit-enfant parmi les membres du ménage. Dans cette hypothèse, il est alors possible de calculer un âge au moyen d'une matrice d'imputation utilisant la différence d'âge. La différence d'âge entre le chef de ménage et le petit-enfant varie beaucoup plus qu'entre époux ou encore entre chef de ménage et enfant. Pour cette raison le programme effectue cette vérification seulement en cas d'échec de la méthode utilisant la différence d'âge mari/femme et la différence chef de ménage/enfant. L'âge calculé doit tenir compte des caractéristiques de formation et de la dernière année d'étude accomplie, de la durée de résidence dans la circonscription administrative, de l'état matrimonial, de la fécondité et de l'activité économique. Le programme doit supposer qu'une personne a atteint l'âge minimum acceptable si il/elle a déjà été mariée(e), a des enfants ou déclare une activité économique quelconque.

i) Vérification de l'âge du chef de ménage en l'absence de toute autre donnée d'âge

315. Lorsqu'une personne ne peut être classée dans l'une des catégories décrite ci-dessus, le programme peut vérifier la présence parmi les membres du ménage d'une autre personne apparentée ou non au chef de ménage. Si une telle personne existe, et si son âge est indiqué, l'équipe de vérification doit décider d'utiliser toute information disponible au moyen d'une matrice d'imputation tenant compte de la différence d'âge. Toutefois, les différences d'âge entre le chef de ménage et les autres personnes apparentées ou non présentent de telles variations que l'équipe de vérification peut très bien décider de renoncer purement et simplement et de se contenter d'utiliser d'autres variables pour procéder à l'imputation dynamique de l'âge du chef de ménage. De toutes façons, cette vérification est exécutée uniquement après échec des méthodes utilisant la différence d'âge entre mari et femme, chef de ménage et enfant, chef de ménage et parent et chef de ménage et petit-enfant. Quelles que soient les modalités de calcul de l'âge, le résultat doit tenir compte des caractéristiques d'éducation, notamment de la dernière année d'études accomplie, ainsi que de la durée de séjour dans la circonscription administrative, de l'état matrimonial, de la fécondité, et de l'activité économique. Le programme doit supposer qu'une personne a atteint l'âge minimum acceptable si il/elle a déjà été mariée, a des enfants ou déclare une activité économique quelconque.

j) Vérification de l'âge du conjoint après détermination de l'âge du chef de ménage

316. La vérification de l'âge du conjoint intervient généralement en même temps que celle relative au chef de ménage, puisque la vérification de l'âge de chacun utilise des informations concernant l'autre. Si toutefois la vérification

est effectuée séparément lorsque l'âge du conjoint est non valide ou incohérent avec d'autres variables, la meilleure estimation de l'âge du conjoint doit se faire au moyen d'une matrice d'imputation dynamique, en fonction de différentes variables dont la différence d'âge avec le chef de ménage. Comme ci-dessus, l'âge calculé doit tenir compte des caractéristiques de formation et de la dernière année d'étude accomplie, de la durée de résidence dans la circonscription administrative, de l'état matrimonial, de la fécondité et de l'activité économique. Le programme doit supposer qu'une personne a atteint l'âge minimum acceptable si il/elle a déjà été mariée(e), a des enfants ou déclare une activité économique quelconque.

k) Vérification de l'âge des autres couples mariés du ménage, lorsque l'âge de l'une des personnes est connu

317. La vérification doit en premier lieu établir si l'enregistrement considéré est celui d'une personne mariée. Si tel est le cas, le programme peut alors rechercher dans les autres enregistrements relatifs au ménage le conjoint de la personne en question. S'il ne trouve aucun conjoint, le programme passe à la deuxième partie de la vérification. S'il trouve un conjoint, il doit alors vérifier la validité de son âge (âge d'au moins x années, selon l'âge minimum à la date du mariage défini par le pays considéré). Si l'âge n'est pas cohérent avec cette donnée, et si l'on procède par imputation dynamique, le programme attribue alors une valeur d'imputation spéciale, calculée à partir de la différence d'âge du mari et de la femme. Les fluctuations des différences d'âge étant inférieures à celles des âges proprement dits, la matrice d'imputation contenue dans le programme enregistrera la différence d'âge (d'après les enregistrements précédents) d'un mari et d'une femme. Cette valeur est ajoutée ou soustraite à l'âge du conjoint de la personne considérée pour obtenir un âge calculé.

318. Pour assurer la cohérence de cet âge calculé avec d'autres caractéristiques, la matrice d'imputation doit également utiliser des variables telles que l'état matrimonial, la durée de résidence et la dernière année d'études accomplie. Lorsque ces variables ne sont pas prises en compte il peut en résulter un âge calculé inférieur au nombre d'années vécues dans la localité par la personne ou inférieur au nombre d'années de scolarité impliqué.

l) Vérification de l'âge d'un enfant après détermination de l'âge du chef de ménage

319. S'il s'agit du fils ou de la fille du chef de ménage, il est possible d'obtenir un âge calculé à partir de l'âge du chef de ménage, de la différence d'âge, de la durée de séjour et du niveau de formation. Comme ci-dessus, l'âge calculé doit tenir compte des caractéristiques de formation et de la dernière année d'étude accomplie, de la durée de résidence dans la circonscription administrative, de l'état matrimonial, de la fécondité et de l'activité économique. Le programme doit supposer qu'une personne a atteint l'âge minimum acceptable si il/elle a déjà été marié(e), a des enfants ou déclare une activité économique quelconque.

m) *Vérification de l'âge d'un parent
après détermination de l'âge du chef de ménage*

320. S'il s'agit d'un parent du chef de ménage, il est possible d'obtenir un âge calculé à partir de l'âge du chef de ménage, de la différence d'âge, de la durée de séjour et du niveau de formation. L'âge calculé doit tenir compte des caractéristiques de formation et de la dernière année d'étude accomplie, de la durée de résidence dans la circonscription administrative, de l'état matrimonial, de la fécondité et de l'activité économique. Le programme doit supposer qu'une personne a atteint l'âge minimum acceptable si il/elle a déjà été marié(e), a des enfants ou déclare une activité économique quelconque.

n) *Vérification de l'âge d'un petit-enfant
après détermination de l'âge du chef de ménage*

321. S'il s'agit d'un petit-enfant du chef de ménage, il est possible d'obtenir un âge calculé à partir de l'âge du chef de ménage, de la différence d'âge, de la durée de séjour et du niveau de formation. Le programme doit supposer qu'une personne est âgée d'au moins 12 ans si il/elle a déjà été marié(e), a des enfants ou participe à une activité économique quelconque.

o) *Vérification de l'âge de toutes les autres personnes*

322. L'équipe de vérification doit définir des matrices d'imputation appropriées concernant les autres personnes du ménage, apparentées et non apparentées. Les principes directeurs adoptés à cet effet dépendront du type particulier de recensement ou d'enquête et des caractéristiques socio-économiques du pays. Par exemple, une personne qui a déjà été mariée, qui a eu des enfants ou qui a participé à une activité économique, est susceptible d'avoir au moins l'âge minimum défini pour le pays considéré. Compte tenu de cette information, en cas d'utilisation de la méthode d'imputation dynamique, la valeur imputée de cette façon ne doit pas être inférieure à l'âge minimum. De manière analogue, si une personne fréquente un établissement scolaire, possède une certaine instruction ou est capable de lire et d'écrire, mais n'a pas la qualité de chef de ménage, n'a jamais été mariée ou n'exerce aucune activité économique, alors il convient de la classer dans un groupe dont l'âge est inférieur à l'âge minimum défini pour les adultes, mais supérieur ou égal à l'âge minimum requis pour l'école. La valeur de la matrice d'imputation peut ensuite être déterminée pour les personnes dont l'âge est inférieur à l'âge minimum de fréquentation de l'école. Bien qu'imparfaite, cette technique limite le domaine de variation des valeurs possibles de la matrice d'imputation.

4. ETAT MATRIMONIAL (P3C)

323. D'après la définition figurant dans la publication *Principes et recommandations* (Nations Unies, 1998, par. 2.96-2.98), l'état matrimonial désigne la situation d'une personne au regard des lois ou coutumes de son pays concernant le mariage. Il y a lieu d'adopter pour cette caractéristique, une classification comprenant au moins les catégories

suyvantes, mais non limitées à celles-ci : a) célibataires (personnes n'ayant jamais été mariées); b) personnes mariées; c) veufs non remariés; d) personnes divorcées non remariées et e) personnes mariées mais séparées. Dans certains pays, il faudra peut-être établir une sous-catégorie de la catégorie b pour désigner les personnes contractuellement mariées, mais ne vivant pas encore maritalement. Dans tous les pays, la catégorie e doit comprendre à la fois les séparations de fait et les séparations de droit qui peuvent, si on le souhaite, constituer deux catégories distinctes. Bien que les couples séparés puissent être considérés comme toujours mariés (vu qu'ils ne sont pas libres de se remarier), ni l'une ni l'autre des sous-catégories de la catégorie e ne doit être incorporée dans la catégorie b. Dans certains pays, il faudra tenir compte des unions coutumières (auxquelles le droit coutumier confère un caractère légal et obligatoire) et des unions extra-légales, ces dernières étant souvent désignées sous le nom d'unions de fait (consensuelles).

a) *Vérification de l'état matrimonial*

324. L'équipe de vérification doit décider de la valeur appropriée de l'âge minimum au premier mariage à prendre en compte pour le recensement ou l'enquête. L'âge minimum au premier mariage (un certain âge x) peut varier en fonction de la région, du pays considéré ou du groupe ethnique. Si par exemple, les populations rurales se marient plus tôt que les populations urbaines, les règles de vérification doivent en tenir compte. Normalement, le bureau national de recensement/statistique détermine l'âge minimum d'un mariage avant de procéder au dénombrement, de telle sorte que la question est posée uniquement aux personnes plus âgées. Les personnes plus jeunes sont classées automatiquement dans la catégorie « personnes jamais mariées ». Par contre, si la question de l'état matrimonial est posée à tout le monde, l'équipe de vérification doit alors mettre au point une vérification concernant l'ensemble de la population.

b) *Attribution de la variable état matrimonial
sans imputation dynamique*

325. Bien que les données sur l'état matrimonial ne doivent être exploitées que pour les personnes âgées d'au moins x années, avec x âge minimum au premier mariage, les équipes de vérification doivent décider de l'opportunité et de l'importance des contrôles effectués. Si le pays utilise uniquement les mentions « non indiqué » ou « inconnu » pour les réponses non valides ou incohérentes, alors toute réponse inadéquate doit être remplacée par le code correspondant à la mention « non indiqué ». Si pour des personnes âgées de moins de x années la réponse « jamais mariée » ne figure pas, elle doit alors être imputée.

c) *Attribution de l'état matrimonial
avec imputation dynamique*

326. Avec imputation dynamique, la vérification de l'état matrimonial doit consister : a) à imputer une valeur lorsqu'une donnée n'est pas comprise dans le domaine de

variation; et b) à contrôler la cohérence entre l'état matrimonial déclaré, le lien avec le chef de ménage et l'âge.

d) *Le conjoint devrait être marié*

327. Toutes les personnes auxquelles le code « conjoint » du lien avec le chef de ménage, devraient avoir le code « personne mariée » en ce qui concerne la rubrique « état matrimonial ».

e) *Conjoint dans un couple marié*

328. Si le numéro de ligne du conjoint de la personne A (personne B) est une variable, alors la personne A devrait être indiquée en tant que conjoint de la personne B. En outre, A et B devraient être l'un et l'autre mariés (tel qu'indiqué au paragraphe 327) et de sexes opposés.

f) *S'il a la qualité de conjoint, le chef de ménage doit être marié*

329. Si aucune valeur n'a été attribuée à l'état matrimonial, mais si la valeur du lien avec le chef de ménage est « chef de ménage », le programme doit vérifier la présence éventuelle du conjoint (en vérifiant la variable « lien avec le chef de ménage » des autres membres). Dans l'affirmative, le programme attribue l'état matrimonial « marié » au chef de ménage.

g) *Chef de ménage, sans conjoint, ni enfant*

330. Si le conjoint ne figure pas parmi les membres du ménage et si la personne considérée est de sexe masculin avec des enfants présents, le programme impute l'état matrimonial en fonction de l'âge et de la présence d'enfant. Si le ménage ne comprend pas d'enfant, le programme peut alors imputer l'état matrimonial en fonction de l'âge et de l'absence d'enfants. Un individu de sexe masculin recensé en tant que chef de ménage, mais dont la femme est absente du ménage en question, est très vraisemblablement divorcé, séparé ou veuf.

h) *Imputation en cas d'échec de toutes les autres tentatives*

331. Pour les personnes dont les codes sont hors limites et auxquelles les tests ci-dessus ne permettent pas d'attribuer un code approprié, la vérification suivante doit porter sur l'âge. Si un âge valide inférieur à l'âge x est mentionné, l'état matrimonial « personne jamais mariée » doit être attribué. Dans tous les autres cas, il convient d'attribuer une valeur au moyen d'une matrice d'imputation. La matrice d'imputation devrait être définie par sexe et par âge (deux dimensions); par sexe, par âge et par lien avec le chef de ménage (trois dimensions) ou encore, par sexe, par âge, par lien avec le chef de ménage et par nombre d'enfants (quatre dimensions). Là encore, les équipes de vérification devant avoir déterminé l'ordre dans lequel s'effectuent les contrôles, il importe lors de la définition des matrices d'imputation de bien distinguer les rubriques qui ont été vérifiées de celles qui ne l'ont pas été. Si seule la vérification du sexe et du lien avec le chef de ménage a précédé celle de l'état matrimonial,

alors la matrice d'imputation doit autoriser la mention « inconnu » pour les autres rubriques.

i) *Relation entre l'âge et l'état matrimonial des jeunes*

332. Pour toutes les personnes qui déclarent un état matrimonial valide autre que « jamais marié », il y a lieu d'effectuer un contrôle de cohérence avec l'âge. Toutes les personnes qui ont déjà été mariées doivent avoir au moins x années, x étant l'âge minimum convenu dans le pays considéré pour une personne dite mariée. Si cet âge est inférieur à x ou s'il est laissé en blanc, d'autres contrôles de cohérence doivent être effectués en fonction des valeurs d'autres variables pertinentes (telles que le nombre d'enfants ou l'activité économique). Si les valeurs attribuées à ces variables ne sont pas valides, il y a lieu d'attribuer la valeur « jamais marié » ou « célibataire » à l'état matrimonial; dans tous les autres cas, l'état matrimonial ne doit pas être modifié.

5. AGE AU PREMIER MARIAGE (P4F)

333. D'après la publication *Principes et recommandations* (Nations Unies, 1998, par. 2.142), la « date du premier mariage est le jour, le mois ou l'année où le premier mariage a eu lieu. Dans les pays où la date du premier mariage est difficile à obtenir, il est recommandé de recueillir des renseignements sur l'âge au mariage ou sur la durée écoulée depuis le mariage (durée du mariage). Il faut prendre en compte non seulement les premiers mariages contractuels et les unions de fait, mais également les mariages coutumiers et religieux. Pour les femmes qui sont veuves, séparées ou divorcées au moment du recensement, il faut obtenir la date de la dissolution du premier mariage, l'âge à la dissolution ou le nombre d'années écoulées depuis. Les renseignements éventuels sur la dissolution du premier mariage livrent les données nécessaires pour calculer indirectement, lors du traitement, la durée du premier mariage. Dans les pays où les déclarations sur la durée du mariage sont plus fiables que l'âge de la femme concernée, les tableaux sur les naissances totales en fonction de la durée du mariage donnent de meilleures estimations de la fécondité que celles que l'on peut calculer à partir des données relatives aux enfants nés vivants, ventilées selon l'âge de la femme. Les données sur la durée du mariage peuvent être obtenues en soustrayant de l'âge du moment, l'âge au mariage, ou directement en calculant le nombre d'années depuis que le mariage a eu lieu.

334. La date du premier mariage doit être mentionnée pour toutes les personnes déjà mariées (ou uniquement pour les personnes de sexe féminin, selon la publication *Principes et recommandations*). Le programme doit vérifier la correspondance suivante : aucune information ne doit figurer dans le cas des personnes qui n'ont jamais été mariées, mais une indication valide de jour, de mois et d'année doit figurer dans le cas des personnes qui l'ont déjà été. Les équipes de vérification sont tenues de décider de l'éventuelle nécessité de mentionner un jour et un mois valides. Les pays n'utilisant pas la méthode d'imputation dynamique peuvent attribuer la valeur « inconnu » au jour et au mois; les pays utilisant la méthode d'imputation dynamique peuvent imputer

une valeur du jour et du mois lorsque cette donnée est manquante.

a) *L'état matrimonial des personnes qui n'ont jamais été mariées doit être laissé en blanc*

335. Les personnes qui n'ont jamais été mariées ne doivent pas déclarer d'âge au premier mariage. Lorsqu'une donnée valide est inscrite dans le cas d'une personne qui n'a jamais été mariée, l'équipe de vérification doit décider de modifier l'état matrimonial ou de laisser en blanc l'âge de la personne. S'il faut modifier l'état matrimonial, les pays utilisant uniquement la mention « non indiqué » inscriront ce code. Les pays utilisant la méthode d'imputation dynamique doivent généralement utiliser l'âge et le sexe pour obtenir une réponse adéquate à la question sur l'état matrimonial.

b) *Une valeur doit être indiquée dans le cas des personnes qui ont déjà été mariées*

336. Quant à l'année du premier mariage, les pays qui n'utilisent pas la méthode d'imputation dynamique peuvent attribuer la valeur « non indiqué » ou « inconnu ». Avec la méthode d'imputation dynamique, les pays ont la possibilité d'utiliser d'autres variables (par exemple, âge du conjoint ou différence d'âge entre conjoints, nombre d'enfants et naissances au cours de l'année écoulée) pour déterminer une valeur adéquate de l'année du premier mariage.

6. FÉCONDITÉ : NOMBRE D'ENFANTS (P4A)
ET NOMBRE D'ENFANTS VIVANTS (P4B)

337. Le « nombre d'enfants » est égal au nombre total d'enfants nés vivants, en excluant par conséquent les enfants mort-nés, les fausses couches et les avortements. Parfois, les démographes utilisent l'expression « nombre d'enfants vivants », mais nous utiliserons ci-après le terme « nombre d'enfants ».

338. La population sur laquelle il convient de recueillir des données pour chacun des sujets énumérés dans cette section est constituée des femmes de 15 ans et plus (ou ayant atteint l'âge minimum acceptable défini le cas échéant), quel que soit leur état matrimonial ou de celles qui entrent dans certaines sous-catégories, telles que les femmes non célibataires. Toutefois, pour les pays n'utilisant pas de données concernant les femmes de 50 ans et plus, il faudra s'efforcer d'obtenir des renseignements concernant les femmes âgées de 15 à 50 ans; pour les enquêtes sur les taux de fécondité récents, il peut être indiqué en revanche d'abaisser de plusieurs années la limite d'âge inférieure (Nations Unies, 1998, par. 2.121).

a) *Informations recueillies sur la fécondité*

339. La publication intitulée *Principes et recommandations concernant les recensements de la population et de l'habitat*, première révision, (Nations Unies, 1998) recommande l'obtention de renseignements concernant trois informations touchant à la fécondité : nombre total d'enfants, date de naissance du dernier enfant né vivant et âge de la mère à la naissance du premier enfant né vivant. Les répon-

ses concernant l'âge au mariage, la date du mariage ou sa durée, permettent d'améliorer les estimations de la fécondité, fondées sur le nombre total d'enfants. De plus, de nombreux pays poursuivent la collecte d'informations sur le nombre d'enfants, ce qui facilite en particulier les études rétrospectives de fécondité.

340. Les recensements et les enquêtes recueillent auprès de toutes les femmes des données sur la fécondité, en tenant compte d'un âge minimum défini par pays et parfois également un âge maximum.

b) *Principes généraux applicables à la vérification des données sur la fécondité*

341. La vérification des données concernant la fécondité vise à assurer leur cohérence interne vis-à-vis des données d'âge:

a) Le nombre total d'enfants nés vivants ne peut être supérieur à l'âge de la personne, majoré d'un âge minimum défini par pays (voir la section ci-dessous intitulée « âge à la première naissance » pour la vérification visant à déterminer la différence d'âge minimum entre la mère et l'aîné des enfants nés vivants);

b) Le nombre total d'enfants ne peut être supérieur à la somme du nombre d'enfants qui vivent dans l'unité d'habitation, ou dans un autre lieu et de ceux qui sont décédés. Lorsque le nombre total est supérieur à la somme des éléments constitutifs, les équipes de vérification doivent décider des éléments prioritaires à prendre en compte de façon à pouvoir effectuer les ajustements nécessaires;

c) Si les questions portent conjointement sur les enfants vivants et les enfants décédés, le nombre total de ces derniers ne peut être supérieur au nombre total d'enfants;

d) Le nombre d'enfants ne peut être inférieur à la valeur indiquée sous la rubrique « naissances au cours des 12 derniers mois »;

e) Selon le pays et le nombre réel d'enfants, ainsi que le nombre d'enfants vivants, il est possible d'utiliser une matrice d'imputation du nombre de naissances au cours des 12 derniers mois, afin de pouvoir attribuer une réponse en fonction de l'âge et du nombre d'enfants; il faut toutefois procéder avec le plus grand soin, lorsque cette rubrique est laissée en blanc. Pour la plupart des pays, un blanc signifie l'absence de naissance. L'attribution de valeurs particulières risque de biaiser les données;

f) Les pays recueillent parfois le nombre d'enfants, le nombre d'enfants vivants et différentes données sur la fécondité, en les classant par sexe. Dans ces conditions, les vérifications figurant dans le présent document portent sur les agrégats, mais les pays souhaitent le cas échéant effectuer des contrôles supplémentaires pour tenir compte des autres informations disponibles. Ces contrôles supplémentaires consistent notamment à vérifier que le nombre d'enfants de sexe masculin est égal à la somme du nombre d'enfants vivants et d'enfants décédés de sexe masculin et que par ailleurs, le nombre d'enfants de sexe féminin est égal à la somme du nombre d'enfants de sexe féminin vivants et décédés. Puisque les vérifications portant sur le nombre d'en-

fants ne distinguent pas les sexes, des ajustements appropriés doivent être introduits lorsque les totaux ne correspondent pas à la somme des parties.

c) *Relation entre le nombre total d'enfants vivants et le nombre d'enfants*

342. Les données sur le nombre d'enfants et le nombre d'enfants vivants permettent d'établir des estimations indirectes des taux de fécondité et de mortalité. Les résultats des recensements ou des enquêtes sont classés par cohorte de femmes d'une année ou de cinq années. Différents algorithmes établissent des estimations de taux de mortalité constant ou évolutif. Toutefois, afin d'obtenir les meilleurs résultats possibles, il incombe aux équipes de vérification de déterminer la procédure appropriée de contrôle des données disponibles.

343. La difficulté à laquelle se heurte la mise au point d'une vérification générale vient du fait que les différents pays n'exigent pas les mêmes informations. Ainsi, les jeux d'informations suivants sont recueillis dans différents pays :

- a) Nombre d'enfants uniquement;
- b) Nombre d'enfants et nombre d'enfants vivants (sexes réunis ou séparés);
- c) Nombre d'enfants, nombre d'enfants vivants, nombre d'enfants décédés (sexes réunis ou séparés);
- d) Nombre d'enfants, nombre d'enfants vivants à la maison, nombre d'enfants vivants ailleurs et nombre d'enfants vivants et décédés (sexes réunis).

d) *Vérification lorsque seul le nombre d'enfants est indiqué*

344. Si le pays n'utilise pas la méthode d'imputation dynamique, il convient d'attribuer la valeur « inconnu » lorsque le nombre d'enfants indiqué est non valide ou manquant. Dans les pays qui utilisent la méthode d'imputation dynamique, les spécialistes doivent décider d'appliquer ou non ladite méthode à tous les éléments d'information. Dans ce dernier cas, il est possible d'obtenir le nombre d'enfants, d'après le seul âge de la femme et d'après au moins une autre caractéristique. L'emploi d'un tableau unidimensionnel de l'âge de la mère offre une autre possibilité. Parmi les

autres caractéristiques à utiliser figurent des informations, telles que le niveau de formation ou la religion, étant entendu que les différences à cet égard se traduisent par des taux de fécondité différents et ce dans nombre de pays.

e) *Vérification en cas d'indication du nombre d'enfants et du nombre d'enfants vivants*

345. Si les rubriques concernant le nombre d'enfants et le nombre d'enfants vivants comportent conjointement des réponses, le programme de vérification doit établir les points suivants :

- a) Cohérence interne des informations fournies (le nombre d'enfants est-il supérieur ou égal au nombre d'enfants vivants ?);
- b) Compatibilité de chaque élément d'information avec l'âge de la femme;
- c) Compatibilité du nombre d'enfants avec le nombre d'enfants nés au cours de l'année passée (ou dernière naissance), si cette information est recueillie.

346. Les données concernant le nombre d'enfants et le nombre d'enfants vivants permettent aux démographes d'établir des estimations indirectes des taux de mortalité. Aussi les vérifications doivent-elles conserver le lien entre ces deux rubriques. Parfois, l'une des deux réponses est fournie, l'autre inconnue. Une vérification simple consisterait à supposer l'absence de décès parmi les enfants et à considérer que ces deux nombres sont identiques. Toutefois, dans cette hypothèse, l'estimation indirecte du taux de mortalité ne tiendrait pas compte des nouveaux-nés décédés après la naissance, sous-estimant ainsi le taux de mortalité et surestimant l'espérance de vie. Lorsque les cas de ce type rencontrés dans le cadre du recensement ou de l'enquête sont peu nombreux, la qualité des données est peu altérée. Toutefois, si leur fréquence est relativement élevée, comme cela devrait être le cas dans les pays procédant par estimation indirecte, les répercussions pourraient être importantes. La figure 27 donne un exemple de ce type de situation.

347. Dans cet exemple le conjoint déclare 5 enfants, mais pour une raison quelconque, le nombre d'enfants vivants n'a pas été mentionné. La personne interrogée ou l'agent de recensement n'a pas indiqué la valeur en ques-

FIGURE 27. Exemple de ménage comportant des données sur la fécondité

Personne	Lien avec le chef de ménage	Sexe	Age	Nombre d'enfants	Nombre d'enfants vivants
1	Chef de ménage	1	60		
2	Conjoint	2	60	5	99
3	Fille	2	40	3	3
4	Petite-fille	2	20	1	1
5	Petite-fille	2	18	0	0
6	Petite-fille	2	1		

Note : 99 = Données manquantes ou non valides.

tion, à moins que l'opérateur de saisie l'ait mal introduite. De nombreux pays mettent au point une vérification ayant pour effet d'attribuer la valeur 5 au nombre d'enfants vivants, d'après le nombre total d'enfants. Toutefois, cette façon de procéder introduit un biais dans les données.

348. En fait, il n'y a aucunement lieu de modifier la valeur indiquée. Les pays qui n'utilisent pas la méthode d'imputation dynamique peuvent choisir de conserver la mention « inconnu »; évidemment, ce choix introduit également un biais, puisqu'il est ainsi convenu que les réponses « inconnues » et « connues » ont la même distribution dans les tableaux de résultats. Un pays qui a besoin de données concernant le nombre total d'enfants et le nombre d'enfants vivants pour établir des estimations indirectes du taux de mortalité, a par ailleurs vraisemblablement des problèmes de déclaration des données. En l'occurrence, le maintien de valeurs inconnues au sein du jeu de données risque de biaiser l'analyse finale. Ainsi, les enregistrements de femmes dont le nombre total d'enfants ou le nombre d'enfants vivants est codé « inconnu » ne peuvent servir à l'estimation d'un taux de mortalité estimé, parce que la différence entre nombre total d'enfants et nombre d'enfants vivants ne peut être déterminée.

349. Les pays qui procèdent par imputation dynamique devraient envisager la détermination de l'élément d'information manquant en s'appuyant au moins sur l'autre donnée sur la fécondité et sur l'âge de la femme. Les matrices d'imputation peuvent être mises à jour en présence d'informations valides concernant l'âge de la femme, le nombre d'enfants et le nombre d'enfants vivants et peuvent ensuite être utilisées lorsque l'élément d'information est manquant. Ainsi, en l'absence d'information sur le nombre total d'enfants, la matrice d'imputation pourra utiliser l'âge de la femme et le nombre d'enfants vivants. Par ailleurs, en l'absence d'information sur le nombre d'enfants vivants, la matrice d'imputation utilisera l'âge de la femme et le nombre total d'enfants.

350. De plus, il importe, lors de l'établissement des matrices d'imputation de ne pas oublier la nécessité de veiller à ce que le nombre d'enfants et le nombre d'enfants vivants soient compatibles avec la différence d'âge entre la mère et l'enfant aîné (lorsque cette information est disponible), ainsi

que le nombre total d'enfants pour un âge particulier de la mère.

351. Par exemple, le nombre d'enfants imputé et l'âge de la mère pourraient différer d'au moins 12. Dès lors, une matrice d'imputation utilisant des cohortes de femmes de 5 ans imputerait de façon quasi certaine des informations parfois incompatibles.

352. La matrice d'imputation correspondant à la figure 28 fait apparaître l'âge des femmes (en ligne) et le nombre d'enfants (en colonne). Le contenu de la matrice correspond aux valeurs imputées du nombre d'enfants vivants. Les réponses recueillies seront parfois appropriées et parfois ne le seront pas. Si le programme de vérification détecte une femme de 19 ans qui a 5 enfants, le nombre de 5 enfants vivants devrait sans doute vérifier le critère de différence d'âge (une différence d'âge de 15 d'après le nombre d'enfants vivants et l'âge déclaré). En revanche, pour une femme de 15 ans, un nombre d'enfants égal à 5 (pour une différence d'âge égale à 10) comme un nombre d'enfants vivants égal à 4 (différence d'âge égale à 11) seraient l'un et l'autre inacceptables.

353. La matrice d'imputation donne de meilleurs résultats lorsque les âges concernent des femmes jeunes. Alors, seules les différences d'âge valides concernant cet âge particulier sont inscrites dans la matrice d'imputation et seules les réponses valides peuvent en être extraites.

f) *Vérification en cas de mention du nombre d'enfants, du nombre d'enfants vivants et du nombre d'enfants décédés*

354. Le « nombre d'enfants » est égal à la somme « nombre d'enfants vivants » et du « nombre d'enfants décédés ». Tout défaut de cohérence peut être corrigé tel qu'indiqué ci-dessous.

i) *Indication des trois informations*

355. En présence des trois informations, le programme de vérification doit déterminer les points suivants :

a) Cohérence interne des trois informations, le nombre d'enfants devant être égal à la somme du nombre d'enfants vivants et du nombre d'enfants décédés;

FIGURE 28. Valeurs initiales pour la détermination du nombre d'enfants vivants avec des données valides d'âge et de nombre d'enfants vivants

Nombre d'enfants	Age												
	15	16	17	18	19	20	21	22	23	24	25-29	30-34	35 +
0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	1	1	1	1	1	1	1	1	1	1	0	0	0
2		2	2	2	2	2	2	2	2	2	1	1	1
3			3	3	3	3	3	3	3	3	2	2	2
4					4	4	4	4	4	4	3	3	3
5						5	5	5	5	5	4	4	4

b) Cohérence de chacune des trois informations avec l'âge de la femme;

c) Cohérence du nombre d'enfants avec le nombre d'enfants nés au cours de l'année écoulée (ou de la dernière naissance, si cette information est recueillie).

356. Si les trois points ci-dessus sont établis, la vérification est terminée. Par contre, si l'un d'eux ne l'est pas, le programme de vérification doit corriger le défaut de cohérence correspondant. Les trois informations ne sont pas nécessairement compatibles entre elles. Par exemple, une femme peut avoir eu 5 enfants, dont seulement 2 vivants et 2 décédés. L'équipe de vérification doit choisir la variable qui aura priorité sur les autres. Dans de nombreux cas, la femme se souviendra très vraisemblablement du nombre total d'enfants qu'elle a eu, mais pourra oublier le nombre exact d'enfants décédés. Ensuite, l'équipe de vérification peut décider d'accepter les valeurs déclarées du nombre total d'enfants et du nombre d'enfants vivants et de déduire ce dernier pour obtenir une nouvelle valeur cohérente du nombre d'enfants décédés.

ii) Indication de deux informations

357. Puisque le nombre total d'enfants (NTE) est égal à la somme du nombre d'enfants vivants (EV) et du nombre d'enfants décédés (ED), si deux des trois informations sont connues, le programme informatique peut déterminer la troisième :

Si NTE et EV sont connus, $ED = NTE - EV$.

Si EV et ED sont connus, $NTE = EV + ED$.

Si NTE et ED sont connus, $EV = NTE - ED$.

Ces vérifications devraient normalement être faites en premier. Lorsque le programme a établi la validité et la cohérence des trois informations, la vérification est terminée.

iii) Indication d'une information

358. Quand une seule des trois informations est indiquée et lorsque le pays considéré n'utilise pas la méthode d'imputation dynamique, les deux autres doivent être mises à la valeur « inconnu ». Par contre, si le pays utilise la méthode d'imputation dynamique, les équipes de vérification doivent définir une façon de déterminer au moins une autre information, la troisième devant ensuite être obtenue par soustraction ou par addition. Une matrice à deux dimensions peut servir à déterminer la deuxième donnée sur la fécondité, en s'appuyant sur la première et sur l'année d'âge des femmes. Si le nombre d'enfants est connu par exemple, le nombre d'enfants vivants peut être déduit de la matrice d'imputation, tel qu'indiqué ci-dessus, puis le nombre d'enfants décédés en procédant par soustraction. De manière analogue, si le nombre d'enfants vivants est connu, le nombre d'enfants est tiré de la matrice d'imputation établie en fonction de l'âge de la femme et du nombre d'enfants vivants, le nombre d'enfants décédés étant ensuite obtenu par soustraction.

iv) Indication d'aucune information

359. Si aucune des trois informations n'est indiquée, l'équipe de vérification doit prendre des décisions quant à la marche à suivre. Lorsque le pays n'utilise pas la méthode d'imputation dynamique, les trois informations doivent être mises à la valeur « inconnu » et aucune ne doit être utilisée par les méthodes d'estimation indirecte des taux de mortalité ou de fécondité. Dans les pays utilisant la méthode d'imputation dynamique, l'application de cette méthode aux trois informations en question doit être décidée par les spécialistes.

360. Si les spécialistes décident d'utiliser l'imputation dynamique, le nombre d'enfants peut être obtenu d'après l'âge de la femme et au moins une autre caractéristique. Il est également possible d'utiliser un tableau de valeurs à une seule dimension (année d'âge de la femme). Le niveau de formation ou la religion pourraient figurer parmi les autres caractéristiques à prendre en compte.

361. Une fois la première information obtenue, l'obtention de la deuxième donnée sur la fécondité peut s'effectuer suivant les étapes décrites ci-dessus pour la procédure de vérification en cas d'indication d'une seule information. La troisième peut ensuite être déduite des deux premières. Les trois informations sont en principe compatibles parce que la mise à jour des matrices d'imputation exige au préalable la compatibilité de toutes les données. Enfin, les données sur la fécondité obtenues doivent également être compatibles avec celles de toutes les autres femmes de la même zone géographique, puisque les données concernant ces dernières servent à la mise à jour de la matrice d'imputation.

g) *Vérification en cas d'indication du nombre d'enfants, du nombre d'enfants qui vivent à la maison, du nombre d'enfants qui vivent à l'extérieur et du nombre d'enfants décédés*

i) Indication des quatre informations

362. En présence des quatre informations, le programme de vérification doit déterminer les points suivants :

a) Cohérence interne des quatre informations, le nombre d'enfants devant être égal à la somme du nombre d'enfants vivant au foyer familial, du nombre d'enfants vivant à l'extérieur et du nombre d'enfants décédés;

b) Cohérence de chacune des quatre informations avec l'âge de la femme;

c) Cohérence du nombre d'enfants avec le nombre d'enfants nés au cours de l'année écoulée (ou de la dernière naissance, si cette information est recueillie).

363. Si les trois points ci-dessus sont établis, la vérification est terminée. Par contre, si l'un d'eux ne l'est pas, le programme de vérification doit corriger le défaut de cohérence correspondant. Comme dans le cas décrit ci-dessus de déclaration de trois informations, les 4 données indiquées n'ont pas nécessairement une cohérence interne. Dans ce cas également l'équipe de vérification doit choisir la variable prioritaire. Dans de nombreux cas la femme interrogée se souvient vraisemblablement du nombre total d'enfants mis au monde, bien qu'elle puisse oublier le nombre exact de

ceux qui ont quitté le foyer familial ou le nombre d'enfants décédés. L'équipe de vérification peut alors convenir d'accepter la valeur indiquée du nombre d'enfants et du nombre d'enfants survivants (en additionnant le nombre d'enfants vivants au foyer familial et en dehors) et de soustraire celui-ci pour obtenir, pour les autres variables, de nouvelles valeurs cohérentes. L'équipe de vérification peut à cet effet devoir mettre au point des algorithmes décrivant les différentes combinaisons possibles d'événements.

ii) *Indication de trois des quatre informations*

364. Puisque le nombre total d'enfants (NTE) est égal à la somme du nombre d'enfants vivant au foyer familial (EVF), du nombre d'enfants vivant à l'extérieur (EVE) et du nombre d'enfants décédés (ED), si trois des quatre informations sont connues, le programme informatique peut déterminer la quatrième :

Si NTE, EVF et EVE sont connus $ED = NTE - EVF - EVE$.

Si EVF, EVE et ED sont connus, $NTE = EVF + EVE + ED$.

Si NTE, EVF et ED sont connus, $EVE = NTE - EVF - ED$.

Si NTE, EVE et ED sont connus, $EVF = NTE - EVE - ED$.

iii) *Indication de deux informations sur quatre*

365. Si deux informations seulement sont connues l'équipe de vérification doit décider de la marche à suivre. Par exemple, dans de nombreux pays, les femmes ne signalent pas le nombre d'enfants décédés. L'autre information particulièrement susceptible d'être omise est le nombre d'enfants vivant en dehors de l'unité d'habitation, donnée ne pouvant être obtenue directement. Il faut donc procéder soigneusement à l'élaboration du questionnaire, à la réalisation du dénombrement, ainsi qu'au dépouillement pour obtenir la meilleure qualité possible de toutes les données sur la fécondité.

366. Le nombre d'enfants qui résident dans l'unité d'habitation considérée (EVF) peut être obtenu en dénombrent les enfants recensés dans cette même unité d'habitation. Dans la mesure où une seule femme du foyer en question est caractérisée par la relation appropriée, un simple comptage doit indiquer le nombre d'enfants qui vivent au sein de l'unité d'habitation. Lorsque plusieurs femmes sont caractérisées par ce type de relation, le programme de vérification est encore applicable, en supposant que les enfants sont recensés immédiatement après la mère au cours de la collecte de données. En cas d'échec de toutes les tentatives, les pays procédant par imputation dynamique pourraient attribuer le nombre d'enfants vivant dans l'unité d'habitation considérée d'après l'âge de la mère et d'après l'une des autres variables connues. Voir les règles générales énoncées ci-dessous en matière d'imputation des données individuelles sur la fécondité à partir des autres informations disponibles et de l'âge de la mère. Il importe d'utiliser l'année d'âge de la femme dans tous les cas où cela est possible, ainsi que le nombre d'enfants qui vivent dans l'unité d'habitation, qui vivent en dehors ou qui sont décédés.

367. A titre d'exemple, supposons que le nombre d'enfants et le nombre d'enfants décédés soient des données va-

lides, et que le nombre d'enfants vivant dans le ménage considéré et le nombre d'enfants vivants à l'extérieur ne le soient pas. Dans ce cas, il est possible de déterminer le nombre d'enfants vivant au foyer familial en dénombrent les enfants caractérisés par le lien approprié avec la mère, (en supposant que la mère soit le chef du ménage). Trois des quatre informations seront alors connues, et la quatrième, le nombre d'enfants vivant à l'extérieur, peut être déterminé en effectuant la soustraction :

$$EVE = NE - EVF - ED.$$

368. Toutefois, lorsque deux informations seulement sont connues, le nombre d'enfants, ainsi que le nombre d'enfants vivant au foyer familial devront plus vraisemblablement faire l'objet d'un nouveau codage. D'ordinaire, les femmes signalent volontiers le nombre d'enfants qu'elles ont eu, tandis que les renseignements concernant le nombre d'enfants vivant au foyer familial peuvent être obtenus par observation ou auprès des personnes interrogées lors du dénombrement; or, ces possibilités n'existent pas dans le cas des enfants vivant à l'extérieur ou des enfants décédés. Les opérations de vérification peuvent alors faire appel à une matrice d'imputation à partir de l'âge de la femme et du nombre d'enfants (NE) ou mieux encore de l'âge de la femme, du nombre d'enfants (NE) et du nombre d'enfants vivant au foyer familial (EVF). Ces variables permettront de déduire les informations requises à partir des renseignements concernant une femme présentant les caractéristiques identiques quant au nombre d'enfants vivant à l'extérieur (EVE).

369. Les pays utilisant uniquement la matrice à deux dimensions [âge de la femme et nombre d'enfants (NE)], n'incluant donc pas la troisième dimension du nombre d'enfants vivant au foyer familial (EVF), risquent d'obtenir pour le nombre d'enfants vivant à l'extérieur (EVE) une valeur incompatible avec les deux autres éléments d'information. Dans le cas d'une femme âgée de 25 ans et d'un nombre d'enfants $NE = 5$, la matrice d'imputation peut indiquer un nombre d'enfants vivant à l'extérieur égal à 3. Si le nombre d'enfants vivant au foyer familial est de 2, la vérification ne soulève aucun problème. La valeur à attribuer au nombre d'enfants décédés est égale à 0 et les données sur la fécondité doivent alors être les suivantes :

$$NE = 5, EVF = 2, EVE = 3 \text{ et } ED = 0.$$

370. Toutefois, le nombre d'enfants vivant au foyer familial pourrait prendre en fait la valeur 4, seuls l'âge de la femme et le nombre d'enfants étant utilisés pour déterminer le nombre d'enfants vivant à l'extérieur. Le choix d'une valeur égale à 3 pour le nombre d'enfants vivant à l'extérieur entraînerait ensuite une incompatibilité entre les différentes informations. En effet, le nombre d'enfants (5) serait alors inférieur à la somme du nombre d'enfants vivant au foyer familial (4) et à l'extérieur (3), soit 7 au total. Aussi conviendrait-il d'utiliser une matrice à 3 dimensions : pour un nombre d'enfants égal à 5 et 4 enfants vivant au foyer familial, la matrice d'imputation pourrait contenir une valeur égale à 1 pour le nombre d'enfants vivant à l'extérieur (et la valeur 0 serait obtenue par soustraction pour le nombre d'enfants décédés). Sinon, la valeur contenue dans la ma-

trice d'imputation devrait être égale à 0 pour le nombre d'enfants vivant à l'extérieur et à 1 (par soustraction) pour le nombre d'enfants décédés. Il faudrait établir des matrices d'imputation analogues pour les autres couples de valeurs connues, tel qu'indiqué à la figure 29.

371. Dans chaque cas, deux des quatre informations sont connues. La troisième est obtenue par imputation dynamique et la quatrième par soustraction ou addition. Les équipes de vérification doivent choisir le meilleur chemin à suivre en fonction des caractéristiques culturelles en présence.

iv) *Indication d'une seule information*

372. Lorsqu'une seule des quatre informations est connue, la situation est encore plus problématique. Les pays doivent alors décider d'une procédure à suivre lorsque si peu d'informations sont disponibles. En cas de recours à la méthode d'imputation dynamique, la première matrice d'imputation utiliserait, tel qu'indiqué plus haut, une donnée comme l'année d'âge de la femme considérée, ainsi qu'une donnée connue pour créer une matrice à deux dimensions permettant d'imputer l'une des autres variables. Lorsque deux éléments sont déterminés, les deux autres restent par définition inconnus. Aussi, le fait de continuer à utiliser la méthode d'imputation dynamique pour le troisième élément ne devrait pas introduire une incompatibilité avec les autres éléments puisqu'ils sont inconnus. Le processus décrit ci-dessus pour deux éléments d'information connus et pour deux éléments inconnus est utilisé pour déterminer un troisième élément. Ensuite, le quatrième élément est obtenu par soustraction. Les quatre informations ainsi déterminées devraient être compatibles. Voir au paragraphe 376 la solution éventuellement applicable lorsqu'une seule information est connue.

v) *Indication d'aucune information*

373. Lorsqu'aucune de ces quatre informations n'est indiquée, l'équipe de vérification doit décider de la marche à

suivre en l'absence de toute donnée connue. Sans recours à la méthode d'imputation dynamique, tous les éléments doivent être mis à la valeur « inconnu », et aucun d'entre eux ne doit être utilisé par des méthodes indirectes d'estimation du taux de mortalité ou de fécondité. En cas d'imputation dynamique, les spécialistes doivent décider s'ils souhaitent appliquer cette méthode à tous les éléments en question. Voir au paragraphe 376 la description d'une autre solution envisageable lorsqu'aucune information n'est connue.

374. Si les spécialistes optent pour la méthode d'imputation dynamique, le nombre d'enfants peut être obtenu d'après l'année d'âge de la femme et au moins une autre caractéristique. Il est également possible d'utiliser un tableau à une seule dimension (uniquement l'année d'âge de la mère). Parmi les autres caractéristiques à utiliser figurent des informations, telles que le niveau de formation ou la religion, étant entendu que les différences à cet égard se traduisent par des taux de fécondité différents et ce dans nombre de pays.

375. Une fois la première information obtenue, l'obtention de la deuxième donnée sur la fécondité peut s'effectuer suivant les étapes décrites ci-dessus pour la procédure de vérification avec indication d'une seule information. La troisième peut ensuite être déduite des deux premières et le quatrième élément peut être obtenu par soustraction. Les quatre informations sont en principe compatibles parce que la mise à jour des matrices d'imputation exige au préalable la compatibilité de toutes les données. Enfin, les données sur la fécondité obtenues doivent également être compatibles avec celles de toutes les autres femmes de la même zone géographique, puisque les données concernant ces dernières servent à la mise à jour de la matrice d'imputation.

h) *Importance de l'utilisation d'un donneur unique pour toutes les données sur la fécondité*

376. Si tant est que cela soit possible il importe tout particulièrement d'imputer toutes les informations à partir d'une seule femme lorsqu'aucune indication n'est fournie.

FIGURE 29. Exemple de matrice d'imputation à définir pour des couples de valeurs connues

<i>Si ces valeurs sont connues...</i>		<i>Utiliser la matrice d'imputation dynamique pour connaître l'une des valeurs suivantes (et procéder ensuite par soustraction ou addition)</i>	
Nombre d'enfants	Nombre d'enfants vivant au foyer familial	Nombre d'enfants vivant à l'extérieur	Enfants décédés
Nombre d'enfants	Nombre d'enfants vivant à l'extérieur	Nombre d'enfants vivant au foyer familial	Enfants décédés
Nombre d'enfants	Enfants décédés	Nombre d'enfants vivant au foyer familial	Nombre d'enfants vivant à l'extérieur
Nombre d'enfants vivant au foyer familial	Nombre d'enfants vivant à l'extérieur	Nombre d'enfants	Nombre d'enfants vivant à l'extérieur
Nombre d'enfants vivant au foyer familial	Enfants décédés	Nombre d'enfants	Nombre d'enfants vivant à l'extérieur
Nombre d'enfants vivant à l'extérieur	Enfants décédés	Nombre d'enfants	Nombre d'enfants vivant au foyer familial

Pour être sûr que toutes les informations proviennent de la même source féminine, il faut parfois établir des matrices d'imputation qui utilisent toutes les données sur la fécondité. En pareille circonstance, la mise à jour des matrices d'imputation pourrait intervenir uniquement lorsque le programme de vérification a établi la compatibilité de toutes les informations sur la fécondité.

7. FÉCONDITÉ : DATE DE NAISSANCE
DU DERNIER ENFANT NÉ VIVANT (P4C)

377. Les renseignements sur la date de naissance (jour, mois et année) du dernier enfant né vivant et le sexe de l'enfant sont utilisés pour estimer la fécondité du moment. Ultérieurement, au moment du traitement de l'information, on pourra utiliser le nombre d'enfants nés vivants au cours des 12 derniers mois précédant immédiatement la date du recensement pour estimer les naissances vivantes au cours des douze derniers mois. On obtient ainsi des données plus précises qu'à partir des informations sur le nombre d'enfants qu'a eu la femme au cours des 12 mois précédant immédiatement le recensement, pour calculer les taux de fécondité par âge et d'autres indicateurs de fécondité (Nations Unies, 1998, par. 2.134).

378. Il convient de noter que les informations relatives à la date de naissance du dernier enfant né vivant ne donnent pas le nombre total d'enfants nés vivants au cours de ladite période de 12 mois. Même si les données sur le dernier enfant né vivant ne comportaient pas d'erreurs, la question ne permettrait d'établir que le nombre de femmes qui ont eu au moins un enfant né vivant au cours de la dernière période de 12 mois et non le nombre des naissances, puisqu'une faible proportion des femmes peuvent avoir eu plus d'un enfant au cours d'une même année (Nations Unies, 1998, par. 2.135).

379. On ne collectera de données que sur les femmes âgées de 15 à 50 ans qui ont indiqué avoir eu au moins un enfant vivant au cours de leur existence. De même, les renseignements doivent porter sur toutes les catégories de femmes, quelle que soit leur état matrimonial, pour lesquelles on collecte des données sur les naissances totales ventilées selon le sexe. Si les données sur la descendance finale ne sont collectées que pour un échantillon de femmes, le même échantillon doit servir pour la collecte de données sur la fécondité du moment (Nations Unies, 1998, 2.136).

380. Le programme de vérification devrait porter sur les points suivants : la date du dernier enfant devrait être mentionnée pour toutes les femmes d'un âge compris entre les âges minimum et maximum définis pour le pays. Le programme devrait vérifier l'existence de cette correspondance. Par exemple, aucune information ne devrait figurer pour les hommes et les femmes n'appartenant pas au groupe d'âge ainsi défini. De plus, pour les femmes du groupe d'âge considéré et ayant au moins un enfant, le jour, le mois et l'année de la dernière naissance devraient être valides. Il incombe à l'équipe de vérification de décider si le jour et le mois doivent être valides : si la méthode d'imputation est de type dynamique, il est possible d'imputer le jour et le mois lorsque ces informations sont manquantes; sinon, la valeur « inconnu » devrait être attribuée au jour et au mois.

381. Si l'année de naissance du dernier enfant est manquante ou non valide, la valeur « non indiqué » ou « inconnu » peut être attribuée lorsque le pays n'utilise pas la méthode d'imputation dynamique; par contre, les pays qui l'utilisent peuvent déterminer la date de naissance du dernier enfant à partir d'autres variables telles que l'âge et le nombre d'enfants.

8. FÉCONDITÉ : ÂGE À LA PREMIÈRE NAISSANCE (P4G)

382. L'âge de la mère à la première naissance vivante sert à calculer indirectement la fécondité à partir des premières naissances, et renseigne sur le début de la période de procréation. Si ce sujet fait partie du recensement, il convient d'obtenir des informations pour chaque femme qui a eu au moins un enfant né vivant. (Nations Unies, 1998, par. 2.143).

383. L'âge à la première naissance est déterminé soit directement, au moyen d'une donnée explicite intitulée « âge à la première naissance », ou au moyen de la différence entre l'âge du moment de la mère et l'âge de l'aîné. L'âge le plus précoce défini par chaque pays pour la procréation n'est pas l'âge biologique le plus précoce. Si un pays fixe par exemple à 13 ans l'âge minimum acceptable à la première naissance, les personnes interrogées peuvent déclarer, ou les agents de recensement peuvent noter un âge à la première naissance de 11 ou 12 ans pour telle ou telle personne. Il incombe alors aux équipes de vérification de décider s'ils doivent modifier l'âge acceptable minimum, supprimer la naissance déclarée ou modifier l'âge de la mère ou encore son âge à la première naissance (sur la base, soit de l'âge d'un enfant, soit de son âge, en fonction des variables employées pour déterminer la différence d'âge). De manière analogue, il incombe aux équipes de vérification de contrôler l'âge maximum à la première naissance. Puisque les femmes peuvent procréer jusqu'à 50 ans et plus, la question ne se pose pas très souvent. De telle sorte que pour corriger les erreurs éventuelles, les vérificateurs doivent déterminer la véracité des valeurs aberrantes relevées.

384. Il ne faut pas oublier que les âges minimum et maximum à la première naissance (ainsi que la différence d'âge entre la mère et l'enfant le plus âgé vivant au foyer familial) doivent être conformes aux coutumes et aux traditions nationales. Les spécialistes doivent donc décider lorsqu'une valeur s'apparente à un bruit et non à une valeur fondée de l'âge à la première naissance. Une fois les règles à suivre établies, les spécialistes doivent décider de la façon de remédier aux problèmes rencontrés. Si l'on ne procède pas par imputation dynamique, le programme doit attribuer la valeur « inconnu ». En revanche, si l'on utilise la méthode d'imputation dynamique, celle-ci permet de déterminer l'âge à la première naissance d'après les données d'autres femmes d'âge voisin et qui ont eu sensiblement le même nombre d'enfants. Par ailleurs, les spécialistes qui établissent la matrice d'imputation souhaitent parfois tenir compte de facteurs tels que la résidence en zone urbaine/ rurale (lorsque la fécondité diffère d'une zone à l'autre), la participation des femmes à la population active (bien que la situation du moment à cet égard ne soit pas nécessairement la

même qu'au moment de la première naissance) et enfin, le niveau de formation.

9. MORTALITÉ (P4D)

385. Les renseignements sur les décès survenus au cours des 12 derniers mois servent à estimer le niveau et la structure de la mortalité par sexe et par âge dans les pays où l'enregistrement des faits d'état civil ne livre pas de statistiques suivies et de bonne qualité sur la mortalité. Pour tirer des estimations fiables de cette rubrique, il faut absolument obtenir les déclarations les plus complètes et les plus exactes possibles sur les décès par sexe et par âge au cours des douze derniers mois. C'est parce que les questionnaires des recensements réalisés les dernières décennies comportaient des questions détaillées sur la mortalité qu'on a vu s'améliorer les résultats des méthodes d'estimation indirectes pour le calcul de la mortalité adulte (Nations Unies, 1998, par. 2.137).

386. En théorie, il faudrait pour étudier la mortalité déterminer le nombre total de décès survenus au cours des douze mois précédant la date du recensement. Lorsqu'il n'est pas possible d'obtenir des renseignements sur les décès au cours des douze derniers mois, il est souhaitable d'en recueillir au moins sur les décès d'enfants de moins d'un an. Pour chaque personne décédée, il faut obtenir le nom, l'âge, le sexe et la date du décès (jour, mois, année). Il convient de bien préciser la période de référence pour éviter que les recensés commettent des erreurs d'interprétation à ce sujet. On peut par exemple définir une période de référence précise en se servant de la date d'une fête ou d'un événement historique particulier au pays (Nations Unies, 1998, par. 2.138).

387. La publication *Principes et recommandations* (Nations Unies 1998) propose de recueillir le nom, l'âge et le sexe, le jour, le mois et l'année du décès des personnes survenu pendant l'année précédant le recensement. Les pays n'utilisant pas la méthode d'imputation dynamique peuvent attribuer la valeur « inconnu » à chacune de ces variables en cas de non-validité des données. Par contre, les pays qui utilisent l'imputation dynamique pourraient employer l'âge (définition de groupes d'âge), le sexe et l'année du décès comme dimensions des matrices d'imputation des autres variables. Les autres matrices d'imputation sont vraisemblablement propres au pays et l'équipe de vérification devra se concerter pour établir les matrices d'imputation appropriées.

10. ORPHELINS DE PÈRE OU DE MÈRE (P4E) ET NUMÉRO DE LIGNE DE LA MÈRE

388. Pour collecter des renseignements sur les orphelins il faut, par deux questions directes, établir : *a*) si la mère naturelle de la personne recensée dans le ménage est encore vivante au moment du recensement; et *b*) si le père naturel de la personne recensée est encore vivant au moment du recensement. Le but est d'obtenir des informations sur les parents biologiques et donc de veiller à ne pas inclure les parents d'adoption et les parents nourriciers. Du fait qu'il existe en général au sein d'une fratrie plus d'un enfant survi-

vant qui peut indiquer s'il est orphelin, il est nécessaire de donner aux questions une forme telle qu'on évite ensuite de compter autant de pères et de mères que d'enfants. On pose pour cela deux questions supplémentaires : *c*) le recensé est-il l'enfant survivant le plus âgé de sa mère; et *d*) le recensé est-il l'enfant survivant le plus âgé de son père. Les tableaux ne devront par conséquent concerner que l'enfant survivant le plus âgé (Nations Unies, 1998, par. 2.140).

389. Les vérifications concernant la « survie de la mère » et le « numéro de ligne de la mère » sont interdépendantes et doivent être réalisées conjointement. Pour les personnes répondant autrement que par oui à la question de savoir si leur mère est vivante, il faut vérifier la validité du numéro de ligne de la mère; si cette donnée est valide, le code correspondant à la réponse « oui » doit être attribué à la rubrique « survie de la mère »; pour ces mêmes personnes, il faut également vérifier si le numéro de ligne de la mère est égal à 00 ou au numéro de ligne d'une femme d'au moins 12 ans. Dans l'une ou l'autre de ces deux situations, le programme suppose que la personne a une mère et attribue la réponse « oui » à la question concernant la survie de la mère. Par contre, si le numéro de ligne de la mère n'est pas valide et si la réponse à la question concernant la survie de la mère est « non » ou « ne sait pas », il faut alors supprimer le numéro de ligne indiqué. Dans tous les autres cas, il convient d'attribuer le code correspondant à la réponse « ne sait pas » à la rubrique « survie de la mère » et de supprimer toute valeur indiquée pour le numéro de ligne.

390. Le pays peut décider de ne pas vérifier le numéro de ligne de la mère relatif aux personnes dont la réponse est « non » ou « ne sait pas » à la question de survie de la mère. Dans tous les autres cas, le numéro de ligne pourrait être vérifié quant à sa cohérence ou devrait être imputé au moyen du lien avec le chef de ménage de la personne considérée, ainsi que du numéro de ligne, du sexe, du lien avec le chef de ménage et de l'âge de la personne indiquée comme étant la mère. S'il y a des incohérences, ou si la mère ne peut être identifiée, le code correspondant à la valeur « vivant ailleurs » pourrait alors être attribué.

B. — Caractéristiques des migrations

391. Les caractéristiques démographiques d'un pays évoluent dans le temps du fait de l'accroissement naturel (fécondité et mortalité) et du solde migratoire. Les migrations incluent les migrations à long terme (depuis la naissance) ou à court terme, déterminées soit en fonction du précédent lieu de résidence et de la durée de séjour, soit à une époque antérieure bien définie. Puisque ces informations sont souvent interdépendantes, il y a lieu parfois, pour certains pays, d'effectuer une vérification conjointe semblable à celle décrite pour les variables démographiques de base. Suivant l'approche descendante, l'ordre de réalisation des vérifications joue un rôle important puisque certains éléments doivent être vérifiés avant d'autres.

392. Les informations sur les migrations exigent souvent des codes plus détaillés que d'autres éléments, puisque pour les besoins stratégiques et pratiques des pouvoirs publics l'utilisation de subdivisions géographiques plus petites

est parfois nécessaire. Ainsi, la planification des effectifs pour une nouvelle école, ou un nouveau centre de soins peut exiger des informations détaillées sur des zones peu étendues. De plus, le recours à des vérifications et à des systèmes de codage différents peut s'avérer indispensable pour les lieux situés à l'intérieur et à l'extérieur du pays.

393. Compte tenu de l'existence de liens évidents entre ces deux rubriques, il faut vérifier la cohérence des données sur le pays de naissance et sur le nombre d'années passées dans la circonscription administrative. De plus, il y a un certain lien logique entre les réponses fournies par les différents membres du ménage. Par exemple, si aucune réponse n'est mentionnée quant au nombre d'années passées dans la circonscription administrative par un enfant, il est possible d'imputer cette information à partir de la réponse obtenue pour la mère, puis de vérifier que la valeur imputée ne dépasse pas l'âge de l'enfant.

1. LIEU DE NAISSANCE (PIC)

394. Le lieu de naissance est d'abord le pays où la personne est née. Il n'existe pas forcément de rapport entre le pays de naissance et la nationalité, qui constitue un sujet distinct (voir par. 405-409). Pour les personnes nées dans le pays où est effectué le recensement (autochtones), la notion de lieu de naissance désigne également la subdivision géographique précise où résidait la mère au moment de la naissance. Dans certains pays toutefois, le lieu de naissance pour les autochtones désigne la subdivision géographique où la naissance a effectivement eu lieu. Chaque pays aura à spécifier la définition qu'il aura choisie pour le recensement (Nations Unies, 1998, par. 2.29).

a) *Liens entre les données concernant le lieu de naissance et le nombre d'années passées dans la circonscription administrative*

395. La cohérence des données concernant le lieu de naissance (PIC) et la durée de séjour (PID) peut être vérifiée en raison des rapports étroits entre ces deux rubriques. De plus, les informations concernant les différents membres d'un ménage sont liées entre elles et des hypothèses peuvent être faites d'après des renseignements sur les autres membres de la famille sur le point de savoir si la personne considérée a ou non émigré à l'étranger.

b) *Attribution de la valeur « inconnu » aux données non valides concernant le lieu de naissance*

396. Si un pays décide de ne pas utiliser la méthode d'imputation dynamique, la valeur « inconnu » doit être attribuée à toute réponse non valide concernant le lieu de naissance. D'ordinaire, sauf omission du code, il n'y a pas lieu de procéder à une vérification de la cohérence des réponses incohérentes au sein d'une même famille ou à une vérification des zones géographiques mentionnées.

c) *Utilisation de la méthode d'imputation statique pour le lieu de naissance*

397. L'indication du pays de naissance doit être modifiée uniquement lorsqu'elle est aberrante; si la rubrique « années de séjour dans la localité » est codée par la mention « toujours », le code du pays doit être attribué au pays de naissance. Si cette rubrique contient une indication différente de « toujours », il est possible d'utiliser l'information relative à une personne précédente. Par exemple, si la personne précédente prise en compte est la mère, le nombre d'années passées par la mère dans la circonscription administrative peut être comparé à l'âge de la personne. Si ce nombre d'années est supérieur ou égal à l'âge de la personne, le code correspondant à « ce pays » doit être attribué; sinon, il y a lieu d'attribuer le « pays de naissance de la mère ». S'il est impossible d'attribuer le pays de naissance d'après les données concernant la mère, les données concernant les autres personnes apparentées peuvent être utilisées de manière analogue. S'il est impossible d'attribuer une valeur après ces différentes vérifications, la rubrique « pays de naissance » peut être mise à la valeur « inconnu ».

d) *Utilisation de la méthode d'imputation dynamique pour le lieu de naissance*

398. Comme ci-dessus, l'indication du pays de naissance doit être modifiée uniquement si elle est aberrante. Si la mention « toujours » correspond au nombre d'années passées dans la circonscription administrative, il y a lieu d'attribuer le code « ce pays » au pays de naissance. Lorsqu'une autre mention figure, les informations concernant les autres personnes du ménage doivent alors être examinées pour obtenir des indices permettant de déterminer le pays de naissance de la personne considérée.

e) *Attribution du lieu de naissance lorsque la mère de la personne figure parmi les membres du ménage*

399. Si le pays de naissance est laissé en blanc ou n'est pas valide, et si la durée de résidence ne porte pas la mention « toujours », il est possible de vérifier la présence éventuelle de la mère de la personne considérée. Si celle-ci fait partie du ménage, la durée de résidence de la mère est alors examinée. Si la mention « toujours » a été indiquée pour le nombre d'années passées dans la circonscription, la variable « pays de naissance » doit être mise à la valeur « ce pays ». Si la mère de la personne considérée n'a pas toujours vécu dans la circonscription administrative, mais si l'âge de la personne est inférieur ou égal à la durée de séjour de la mère dans la circonscription administrative en question, le programme peut également mettre la variable « pays de naissance » à la valeur en question. Si l'âge de la personne est supérieur au nombre d'années passées par la mère dans la circonscription en question, et si le pays de naissance de la mère est valide, alors la personne se voit attribuer le même pays de naissance que sa mère.

f) *Attribution du lieu de naissance dans le cas d'un enfant du chef de ménage*

400. Si la mère de la personne ne figure pas parmi les membres du ménage, mais que la personne considérée est un fils ou une fille du chef de ménage, les renseignements tirés de l'enregistrement concernant ce dernier permettent d'effectuer plusieurs vérifications pour obtenir le lieu de naissance. Si la variable « nombre d'années passées dans la circonscription administrative » est mise à la valeur « toujours », pour le chef de ménage, le programme doit attribuer la valeur « ce pays » à la rubrique « pays de naissance » de l'enregistrement relatif à la personne. Si le « nombre d'années passées par le chef de ménage dans la circonscription administrative » n'est pas mis à la valeur « toujours », mais s'il est supérieur à l'âge de la personne, alors le programme doit attribuer la valeur « ce pays » à la rubrique « pays de naissance » de la personne considérée. Toutefois, si l'âge de cette dernière est supérieur au nombre d'années passées par le chef de ménage dans cette circonscription, alors le programme doit lui attribuer le pays de naissance du chef de ménage, si le code correspondant est correct.

g) *Attribution d'un lieu de naissance à un enfant autre qu'un enfant du chef de ménage*

401. Les imputations réalisées peuvent être relativement différentes selon que l'âge de la personne considérée est supérieur ou égal à une valeur donnée (âge x) fixée par l'équipe de vérification du pays. Si l'âge de la personne est inférieur à l'âge x, l'imputation du pays de naissance doit se faire d'après le premier enregistrement précédent concernant un enfant de moins de x années, par âge et par sexe.

h) *Attribution d'un lieu de naissance pour des femmes adultes mariées*

402. S'il s'agit d'une personne d'au moins x années, et de sexe féminin, le programme doit vérifier la présence d'un mari parmi les membres du ménage. S'il y a un mari et si le code de son pays de naissance est valide, le programme doit attribuer le code de son pays de naissance à l'enregistrement de son épouse. Par contre, si le code du pays de naissance du mari n'est pas valide, il faut examiner la valeur indiquée pour le nombre d'années passées dans la circonscription. Si cette donnée relative au mari a été mise à la valeur « toujours », il faut attribuer la valeur « ce pays » au pays de naissance de la femme. Enfin, si le nombre d'années passées par le mari dans la localité n'est pas mis à la valeur « toujours », alors il faut imputer le pays de naissance de la femme, en fonction de l'âge et du sexe.

i) *Attribution d'un lieu de naissance pour des femmes adultes non mariées*

403. Une femme d'un âge supérieur à une valeur minimum fixée par l'équipe de vérification et ne comptant pas de mari parmi les membres du ménage peut cependant être la mère d'un enfant vivant sous le même toit. Dans ce cas, le programme doit vérifier la présence de l'aîné de ses enfants.

Si cette opération n'aboutit pas, le programme peut alors imputer le pays de naissance en fonction de l'âge et du sexe. Par contre, si le pays de naissance de l'enfant est correctement codé et si le nombre d'années passées dans la circonscription administrative déclaré par la mère est supérieur à l'âge de l'enfant, le programme doit imputer le pays de naissance en fonction de l'âge et du sexe. Enfin, si le nombre d'années passées dans la circonscription administrative par la mère est inférieur ou égal à l'âge de l'enfant, le programme doit lui attribuer le pays de naissance de l'enfant.

j) *Attribution d'un lieu de naissance à des individus de sexe masculin*

404. Afin de déterminer le lieu de naissance d'un homme, le programme de vérification peut chercher à identifier son épouse ou s'il s'agit du chef d'un ménage, ses enfants. Dans un premier temps, le programme cherche à identifier l'épouse de l'homme en question. Si cette opération aboutit et si l'homme a passé dans la circonscription un nombre d'années inférieur ou égal à la durée de résidence de sa femme, alors le pays de naissance de la femme lui est attribué. Par ailleurs, si le nombre d'années passées par l'homme dans la circonscription est supérieur à la durée de résidence de sa femme, son pays de naissance doit lui être imputé en fonction de l'âge et du sexe au moyen d'une matrice d'imputation. Si l'homme en question est le chef de ménage, s'il a un fils ou une fille présent sous son toit et s'il a séjourné dans la circonscription un nombre d'années inférieur ou égal à l'âge de l'enfant, le programme doit lui attribuer le même pays de naissance que son enfant. Enfin, si sa durée de séjour dans la circonscription est supérieure à l'âge de son enfant, le programme doit procéder à une imputation du pays de naissance en fonction de l'âge.

2. NATIONALITÉ (P3D)

405. Il importe de recueillir des données sur la nationalité afin d'être en mesure de classer la population en : a) nationaux par la naissance; b) nationaux par naturalisation, que la nationalité ait été acquise par déclaration, option, mariage ou de toute autre manière; et c) étrangers. Il convient en outre de recueillir des renseignements sur la nationalité des étrangers. Il est important de consigner le nom du pays dont les recensés sont des nationaux en tant que tels et de ne pas utiliser l'adjectif de nationalité, car l'adjectif peut parfois s'appliquer également au groupe ethnique. Le codage de l'information portant sur les pays dont les recensés sont les nationaux doit être suffisamment détaillé pour permettre d'identifier à part toutes les nationalités représentées par la population étrangère vivant dans le pays. S'agissant de codage, il est recommandé au pays d'utiliser le système de codage numérique figurant dans Codes standard des pays et des zones à usage statistique (Nations Unies 1999). L'utilisation de codes normalisés pour la classification de la population étrangère par nationalité renforcera l'utilité desdites données et permettra au pays d'échanger des informations sur la population étrangère. Si un pays décide de regrouper les nationalités en catégories plus larges, il lui est recommandé d'adopter les classifications régionales et sous-régio-

nales normalisées définies dans la publication susmentionnée. (Nations Unies, 1998, par. 2.105).

a) *Vérification de la nationalité*

406. La nationalité dépend des définitions établies par chaque pays. Dans la plupart des pays, les personnes nées sur le territoire national sont automatiquement des nationaux par la naissance. Par conséquent, un programme de vérification doit examiner le rapport entre lieu de naissance et nationalité et le cas échéant attribuer ensuite le code « nationaux par la naissance » aux personnes nées dans le pays en question.

b) *Relation entre groupe ethnique/race et nationalité*

407. Certains pays recueillent en outre des informations concernant le « groupe ethnique » ou la « race », susceptibles de fournir des renseignements supplémentaires à utiliser pour déterminer la nationalité, notamment en cas de non-validité de la réponse recueillie. Pour de nombreux pays, l'origine ethnique et la nationalité des migrants de la première génération doivent présenter une cohérence pratiquement parfaite. Quant aux pays qui ont une longue tradition d'accueil de migrations internationales, cette caractéristique est sans doute moins intéressante mais peut néanmoins être prise en considération avec d'autres variables.

c) *Relation entre naturalisation et nationalité*

408. Dans les pays où la nationalité peut être obtenue par voie de naturalisation, les rubriques de recensement ne couvrent pas nécessairement les conditions requises à cet effet. Si par exemple une période de résidence est nécessaire, une rubrique intitulée « durée de résidence » pourrait servir à vérifier si cette condition de naturalisation est remplie. Ensuite, lorsqu'une personne est née à l'étranger et si la réponse à la question sur la nationalité est non valide ou incohérente, les équipes de vérification peuvent convenir d'inscrire « naturalisé » sous la rubrique « nationalité ». Quant aux personnes dont la durée de résidence ne répond pas aux conditions de naturalisation, le code « étranger » leur serait attribué, selon la méthode d'imputation statique.

d) *Rapport entre durée de résidence et nationalité*

409. La rubrique « durée de résidence » ne figure pas sur le questionnaire ou ne permet pas de déterminer de façon catégorique la nationalité, ou encore n'a pas été retenue par l'équipe de vérification en vue de son utilisation. Si la réponse fournie à la question sur la nationalité est non valide ou incohérente, par rapport au lieu de naissance, il convient de lui attribuer la valeur « inconnu » lorsqu'il n'a pas été convenu d'utiliser la méthode d'imputation dynamique. En revanche, les pays qui ont opté pour cette solution pour corriger les valeurs non valides, doivent faire appel à deux caractéristiques au moins (l'une d'entre elles étant vraisemblablement le lieu de naissance) pour obtenir des informations « connues » provenant de personnes similaires de la même zone géographique.

3. DURÉE DE RÉSIDENCE (PID)

410. La durée de résidence est la période de temps arrêtée au moment du recensement et exprimée en années révolues, pendant laquelle une personne a résidé : a) dans la localité qui est son lieu de résidence habituel au moment du recensement; et b) dans la plus ou moins grande circonscription administrative où cette localité est située (Nations Unies, 1998, par. 2.35).

a) *Vérification de la durée de résidence*

411. A l'instar du pays de naissance, la durée de résidence joue un rôle important lors de l'établissement de statistiques sur la mobilité de la population. Dans certains cas, un sous-groupe de la population peut présenter une mobilité nettement inférieure à celle de la nation considérée dans son ensemble. Les opérations de vérification concernant cette rubrique tiennent compte du lieu de naissance de la personne considérée et des réponses concernant les autres membres du ménage. La « durée de résidence » doit être vérifiée conjointement avec « le lieu de résidence antérieure » ou encore le « lieu de résidence à une date déterminée du passé ».

b) *Résidence de fait/de droit et durée*

412. Le caractère de fait ou de droit du recensement peut avoir une incidence sur le type de vérification. Etant donné que le recensement de droit recueille des informations sur la résidence habituelle, la rubrique « durée de résidence » ne couvre pas forcément la même information que dans le cadre d'un recensement de fait à l'occasion duquel les personnes sont dénombrées sur leur lieu de résidence, le soir du recensement. De plus, les codes employés et les vérifications effectuées doivent tenir compte des personnes qui ont « toujours » vécu au même endroit ou qui ne l'ont « jamais quitté ». Dans ces cas particuliers, le programme de vérification devrait sauter différents contrôles, notamment de cohérence.

c) *Relation entre l'âge et durée de résidence*

413. La première phase de la vérification devrait consister à vérifier la cohérence entre l'âge et le lieu de naissance, ainsi que la présence d'une indication valide quant au nombre d'années passées dans la localité ou dans la circonscription administrative en question; en effet, le nombre d'années passées par une personne dans un lieu quelconque ne peut dépasser son âge. De plus, une personne née à l'étranger ne peut avoir toujours vécu dans ladite localité ou circonscription administrative. Le programme doit donc attribuer la valeur « toujours » au nombre d'années passées dans la localité ou la circonscription administrative, lorsque ce nombre d'années est supérieur à l'âge et lorsque la rubrique « pays de naissance » contient la mention « ce pays ». Si le nombre d'années passées dans la localité ou la circonscription administrative est supérieur à l'âge, mais si la mention « ce pays » ne figure pas sous la rubrique « pays de naissance », il convient alors d'affecter l'âge de la personne au nombre d'années passées dans la localité ou dans la circonscription administrative : dans ce cas, on suppose que

la personne, bien que née à l'étranger, avait moins d'un an lorsqu'elle est arrivée.

d) *Relation entre lieu de naissance et durée de résidence*

414. Dans ce cas les réponses aberrantes doivent donner lieu aux mêmes contrôles que pour le lieu de naissance. Il convient à cet effet d'identifier les précédentes personnes apparentées (mère, chef de ménage, mari, enfants). L'imputation des valeurs doit s'appuyer sur les informations ainsi recueillies. Toutefois, avant d'attribuer une valeur il faut s'assurer de sa cohérence avec l'âge et le lieu de naissance de la personne dont l'enregistrement est vérifié.

e) *Pour les personnes qui ont toujours vécu au même endroit*

415. Si la réponse concernant le nombre d'années passées dans la localité ou la circonscription administrative est « toujours », mais si le pays de naissance déclaré n'est pas « ce pays », l'équipe de vérification peut décider d'attribuer l'âge de la personne à la durée de résidence dans la localité ou la circonscription administrative en question. Les spécialistes supposent alors que la personne, bien que née à l'étranger, avait moins d'un an lorsqu'elle est arrivée. L'étape suivante de la vérification consistera à vérifier la validité de la mention figurant au titre du nombre d'années de résidence dans la localité ou la circonscription administrative. Etant donné que la durée de résidence d'une personne ne peut dépasser son âge, l'âge sera donc attribué au nombre d'années passées dans la localité ou dans la circonscription administrative.

f) *Durée de résidence d'une personne d'après celle de la mère*

416. Si un code valide n'a pas été affecté à la catégorie, le programme peut vérifier dans la série d'enregistrements la présence de la mère de la personne considérée parmi les membres du ménage. Si l'issue de la recherche est positive, l'enregistrement de la mère peut fournir des informations utiles pour l'attribution des valeurs manquantes. Si la mère de la personne a toujours vécu dans la localité ou dans la circonscription administrative et si elle a déclaré pour pays de naissance « ce pays », (comme cela devrait être le cas), le programme mentionnera « toujours » sous la rubrique nombre d'années passées par la personne dans la localité ou dans la circonscription administrative. Si le pays de naissance déclaré par la mère n'est pas « ce pays », même lorsque la mention « toujours » a été enregistrée sous la rubrique « nombre d'années passées dans la localité ou dans la circonscription administrative », cela indique une erreur dans les catégories concernant la mère. Le programme ignorera ensuite le pays de naissance de la mère et attribuera la valeur de l'âge à la durée de résidence dans la localité ou la circonscription administrative. Si pour cette rubrique, la mère n'a pas indiqué la mention « toujours », mais un code valide et si l'âge de la personne est inférieur au nombre d'années passées par la mère dans la localité ou ladite circonscription, le programme de vérification effectuera un retour en arrière et vérifiera le pays de naissance de la mère si le pays de nais-

sance est « ce pays ». Le programme attribuera l'âge de cette personne au nombre d'années passées dans la localité ou la circonscription administrative. Toutefois, si l'âge d'une personne est supérieur ou égal à la durée de résidence de la mère dans la localité ou la circonscription administrative, le programme attribuera la durée de résidence en question de la mère à la durée de résidence de la personne considérée dans la localité ou la circonscription administrative.

g) *Durée de résidence d'une personne d'après la durée de résidence d'un enfant*

417. Si la personne en question est un enfant (fils ou fille), le programme de vérification doit contrôler l'enregistrement du chef de ménage pour y trouver les indications susceptibles de faciliter l'attribution de valeurs aux données manquantes concernant la durée de résidence. Si le chef de ménage est né dans « ce pays » et a toujours vécu dans la localité ou la circonscription administrative considérée, le programme attribuera la valeur « toujours » à la rubrique durée de résidence de l'enfant dans la localité ou la circonscription administrative. Si le chef de ménage a toujours vécu dans ce lieu sans néanmoins être né « dans ce pays », l'âge de l'enfant sera attribué à la durée de résidence dans la localité ou la circonscription administrative. Par contre, si le chef de ménage n'a pas fourni la réponse « toujours » à la question nombre d'années de résidence dans la localité ou la circonscription administrative, mais a inscrit un code valide, cette dernière information est utilisable si elle est compatible avec l'âge mentionné dans l'enregistrement de l'enfant faisant l'objet d'une vérification. Si l'âge de l'enfant est égal ou supérieur à la durée de résidence du chef de ménage dans la localité ou la circonscription administrative, le programme considérera cette durée relative au chef de ménage comme la durée de résidence du fils ou de la fille passée dans cette même localité ou circonscription administrative. Si l'âge de l'enfant est inférieur à la durée de résidence du chef de ménage dans la localité ou circonscription administrative, le programme attribuera une valeur selon le pays de naissance du chef de ménage. La valeur attribuée sera « toujours » si le chef de ménage est né dans « ce pays »; sinon, le programme attribuera au nombre d'années de résidence dans ladite localité ou circonscription administrative une valeur égale à l'âge du fils ou de la fille.

h) *Durée de résidence d'une personne en l'absence de toute autre information disponible*

418. En cas d'échec de toutes les tentatives ci-dessus pour obtenir une valeur valide, le programme peut attribuer la valeur « non indiqué » ou « inconnu » au nombre d'années passées dans la localité ou la circonscription administrative par cette personne. Si la valeur reste non valide, il convient d'inscrire « inconnu » sous cette rubrique, en cas de non utilisation de la méthode d'imputation dynamique. En revanche, en procédant par imputation dynamique des valeurs non valides les pays doivent utiliser au moins deux caractéristiques afin d'obtenir des informations « connues » à partir de personnes similaires de la zone géographique considérée.

4. LIEU DE RÉSIDENCE ANTÉRIEURE (PIE)

419. Le lieu où la personne recensée habitait immédiatement avant de s'installer dans la division administrative où elle a maintenant son domicile habituel (Nations Unies, 1998, par. 2.38).

a) *Vérification du lieu de résidence antérieure*

420. La rubrique « lieu de résidence antérieure » doit être vérifiée en même temps que la « durée de résidence ». Si la personne est née à cet endroit (pays, localité ou circonscription administrative, selon la rubrique de recensement) et n'a jamais déménagé, il convient alors soit de laisser en blanc cette rubrique, soit de lui attribuer un code spécifique (« n'a jamais déménagé »). Toutefois, la présence de valeurs en blanc risque de créer des problèmes lors de l'exploitation des résultats; aussi l'équipe de vérification doit-elle décider de la meilleure façon de traiter ce cas particulier.

b) *Lieu de résidence antérieure lorsque les frontières ont changé*

421. Les frontières nationales évoluent dans le temps et il faut donc veiller à ce que les systèmes de codage établissent les correspondances appropriées. De plus, les codes doivent être conçus de manière à permettre des regroupements logiques. Par exemple, tel qu'indiqué plus haut, dans un système de codage à 3 chiffres, le premier chiffre pourrait représenter le continent de résidence, le deuxième la région et le troisième le pays.

c) *Situation dans laquelle la personne n'a pas déménagé depuis sa naissance*

422. Le traitement des données met en tableau des résultats concernant certaines rubriques individuelles. Aussi les spécialistes doivent-ils vérifier qu'un code spécial correspondant à la situation « né dans ce lieu » est utilisé conjointement aux autres codes de localisation. De cette manière, le programme peut distinguer entre les personnes nées à un endroit, et celles nées en un lieu donné, mais qui ont déménagé ailleurs, à l'intérieur de la même zone géographique.

d) *Utilisation d'autres personnes faisant partie de l'unité d'habitation*

423. Lorsque le « lieu de résidence antérieure » est non valide ou incohérent, il convient généralement d'effectuer des vérifications semblables à celles qui ont été faites pour la « durée de résidence ». Le programme de vérification peut à cet effet examiner la résidence antérieure de la mère si elle fait partie de l'unité d'habitation. Il peut ensuite examiner la résidence antérieure du chef de ménage pour les enfants et les adultes dans les pays où les adultes ne déménagent pas fréquemment.

e) *Aucune autre personne appropriée pour la vérification du lieu de résidence antérieure*

424. En cas d'échec de toutes les tentatives ci-dessus pour obtenir une valeur valide, le programme peut attribuer

la valeur « non indiqué » ou « inconnu » au nombre d'années passées dans le lieu de résidence antérieure concernant cette personne. Si la valeur reste non valide, il convient d'inscrire « inconnu » sous cette rubrique, en cas de non utilisation de la méthode d'imputation dynamique. En revanche, en procédant par imputation dynamique des valeurs non valides les pays doivent utiliser au moins deux caractéristiques afin d'obtenir des informations « connues » à partir de personnes similaires de la zone géographique considérée.

5. LIEU DE RÉSIDENCE À UNE DATE DÉTERMINÉE DU PASSÉ (PIF)

425. Le lieu de résidence à une date déterminée du passé désigne la plus ou moins importante division administrative, ou le pays étranger, où la personne recensée habitait à une date déterminée précédant le recensement. La date de référence choisie doit être celle qui répond le mieux aux besoins nationaux. Dans la plupart des cas, on a jugé qu'elle devait remonter à 1 ou à 5 ans avant le recensement. La première date de référence permet d'obtenir les statistiques du moment sur la migration pendant une année; la seconde est peut-être plus commode pour analyser les migrations internationales, mais elle convient moins bien à l'analyse des migrations internes. Il faut également se demander, en choisissant la date de référence, dans quelle mesure les personnes recensées pourront se rappeler avec exactitude quel était leur lieu de résidence habituel un an ou cinq ans avant la date du recensement. Dans les pays qui effectuent des recensements quinquennaux, la plupart des personnes feront sans peine des rapprochements entre cette date et l'époque du recensement précédent. Dans d'autres cas, il sera probablement plus facile pour les personnes recensées de se rappeler des faits remontant à un an qu'à cinq. Toutefois, certains pays devront peut-être choisir un an ou cinq ans avant le recensement, ces intervalles pouvant dans les deux cas présenter des problèmes de mémoire. Les circonstances nationales peuvent faire qu'il soit nécessaire, dans certains pays de prendre pour date de référence une date susceptible d'être associée à un événement important dont chacun se souvient. L'année d'arrivée dans le pays peut, elle aussi, être une donnée utile pour l'étude des migrations internationales (Nations Unies, 1998, par. 2.40).

426. « Le lieu de résidence à une date déterminée du passé » fait l'objet du même type de vérification que le lieu de résidence antérieure. Habituellement, la question posée porte soit sur la « durée de résidence », et le « lieu de résidence antérieure » soit simplement sur le « lieu de résidence à une date déterminée du passé ». Si la personne est née là où a lieu le recensement (pays, localité ou circonscription administrative, suivant la rubrique du recensement) et n'a jamais déménagé, alors cette rubrique peut soit être laissée en blanc, soit recevoir un code spécifique correspondant à la situation « n'a jamais déménagé ». Tel qu'indiqué plus haut, les mentions laissées en blanc peuvent poser des problèmes au stade de l'exploitation des données. Aussi doit-on alors appliquer des procédures identiques à celles concernant le lieu de résidence antérieure, décrites aux trois paragraphes précédents.

C. — Caractéristiques sociales

427. Les caractéristiques sociales varient d'un pays à l'autre, mais correspondent généralement à des données qui rendent compte de divers aspects de la situation socio-culturelle qui prévaut dans le pays considéré. Les données relatives à l'éducation, notamment l'aptitude à lire et à écrire, la fréquentation scolaire et le niveau de formation, ainsi que le domaine d'étude et les diplômes de fin d'études, peuvent être classés d'après les catégories de la version révisée (1997) de la Classification internationale type de l'éducation (CITE), mise au point par l'Organisation des Nations Unies pour l'éducation, la science et la culture (UNESCO) [Nations Unies, 1998, paragraphes 2.144-2.164]. Cette section traite également de l'incapacité, des déficiences et des désavantages, ainsi que des causes de l'incapacité. La Classification internationale des handicaps : déficiences, incapacités et désavantages publiée par l'Organisation mondiale de la santé en 1980 définit un cadre commun et des définitions relatifs aux différents aspects des incapacités. Parmi les différentes caractéristiques sociales passées en revue ci-dessous figurent la religion, la langue et le groupe ethnique.

1. APTITUDE À LIRE ET À ÉCRIRE (P5A)

428. Il est préférable de recueillir des renseignements sur l'aptitude à lire et à écrire pour toutes les personnes âgées de 10 ans et plus. Dans un certain nombre de pays toutefois, certaines personnes dont l'âge est compris entre 10 et 14 ans peuvent être sur le point de savoir lire et écrire grâce à leur scolarité et le taux d'alphabétisation pour ce groupe d'âge risque d'être trompeur. En conséquence, pour une comparaison internationale de l'alphabétisation, les données sur l'alphabétisation devront être exploitées pour toutes les personnes âgées de 15 ans et plus. Lorsque les pays recueillent des données sur les personnes plus jeunes, les tableaux sur l'alphabétisation devront au moins distinguer les moins de 15 ans et les personnes âgées de 15 ans et plus (Nations Unies, 1998, par. 2.147).

429. Il incombe à chaque pays de définir l'âge minimum pour l'établissement de tableaux de résultats sur l'aptitude à lire et à écrire; de manière analogue, les équipes de vérification doivent décider de l'âge minimum pour les vérifications portant sur ce sujet, compte tenu de la nécessité éventuelle de tableaux supplémentaires pour usage interne. Lors de l'élaboration du questionnaire, les équipes de vérification doivent décider de l'âge minimum auquel les données pertinentes sont recueillies, et du niveau de formation à partir duquel la question n'a plus besoin d'être posée. Si la personne interrogée a déjà atteint un certain niveau de scolarité, l'agent de recensement n'a donc vraisemblablement pas besoin de poser la question sur l'aptitude à lire et à écrire.

430. La vérification concernant l'aptitude à lire et à écrire contrôle en premier le nombre d'années d'étude accomplies. Si ce nombre implique l'aptitude à lire et à écrire, le code correspondant à la réponse « oui » doit être attribué. En effet, les personnes ayant atteint un certain niveau de scolarité doivent être considérées comme aptes à lire et à écrire.

Ainsi, lorsque le code inscrit pour l'aptitude à lire et à écrire n'est pas valide, il convient de lui attribuer une valeur. Celle-ci doit être soit « non indiqué », soit établie au moyen d'une matrice d'imputation fondée sur des variables déterminées, telles que le nombre d'années d'étude et le sexe.

2. FRÉQUENTATION SCOLAIRE (P5B)

431. Il convient en principe de recueillir des renseignements sur la fréquentation scolaire pour toutes les personnes quel que soit leur âge. Cela concerne en particulier la population officiellement d'âge scolaire, comprise en générale entre 5 et 29 ans, mais pouvant varier d'un pays à l'autre selon la manière dont l'enseignement est organisé dans le pays. Lorsque la collecte de données est élargie au programme d'enseignement pré-scolaire et/ou à d'autres programmes d'enseignement et de formation réguliers organisés à l'intention des adultes dans des entreprises de production et de services, des organisations communautaires et d'autres institutions que les établissements d'enseignement, la fourchette des âges est à adapter en conséquence (Nations Unies, 1998, par. 2.151).

a) Vérification de la fréquentation scolaire

432. Dans chaque pays l'équipe de vérification doit décider des âges adaptés à la collecte de données sur la fréquentation scolaire. Puisque la plupart des pays subdivisent également la formation scolaire en plusieurs niveaux, lorsque ces niveaux doivent faire l'objet d'une compilation par âge, les spécialistes doivent en outre décider des groupes d'âge qui doivent correspondre aux différents niveaux de scolarisation. Il faut modifier alors les données enregistrées pour toutes les autres personnes. Au cas où le programme de vérification produit des réponses incohérentes pour la catégorie considérée, il faut alors modifier soit l'âge, soit les données de fréquentation scolaire. Généralement l'âge est déterminé au moment où cette vérification est effectuée, de telle sorte que la modification concerne les renseignements sur la fréquentation scolaire. Les agents de recensement doivent avoir pour instruction d'omettre les renseignements sur la fréquentation scolaire pour les personnes dont l'âge dépasse un certain nombre d'années déterminé au préalable; dans certains pays lorsque des personnes poursuivent à l'âge adulte des études secondaires ou supérieures, il peut s'avérer indiqué de ne pas fixer des limites supérieures à l'âge de fréquentation scolaire. On peut supposer que les réponses et les combinaisons de réponses sont testées préalablement au recensement, à l'occasion de tests préalables, de telle sorte que les décisions à ce sujet puissent être prises avant le déroulement du recensement proprement dit.

b) Fréquentation scolaire à temps plein ou à temps partiel

433. Certains pays peuvent souhaiter recueillir des informations sur la fréquentation scolaire à temps partiel ou à temps plein. Si cette rubrique figure sur le questionnaire, il peut s'avérer nécessaire de procéder à sa vérification dans le cadre de la vérification portant sur la fréquentation scolaire ou séparément.

c) *Cohérence entre la fréquentation scolaire et l'activité économique*

434. Il convient d'effectuer en premier lieu les vérifications de cohérence par rapport à d'autres rubriques importantes telles que la principale activité économique. Si la fréquentation scolaire figure parmi les valeurs possibles de la principale activité économique et si une personne a déclaré que la fréquentation de l'école constituait sa principale activité économique, la réponse « oui » doit être attribuée à la question « fréquentation de l'école », tandis que la mention « étudiant » doit figurer sous la rubrique « principale activité économique ». Autrement dit, les réponses fournies à cette question doivent être cohérentes. Dans tous les autres cas, toute réponse valide doit être acceptée.

d) *Attribution d'une valeur en cas de réponse non valide ou incohérente concernant la « fréquentation scolaire »*

435. Si la fréquentation scolaire indiquée est aberrante et si le nombre d'années d'études accomplies est valide, il convient d'attribuer une valeur au moyen d'une matrice d'imputation en fonction de l'âge, du sexe et du nombre d'années d'études. Si le code du nombre d'années d'études n'est pas valide, il convient alors d'utiliser la donnée recueillie concernant l'aptitude à la lecture et à l'écriture, pour attribuer une valeur à la « fréquentation scolaire ». Si le code de l'aptitude à la lecture et à l'écriture n'est pas valide, la valeur à inscrire sous la rubrique « fréquentation scolaire » doit être attribuée uniquement en fonction de l'âge et du sexe.

436. Il est parfois nécessaire de définir des matrices d'imputation correspondant à différentes distributions de la fréquentation scolaire par sexe et par âge (parfois par année d'âge ou par groupe d'âge d'étendue restreinte).

3. NIVEAU DE FORMATION (NOMBRE D'ANNÉES D'ÉTUDE OU NIVEAU D'ÉTUDE ATTEINT) [P5C]

a) *Vérification du niveau de formation*

437. La vérification du niveau de formation (nombre d'années d'étude ou niveau atteint) doit comporter les opérations suivantes : a) contrôle de cohérence entre une réponse valide et l'âge; et b) imputation d'une réponse lorsque la réponse d'origine est aberrante; comme ci-dessus, dans les pays qui n'utilisent pas la méthode d'imputation dynamique il convient d'attribuer la valeur « non indiqué ». Dans les pays qui utilisent par contre la méthode d'imputation dynamique, il faudra tenir compte du sexe et de l'année d'âge pour les individus plus jeunes, et d'autre part du sexe et des groupes d'âge pour les enfants légèrement plus âgés. Les pays dont les données portent conjointement sur le nombre d'années d'étude et le niveau le plus élevé atteint, sont parfois tenus de recourir à des matrices d'imputation multiples.

b) *Age minimum compatible avec le niveau de formation*

438. Les équipes de vérification de chaque pays doivent décider de l'âge minimum de fréquentation de l'école. Une fois cet âge minimum fixé, le nombre d'années d'études accomplies ne doit pas d'ordinaire dépasser l'âge de la per-

sonne considérée, majoré d'une valeur constante (qui correspond à l'âge minimum de fréquentation de l'école). Là encore, il importe d'utiliser l'année d'âge des enfants, en raison du risque d'introduction d'erreurs lors de la mise à jour des matrices d'imputation avec des groupes d'âge très étendus.

c) *Relation entre âge et niveau de formation*

439. L'équipe de vérification doit par ailleurs décider du niveau de bruit acceptable dans le jeu de données. D'ordinaire, il est préférable de modifier quelques cas exceptionnels advenant une incompatibilité entre âge et niveau de formation, plutôt que d'accepter un grand nombre de réponses effectivement incohérentes. Quand la réponse d'origine est aberrante ou incohérente avec l'âge et lorsqu'on ne procède pas par imputation dynamique, la mention « non indiqué » peut donc être inscrite. Pour les pays qui font appel à la méthode d'imputation dynamique, une réponse peut être déterminée d'après l'âge (notamment l'année d'âge des personnes d'âge scolaire), le sexe et la fréquentation scolaire. Puisque selon l'UNESCO, le « degré d'instruction » est distinct du niveau de formation, il convient sans doute de ne pas faire figurer dans la matrice d'imputation la valeur de la variable « aptitude à lire et à écrire ».

4. DOMAINE D'ÉTUDE ET DIPLÔME DE FIN D'ÉTUDES (P5D)

440. On a besoin de renseignements sur les personnes par niveau d'enseignement et domaine d'étude, pour examiner dans quelle mesure l'offre de main-d'œuvre qualifiée, dotée de telle ou telle spécialisation, correspond à la demande sur le marché du travail. Ces renseignements sont tout aussi importants pour planifier et ajuster la capacité des différents niveaux, types et secteurs des établissements d'enseignement et des programmes de formation. (Nations Unies, 1998, par. 2.158).

441. Pour les personnes de moins de 15 ans (ou dont l'âge est inférieur à une valeur pré-déterminée), il ne doit pas y avoir d'information sur le domaine d'étude et/ou le diplôme de fin d'études; pour celles de 15 ans et plus, il doit y avoir une relation entre niveau de formation et domaine d'étude et/ou diplôme de fin d'étude. Dans chaque cas, en présence de réponses non valides, les pays qui n'utilisent pas la méthode d'imputation dynamique peuvent inscrire la mention « inconnu ». Quant aux pays qui utilisent cette méthode, ils peuvent convenir de prendre en considération l'âge, le sexe, le niveau de formation et le cas échéant, l'activité professionnelle pour attribuer une valeur à inscrire sous la rubrique « domaine d'étude » et/ou « diplôme de fin d'étude ».

5. RELIGION (P3E)

442. Aux fins du recensement, la religion peut se définir soit : a) par la croyance ou tendance religieuse ou spirituelle, que cette croyance soit ou non représentée par un groupe organisé; ou b) par l'affiliation à un groupe organisé qui adhère à des dogmes religieux ou spirituels déterminés. Des pays étudiant la religion par le recensement utiliseront la dé-

finition qui répond le mieux à leurs besoins et les indiqueront dans la publication relative au recensement (Nations Unies, 1998, par. 2.109).

a) *Vérification concernant la religion*

443. La religion figure parmi les variables dont les exemples présentés au chapitre II constituent une bonne illustration. Il est possible d'obtenir une donnée valide concernant un individu, soit directement à partir d'un autre membre du ménage, si une donnée appropriée est disponible, soit à partir d'un autre chef de ménage présentant des caractéristiques similaires. L'équipe de vérification doit déterminer la séquence logique de vérification adoptée pour les autres variables sociales. Le chef de ménage doit être désigné, puis faire l'objet d'une première vérification, qu'il ou elle figure ou non en tête de liste de l'unité d'habitation. Lorsque le chef du ménage est une personne dont la religion est non valide ou inconnue, il convient alors de procéder comme suit:

b) *Indication de religion omise dans le cas du chef de ménage, mais présente pour une autre personne de l'unité d'habitation*

444. La première étape consiste à déterminer si une religion valide est mentionnée pour une autre personne de l'unité d'habitation, puis à attribuer la première religion valide identifiée.

c) *Identification de religion omise pour le chef de ménage ou pour toute autre personne de l'unité d'habitation*

445. Si aucune religion n'est mentionnée pour aucune personne du ménage, il faut soit attribuer la mention « inconnu » à cette rubrique (si la méthode d'imputation dynamique n'est pas utilisée dans le pays considéré), soit imputer une religion à partir du cas le plus récemment traité de chef de ménage présentant des caractéristiques similaires, notamment d'âge et de sexe, ou encore de langue, de lieu de naissance ou d'autres caractéristiques éventuellement appropriées suivant le contexte.

d) *Personne autre que le chef de ménage, sans mention de religion*

446. Si la personne considérée n'est pas le chef de ménage et ne déclare aucune religion, il convient alors d'attribuer la religion du chef de ménage.

6. LANGUES (P3F)

447. Il y a trois types de données relatives à la langue que l'on peut recueillir lors d'un recensement (Nations Unies, 1998, par. 2.112); ce sont les suivantes :

a) La langue maternelle, définie comme la langue habituellement parlée au foyer de la personne considérée, dans sa première enfance;

b) La langue habituelle, définie comme la langue parlée actuellement ou celle que l'intéressé emploie maintenant le plus souvent au foyer;

c) Aptitude à parler une ou plusieurs langues déterminées.

a) *Vérification de la langue*

448. Parmi les trois renseignements sur la langue pouvant figurer sur le questionnaire (langue maternelle, langue usuelle et aptitude à parler plusieurs langues déterminées), les deux premières, la langue maternelle et la langue usuelle sont liées. En présence de ces deux indications sur un questionnaire, les équipes de vérification doivent envisager de les contrôler simultanément. Si l'une d'entre elles n'est pas valide, l'autre peut servir à la détermination d'une valeur.

b) *Vérification de la langue : chef de ménage*

449. La langue figure parmi les variables dont les exemples présentés au chapitre II constituent une bonne illustration. Les équipes de vérification doivent définir le schéma logique de vérification employé pour les autres variables dites sociales, en contrôlant tout d'abord les données du chef de ménage. Si la personne dont la langue est non valide ou inconnue (langue maternelle ou langue usuelle) est le chef de ménage, il faut déterminer en premier lieu si une langue valide est indiquée pour une autre personne de l'unité d'habitation et attribuer ainsi la première langue valide. Si aucune langue valide n'est mentionnée, il faut alors soit attribuer la valeur « inconnue » si l'imputation dynamique n'est pas utilisée, soit imputer une langue d'après les données concernant le dernier chef de ménage pris en compte doté de caractéristiques similaires, notamment l'âge et le sexe, ainsi que les autres variables appropriées selon le contexte, concernant la langue et le lieu de naissance en particulier.

c) *Vérification concernant la langue : personnes autres que le chef de ménage*

450. Si la personne n'est pas le chef de ménage et si la langue n'est pas valide, il faut alors lui attribuer la langue du chef de ménage.

d) *Vérification concernant la langue : utilisation de l'origine ethnique ou du lieu de naissance*

451. La langue ou l'origine ethnique et parfois le lieu de naissance sont étroitement liés et, dans le cas de certains pays, peuvent faire l'objet d'une vérification conjointe. Aussi, les équipes de vérification doivent-elles concevoir les codes de façon à ce qu'ils reflètent les liens entre ces variables. Selon le nombre de chiffres du code et la répartition des langues nationales et des groupes ethniques, des correspondances peuvent être définies pour faciliter l'attribution de réponses adéquates en présence de données non spécifiées ou incohérentes.

e) *Vérification concernant la langue : langue maternelle*

452. Si la langue maternelle est inconnue, mais si la personne considérée est native des Philippines, il est possible de lui attribuer une langue équivalente appropriée, le taga-

log, l'ilokano ou une autre langue nationale. D'ordinaire, seul le chef de ménage fait l'objet de ce type d'attribution de la langue et le code correspondant est alors attribué aux autres membres du ménage; toutefois, l'équipe de vérification de chaque pays doit tenir compte du contexte particulier, notamment des données géographiques (par exemple, résidence urbaine ou rurale), de l'âge ou d'autres données.

f) *Vérification concernant la langue : aptitude à parler une langue déterminée*

453. L'aptitude à parler une langue déterminée est une troisième variable dont les exemples présentés au chapitre II constituent une bonne illustration. Là encore, la vérification doit porter initialement sur le chef de ménage. S'il est fait état pour ce dernier d'une langue non valide ou inconnue, la première étape consiste à déterminer si une autre personne de l'unité d'habitation fait l'objet d'un code valide concernant l'aptitude à parler la langue en question, puis à attribuer la première valeur valide observée. Par ailleurs, si aucune personne de ce type n'est identifiée, il convient alors d'attribuer à cette variable la valeur « inconnue » si le pays considéré n'utilise pas la méthode d'imputation dynamique, ou sinon à imputer l'aptitude à parler une langue déterminée en fonction des données concernant le dernier chef de ménage examiné présentant des caractéristiques similaires (par exemple, âge et sexe, mais aussi lieu de naissance et différentes variables le cas échéant, compte tenu du contexte). Si la personne n'est pas le chef de ménage, et si la valeur déclarée pour l'aptitude à parler une langue désignée est non valide, il faut alors attribuer l'aptitude déclarée pour le chef de ménage.

7. GROUPE ETHNIQUE (P3G)

454. Les groupes nationaux ou ethniques de la population sur lesquels on a besoin de renseignements dans les différents pays dépendent des particularités du pays. Parmi les critères servant à identifier les divers groupes ethniques, on peut citer les suivants : nationalité ethnique (c'est-à-dire pays ou région d'origine pour autant qu'ils diffèrent de la citoyenneté ou du pays dont la personne est ressortissante), race, couleur, langue, religion, coutumes relatives à l'habillement ou au mode d'alimentation, tribu ou diverses combinaisons de ces caractéristiques. D'autre part, certains des termes employés comme « race », « origine » ou « tribu », présentent différentes connotations. Les définitions d'experts attribuées pour chaque pays pour l'étude des caractéristiques ethniques de la population doivent donc être déterminées en fonction des groupes qu'ils cherchent à identifier. Etant donnée la nature même du sujet, les groupes varient très sensiblement d'un pays à l'autre, et il est impossible de recommander des critères de valeur internationale (Nations Unies, 1998, par. 2.116).

a) *Vérification concernant l'appartenance ethnique*

455. Plusieurs autres variables peuvent faciliter, du moins si les données correspondantes sont recueillies, la « détermination » du groupe ethnique lorsque celui-ci est non valide ou inconnu. Dans de nombreux pays, il existe une

relation entre le lieu de naissance, à l'intérieur des frontières comme à l'étranger, et le groupe ethnique. De manière analogue la « langue maternelle » est souvent un bon indicateur du groupe ethnique pour nombre de pays, puisque les catégories correspondantes, et par conséquent les codes, seront semblables sinon identiques.

b) *Vérification concernant le groupe ethnique : pour le chef de ménage*

456. Le groupe ethnique figure parmi les variables dont les exemples présentés au chapitre II constituent une bonne illustration. Les équipes de vérification doivent définir le schéma logique de vérification employé pour les autres variables dites sociales, en contrôlant tout d'abord les données du chef de ménage. Si la personne dont le groupe ethnique est non valide ou inconnu est le chef de ménage, il faut déterminer en premier lieu si un groupe ethnique valide est indiqué pour une autre personne de l'unité d'habitation et attribuer ainsi le premier groupe ethnique valide. Si aucun groupe ethnique valide n'est mentionné, il faut alors soit attribuer la mention « inconnu » si l'imputation dynamique n'est pas utilisée, soit imputer un groupe ethnique d'après les données concernant le dernier chef de ménage pris en compte doté de caractéristiques similaires (notamment l'âge, le sexe, et la langue, ainsi que les autres variables appropriées selon le contexte).

c) *Vérification concernant le groupe ethnique : personnes autres que le chef de ménage*

457. Si la personne considérée n'est pas le chef de ménage et si son origine ethnique n'est pas valide, il faut alors attribuer l'origine ethnique du chef de ménage.

d) *Vérification concernant le groupe ethnique : utilisation de la langue et du lieu de naissance*

458. L'origine ethnique, la langue et parfois le lieu de naissance sont étroitement liés et, dans le cas de certains pays, peuvent faire l'objet d'une vérification conjointe. Aussi, les équipes de vérification doivent-elles concevoir les codes de façon à ce qu'ils reflètent les liens entre ces variables. Selon le nombre de chiffres du code et la répartition des langues nationales et des groupes ethniques, des correspondances peuvent être définies pour faciliter l'attribution de réponses adéquates en présence de données non spécifiées ou incohérentes.

459. Par exemple, si l'origine ethnique est inconnue, mais si la personne parle l'une des langues des Philippines et si elle est native de ce pays, il est alors possible de lui attribuer une origine ethnique équivalente appropriée (philippine). D'ordinaire, seul le chef de ménage ferait l'objet de ce type d'attribution du groupe ethnique (tandis que ce même code serait attribué aux autres membres) mais l'équipe de vérification de chaque pays doit examiner le contexte particulier, notamment les données géographiques (par exemple, résidence urbaine ou rurale), l'âge ou d'autres rubriques.

8. INCAPACITÉ (P8A)

460. Aux fins de l'évaluation des incapacités au sein d'une population, une personne souffrant d'une incapacité sera définie comme une personne dont les activités sont limitées en genre et en nombre du fait de difficultés de longue date, d'origine physique ou mentale, ou tenant à son état de santé. Il ne sera pas tenu compte des incapacités momentanées résultant d'un état temporaire (jambe cassée), mais seulement des incapacités d'une durée supérieure à 6 mois (Nations Unies, 1998, par. 2.262).

461. La question posée aux fins du recensement des handicapés devra comporter une liste des grandes catégories d'incapacités, afin que chaque personne interrogée puisse cocher la personne correspondant à la présence ou à l'absence de chacune d'entre elles. Il est recommandé pour cela de se fonder sur la Classification internationale des handicaps, déficiences, incapacités et désavantages : vision (même avec des lunettes le cas échéant); écoute (même avec un appareil auditif); parole; déplacement/locomotion (marcher, monter les escaliers, se tenir debout); mouvements du corps (atteindre les objets, s'accroupir, s'agenouiller); préhension (saisir ou manipuler des objets avec les doigts); apprentissage (difficultés d'ordre intellectuel); arriération; comportement, problèmes psychologiques et émotionnels); soins corporels (se laver, s'habiller, se nourrir); autres (préciser). Sera considérée comme handicapée une personne qui aura fait état d'une ou de plusieurs des incapacités énumérées dans la liste (Nations Unies, 1998, par. 2.264).

a) *Vérification concernant les incapacités*

462. Lorsqu'une personne ne répond pas à cette question il est difficile de déterminer si la rubrique considérée est laissée en blanc en raison de l'absence d'incapacité, ou du fait de la réticence à répondre manifestée pour une raison quelconque par la personne interrogée. L'équipe de vérification d'un pays doit donc décider s'il faut vérifier la rubrique de la manière habituelle, soit en attribuant des codes « inconnu » (sans imputation dynamique), soit sur la base des réponses fournies par d'autres personnes (dans le cas contraire). Sinon, les spécialistes peuvent décider d'accepter uniquement les réponses spécifiant l'existence d'une incapacité et de considérer toute réponse non valide comme équivalente à la réponse « aucune incapacité ». Dans ce dernier cas, la méthode d'imputation dynamique ne serait pas utilisée.

b) *Incapacités multiples*

463. Les pays qui recueillent des informations concernant la présence d'incapacités multiples devront modifier le programme de vérification. Celui-ci devra suivre le nombre d'incapacités totales envisageables, ainsi que les répétitions et la distribution de ces incapacités. Comme ci-dessus, la plupart des pays estimeront qu'ils ne convient pas d'utiliser les données provenant d'autres personnes pour attribuer les valeurs à inscrire sous cette rubrique; aussi, dans les cas de non-validité il sera vraisemblablement nécessaire d'utiliser

les mentions « inconnu » et même « présence non établie d'une incapacité ».

9. DÉFICIENCES ET DÉSAVANTAGES (P8B)

464. On peut estimer le degré de désavantage en comparant les personnes qui font état d'une incapacité et celles qui n'en font pas état, sous l'angle d'autres caractéristiques telles que l'éducation et l'emploi. Il peut être intéressant aussi de collecter des données sur les différentes situations où les handicapés se sentent désavantagés (transport en commun, lieu de travail, manifestations sociales et culturelles). Ce type d'information pourra servir à limiter certains facteurs d'isolement des handicapés (obstacles matériels, absence de services, attitude défavorable de la collectivité ou exclusion). Une question sur les désavantages permettra de préciser la nature des difficultés qui empêchent l'intéressé de participer dans l'égalité aux activités sociales. Il faudra prendre en compte à la fois les dimensions physiques et sociales pour mieux comprendre le cadre de vie du handicapé (Nations Unies, 1998, par. 2.276).

a) *Vérification concernant les déficiences et les désavantages*

465. Dans ce cas également, l'équipe de vérification d'un pays doit décider s'il faut vérifier la rubrique de la manière habituelle, en attribuant des codes « inconnu » (sans imputation dynamique) ou sur la base des réponses fournies par d'autres personnes (dans le cas contraire). Sinon, les spécialistes peuvent décider d'accepter uniquement les réponses spécifiant l'existence d'un désavantage et de considérer toute réponse non valide comme équivalente à la réponse « aucun désavantage ».

10. CAUSES DE L'INCAPACITÉ (P8C)

466. Il est important de connaître les causes de l'incapacité pour bien planifier et évaluer les programmes de prévention. Compte tenu de la place limitée qu'offre le questionnaire de recensement, on s'enquerra des causes en s'informant du contexte dans lequel l'incapacité s'est manifestée; plutôt que de poser des questions précises sur les maladies ou des blessures déterminées. Il est recommandé de classer ces causes en 5 grandes catégories : a) congénitales/prénatales; b) affections/maladies (poliomyélite, lèpre, cataracte, etc.); c) blessure/accident/traumatisme (accident de la route; blessure due à une chute accidentelle, à un accident, à un fait de guerre, intoxication accidentelle, par exemple); d) autres causes; et e) causes inconnues (Nations Unies, 1998, par. 2.277).

a) *Vérification concernant les causes de l'incapacité*

467. L'équipe de vérification d'un pays doit décider s'il faut vérifier la rubrique de la manière habituelle, soit en attribuant des codes « inconnu » (sans imputation dynamique), soit sur la base des réponses fournies par d'autres personnes (dans le cas contraire). Sinon, les spécialistes peuvent décider d'accepter uniquement les réponses spéci-

fiant l'existence d'une cause d'incapacité et de ne pas utiliser de matrice d'imputation dynamique.

D. — Caractéristiques économiques

468. Les données sur la situation d'activité sont à collecter en principe pour l'ensemble de la population, mais dans la pratique, pour chaque personne ayant atteint ou dépassé un âge minimum, qui est déterminé en fonction de la situation particulière de chaque pays. L'âge où s'achève la scolarité obligatoire ne devrait pas être automatiquement retenu comme âge minimum pour la collecte de ces données. Les pays dans lesquels, en règle générale, les enfants sont nombreux à participer aux travaux agricoles ou à d'autres types d'activité économique (extraction minière, tissage, petit commerce) retiendront un âge minimum plus bas que celui des pays où des jeunes enfants travaillent. Les tableaux de caractéristiques économiques devraient au moins faire la distinction entre les personnes (âgées) de moins de 15 ans et celles (âgées) de 15 ans et plus; les pays où la scolarité est obligatoire au-delà de 15 ans et où des enfants n'ayant pas atteint cet âge, exercent une activité économique, devraient s'efforcer d'obtenir des données sur les caractéristiques économiques de ces enfants en vue d'assurer la comparabilité internationale, des données au moins pour les personnes âgées de 15 ans et plus. Par ailleurs, il arrive souvent que l'activité économique des hommes et des femmes ayant dépassé l'âge normal de la retraite ne soit pas prise en compte. Un effort particulier devra donc être fait pour y remédier lorsqu'on cherche à établir l'effectif de la population active. Il ne devrait pas normalement être fixé d'âge maximum, puisque bon nombre de personnes ayant dépassé l'âge de la retraite peuvent exercer une activité de façon régulière, ou occasionnelle (Nations Unies, 1998, par. 2.172).

469. Chaque pays doit fixer un âge minimum de participation à l'activité économique. Les pays désireux de collecter des données sur le travail des enfants doivent parfois choisir un âge minimum plus bas, mais ne doivent pas oublier la présence d'un certain « bruit » lorsque des enfants ne faisant pas partie de la population active y sont recensés par erreur. Une fois l'âge minimum fixé, les rubriques concernant l'activité économique sont vérifiées en vue de la mise en tableau des résultats pour les personnes âgées de x années au plus; aussi la vérification des données concernant les enfants de moins de x années sera indispensable, uniquement pour vérifier que toutes les réponses sont laissées en blanc. Pour faciliter la création de tous les tableaux, il convient d'éliminer toute réponse mentionnée pour des enfants de moins de x années.

1. SITUATION D'ACTIVITÉ (P6A)

470. La situation d'activité économique regroupe plusieurs variables économiques dont certaines sont décrites ci-dessous. Ces variables se prêtent bien à la collecte de données, mais doivent parfois être reclassées dans des catégories différentes en vue du traitement et de l'analyse.

471. La « situation du moment au regard de l'activité » rend compte de la relation d'une personne à l'activité économique, pendant une période de référence courte, une se-

maine par exemple. On considère que l'activité du moment convient davantage aux pays où l'activité économique est faiblement des facteurs saisonniers ou autres facteurs de variation au cours de l'année. On peut utiliser une période de référence d'une semaine, qui pourrait être soit une semaine récente, soit la dernière semaine civile complète, soit les 7 jours précédant le dénombrement (Nations Unies, 1998, par. 2.180).

472. D'après les Nations Unies, 1998, par. 2.182, la catégorie des personnes **pourvues d'un emploi** comprend toutes celles ayant dépassé un âge déterminé qui, pendant une période de référence courte, (une semaine ou un jour) entraient dans les catégories suivantes :

- a) Emploi salarié :
 - i) Personnes au travail : les personnes qui durant la période de référence ont effectué un travail moyennant un salaire, ou un traitement en espèces ou en nature;
 - ii) Personnes qui ont un emploi, mais ne sont pas au travail : les personnes qui ayant déjà travaillé dans leur emploi actuel, en étaient absentes pendant la période de référence et avaient des liens formels avec leur emploi, attestés par exemple par le versement ininterrompu du salaire ou du traitement, l'assurance de retourner au travail à la fin de la situation d'exception, un accord sur la date de retour après l'absence de courte durée.
- b) Emploi non salarié :
 - i) Personnes au travail : personnes qui durant la période de référence ont effectué un travail en vue d'un bénéfice ou d'un gain familial en espèces ou en nature;
 - ii) Personnes ayant une entreprise, mais n'étant pas au travail : les personnes qui durant la période de référence avaient une entreprise, qui peut être une entreprise industrielle, un commerce, une exploitation agricole ou une entreprise de service, mais n'étaient temporairement pas au travail pour une raison précise, quelle qu'elle soit.

473. La population « habituellement inactive » comprend toutes les personnes qui n'étaient classées ni comme pourvues d'un emploi, ni comme chômeurs. Il est recommandé de ventiler cette population selon les quatre catégories ci-après (Nations Unies, 1998, par. 202) :

- a) **Etudiants** : personnes des deux sexes non classées dans la catégorie dite de la population « habituellement inactive », qui pendant la période de référence, ont fréquenté un établissement public ou privé pour y recevoir une instruction complète, à quelque niveau d'enseignement que ce soit;
- b) **Personnes s'occupant du foyer** : personnes des deux sexes non classées dans la catégorie des personnes habituellement actives, (n'exerçant habituellement aucune activité économique);
- c) **Retraités et rentiers** : personnes des deux sexes n'exerçant habituellement aucune activité économique, qui ont un revenu constitué par des biens ou des placements, in-

térêts, loyers, redevances ou pensions provenant d'activités antérieures et qui ne peuvent être classées comme étudiant ou personne s'occupant du foyer;

d) *Autres personnes inactives* : personnes des deux sexes n'exerçant habituellement aucune activité économique qui reçoivent une aide de l'état ou de source privée, et toutes autres personnes non comprises dans les catégories susmentionnées.

a) *Catégories liées à la situation d'activité*

i) *Population au chômage (P6A1)*

474. Les **chômeurs** comprennent d'après la définition des Nations Unies (1998, par. 2.194), toutes les personnes ayant dépassé un âge spécifié, qui au cours de la période de référence, étaient :

a) Sans travail, c'est-à-dire pourvues ni d'un emploi salarié, ni d'un emploi non salarié;

b) Disponibles pour travailler dans un emploi salarié ou non salarié durant la période de référence;

c) A la recherche d'un travail : c'est-à-dire qui avaient pris des dispositions précises au cours d'une période récente spécifiée pour chercher un emploi salarié ou non salarié. Ces dispositions peuvent inclure : l'inscription à un bureau de placement public ou privé; la candidature auprès d'employeurs; les démarches sur les lieux de travail, dans les fermes ou à la porte des usines; sur les marchés ou autres endroits publics; l'insertion ou la réponse à des annonces dans les journaux, les recherches par relations personnelles; la recherche de terrains, d'immeubles, de machines, ou d'équipements pour créer une entreprise personnelle; les démarches pour obtenir des ressources financières, des permis et licences. Il serait utile dans la classification des chômeurs d'établir une distinction entre les personnes à la recherche d'un premier emploi et les autres demandeurs d'emploi.

475. En général, pour être classée dans la catégorie des chômeurs, une personne doit réunir les trois critères ci-dessus. Toutefois, lorsque les moyens classiques de recherche d'emploi présentent peu d'intérêt, que le marché du travail est très peu organisé ou étroit, que le taux d'absorption de la main-d'œuvre est insuffisant au moment considéré ou que la population active est composée en grande partie de travailleurs indépendants (pourvus d'un emploi non salarié), on peut appliquer la définition type du chômage en assouplissant le critère de recherche d'emploi. Cela concerne les pays les moins avancés où ce critère ne permet pas de rendre compte entièrement de l'ampleur du chômage. Cet assouplissement, qui dans les cas extrêmes peut amener à éliminer entièrement ce critère, laisse comme critères fondamentaux, ceux de « sans travail » et de « disponibles » pour travailler (Nations Unies, 1998, par. 2.195).

476. Les vérifications concernant les indications relatives au chômage — « licencié », « à la recherche d'un emploi », « aptitude éventuelle à prendre un emploi » et « dernière année d'emploi » (le cas échéant) — devraient être effectuées conjointement. De plus, les données correspondantes doivent être compatibles avec les réponses concernant l'activité économique et dans la plupart des cas, ne de-

vraient pas être fournies lorsque les rubriques concernant la durée de travail, le secteur d'activité, la profession, la catégorie de travailleurs sont renseignées. Si les spécialistes du sujet déterminent qu'une valeur doit figurer sous la rubrique « licencié » lorsque la réponse est soit laissée en blanc, soit non valide, alors une matrice d'imputation d'après l'âge et le sexe, et éventuellement le niveau de formation de la personne, pourrait être utilisée.

ii) *A la recherche d'un emploi (P6A2)*

477. La vérification concernant la rubrique « à la recherche d'un emploi » devrait être faite parallèlement à celle concernant l'indication « licencié » et « cause de la non-recherche d'emploi ». Les spécialistes du sujet devraient concevoir des vérifications d'après les réponses fournies à ces questions, afin d'imputer les autres données. La vérification devrait tenir compte de la situation locale et régionale ainsi que des différentes variables recueillies par le recensement ou l'enquête.

iii) *Population inactive du moment (P6A3)*

478. La population « inactive du moment » ou en d'autres termes n'appartenant pas à la « main-d'œuvre », comprend toutes les personnes qui n'étaient ni pourvues d'un emploi, ni en chômage pendant la période de référence courte utilisée pour déterminer l'activité du moment. Elles peuvent être classées dans les catégories suivantes, selon la cause de leur inactivité du moment :

a) Personnes fréquentant un établissement d'enseignement;

b) Personnes effectuant des travaux ménagers;

c) Retraités (pensionnés, rentiers);

d) Autres, y compris les handicapés ou les personnes victimes d'incapacités ou de déficiences.

Les vérifications concernant les personnes actuellement inactives ont été intégrées aux vérifications ci-dessus, relatives à l'activité économique.

iv) *Causes de la non-recherche d'emploi (P6A4)*

479. Cette rubrique doit être vérifiée uniquement pour les personnes enregistrées comme n'étant pas à la recherche d'un emploi; pour toutes les autres, aucune mention ne doit y figurer. Sinon, lorsqu'une mention valide figure sous « profession, secteur d'activité et état et situation d'activité », le code associé à la position « pourvu d'un emploi mais ne travaillant pas actuellement » doit être inscrit. Ce code correspond à des personnes économiquement actives, qui ont été pourvues d'un emploi, mais qui ne travaillent pas pendant la période de référence. Dans tous les autres cas, et en l'absence d'imputation dynamique, la mention « inconnu » peut être attribuée. Les pays utilisant la méthode d'imputation dynamique, peuvent par contre lui attribuer une valeur en fonction de l'âge, du sexe et de l'activité principale.

b) *Vérification concernant la situation au regard de l'activité économique*

480. On peut généralement définir les catégories suivantes au regard de l'activité économique :

- 1) Personnes qui ont un emploi et qui sont au travail;
- 2) Personnes qui ont un emploi et ne sont pas au travail;
- 3) Personnes qui ont un emploi non salarié et sont au travail;
- 4) Personnes qui ont un emploi non salarié et qui ne sont pas au travail;
- 5) Personnes à la recherche d'un emploi;
- 6) Etudiants;
- 7) Personnes qui s'occupent d'un foyer;
- 8) Retraités ou rentiers;
- 9) Autres personnes ne faisant pas partie de la population active.

481. En ce qui concerne cette variable, les quatre premières possibilités désignent des personnes économiquement actives et les quatre suivantes, des personnes qui ne sont pas économiquement actives. Les personnes dites « au travail » (catégories 1 et 3) sont pourvues d'un emploi, tandis que celles « qui ne sont pas au travail » (catégories 2 et 4) peuvent être soit en chômage, soit ne pas faire partie de la population active, selon les réponses aux questions relatives au chômage (licencié, à la recherche d'un emploi et dernière année d'emploi).

i) *Personnes pourvues d'un emploi*

482. Si l'on considère l'une des catégories de personnes économiquement actives (catégories 1 à 4), les valeurs des variables correspondant à la durée de travail, à la profession, au secteur d'activité, à la situation vis-à-vis de l'activité économique et au lieu de travail, doivent être spécifiées. Sinon, le programme de vérification doit leur être appliqué et il faut leur assigner soit la valeur « inconnu », soit des valeurs obtenues par imputation statique (*cold deck*) ou dynamique (*hot deck*). Si l'on considère les personnes des catégories 1 à 4, les rubriques « licencié », « à la recherche d'un emploi » et « dernière année d'emploi » doivent être vides. Si elles ne le sont pas, elles doivent être mises à la valeur BLANC.

ii) *Activité économique des personnes sans emploi*

483. Si l'on considère la première catégorie parmi les personnes non économiquement actives, (catégories 5 à 9), les rubriques « licencié », « à la recherche d'un emploi » et « dernière année d'emploi » doivent être complétées. Si elles ne le sont pas par des données valides, le programme de vérification doit leur être appliqué et il faut alors leur affecter soit la valeur « inconnu », soit une valeur d'imputation statique (*cold deck*) ou dynamique (*hot deck*). Si l'on considère les catégories 5 à 9, les rubriques temps de travail, profession, secteur d'activité et situation eu égard à l'activité économique doivent être vides. Si elles ne le sont pas, elles doivent être mises à la valeur BLANC.

iii) *Activité économique des étudiants et des retraités*

484. Si l'on considère la catégorie n° 6 (étudiants), les spécialistes du sujet doivent décider si la rubrique concernant la fréquentation scolaire doit contenir la mention « oui, fréquente l'école ». Si l'on considère la catégorie n° 8 (retraités), les spécialistes du sujet doivent décider si la qualité de retraité exige des personnes concernées qu'elles aient atteint un certain âge.

iv) *Activité économique non valide et déclaration de variables concernant l'emploi*

485. Si la valeur indiquée pour l'activité économique n'est pas valide, et si une valeur est mentionnée pour certaines des variables concernant l'emploi (temps de travail, profession, secteur d'activité et lieu de travail), il convient d'attribuer au code de l'activité économique de la personne interrogée une valeur allant de 1 à 4. Il faudra vraisemblablement utiliser une matrice d'imputation pour déterminer la réponse appropriée.

v) *Activité économique non valide et déclaration de variables concernant le chômage*

486. En cas d'indication d'une valeur pour l'une des variables concernant le chômage (« licencié », « à la recherche d'un emploi » et « dernière année d'emploi »), il convient d'attribuer à la variable « activité économique » un code allant de 5 à 9. Si la personne considérée fréquente l'école, il convient sans doute d'attribuer une valeur égale à 6. S'il s'agit d'une personne âgée, la valeur à attribuer est sans doute de 8. Sinon, les spécialistes du sujet peuvent convenir d'utiliser une matrice d'imputation afin d'attribuer une réponse appropriée.

vi) *Activité économique non valide et aucune déclaration concernant les variables économiques*

487. Si aucune des rubriques concernant l'activité économique ne comporte de réponse, les spécialistes du sujet souhaiteront vraisemblablement utiliser des matrices d'imputation pour déterminer la réponse la plus appropriée et imputer ensuite la valeur des autres rubriques économiques.

2. TEMPS DE TRAVAIL (P6B)

488. Le temps de travail est le temps total consacré effectivement à la production de biens et de services durant la période de référence de l'activité économique dans le recensement et comprend à la fois les heures normales de travail et les heures supplémentaires. Si la période de référence est courte (par exemple, la semaine précédant le recensement), il est recommandé de mesurer le temps de travail en heures. Si elle est longue, (par exemple, les douze mois précédant le recensement), le temps de travail est à mesurer en semaines ou en journées dans la mesure du possible. Le temps de travail devrait également comprendre le temps consacré aux activités, qui, tout en n'aboutissant pas directement à la production de biens ou de services, relève néanmoins des tâches et responsabilités afférant au travail, notamment le temps consacré à la préparation, à la réparation ou à l'entretien du lieu ou des outils de travail. Dans la pratique, on y fera éga-

lement entrer les temps morts survenant durant ces activités, notamment le temps passé à attendre et les autres poses brèves. Les temps de pose consacrés au repas, de même le temps au cours duquel l'intéressé n'a pas travaillé pour cause de vacances, de jours fériés, de maladie ou de conflits (par exemple, grèves et mises à pied) ne seront pas comptés dans le temps de travail (Nations Unies, 1998, par. 2.210).

489. Cette rubrique doit faire l'objet d'une vérification uniquement pour les personnes dont la réponse à la question sur l'activité économique porte la mention « pourvu d'un emploi, au travail » ou « pourvu d'un emploi non salarié, au travail ». Dans certains pays le temps de travail doit également être spécifié pour les personnes s'occupant du foyer. Les catégories choisies au préalable par l'équipe de vérification doivent être acceptées. En l'absence d'imputation dynamique, les rubriques laissées en blanc contenant les 0 ou les codes non numériques doivent être modifiées et mises à la valeur « inconnu »; les spécialistes du sujet peuvent souhaiter remplacer la valeur attribuée à la variable activité économique par la mention « ne travaille pas » si le temps de travail déclaré est égal à 0.

490. En cas de recours à la méthode d'imputation dynamique, les variables utilisées à cet effet comprennent au moins le groupe d'âge et le sexe, bien que d'autres variables telles que le niveau de formation, la profession ou les grandes catégories de secteurs d'activité puissent également servir.

3. PROFESSION (P6C)

491. La profession désigne le genre de travail effectué pendant la période de référence par une personne pourvue d'un emploi (ou dans le cas d'un chômeur le genre de travail effectué précédemment), quelle que soit la branche d'activité économique ou la situation de l'intéressé dans la profession (Nations Unies, 1998, par. 2.214).

492. Cette donnée doit être vérifiée uniquement pour les personnes ayant déclaré au titre de l'activité économique « pourvu d'un emploi, au travail » ou « pourvu d'un emploi non salarié, au travail ». Si la méthode d'imputation dynamique n'est pas utilisée, il convient d'inscrire la mention « non indiqué » à la place des rubriques laissées en blanc, ou comportant des valeurs nulles ou non valides.

493. Les branches d'activité économique donnent lieu généralement à la définition de codes, dont les différents chiffres correspondent à des codes principaux et secondaires de profession. Les mentions manuscrites, pratiquement inévitables dans le cas de la profession, alourdiront le travail de codage.

494. En cas de retour à la méthode d'imputation dynamique, les variables prises en compte dans la matrice d'imputation comprendront au moins le groupe d'âge et le sexe, mais aussi le niveau de formation ou les principales branches d'activité industrielle, le cas échéant.

4. ACTIVITÉ ÉCONOMIQUE (P6D)

495. D'après les Nations Unies (1998, par. 2.221) « l'activité économique se rapporte à l'activité de l'établis-

sement où la personne pourvue d'un emploi travaillait pendant la période de référence fixée pour les données sur les caractéristiques économiques (ou dans lequel elle avait travaillé en dernier lieu, si elle est au chômage). Pour plus d'informations sur le choix des emplois ou activités devant figurer dans la classification, voir le paragraphe 2.212 de la publication *Principes et recommandations*.

496. Cette donnée doit être vérifiée uniquement pour les personnes ayant déclaré au titre de l'activité économique « pourvu d'un emploi, au travail » ou « pourvu d'un emploi non salarié, au travail ». Si la méthode d'imputation dynamique n'est pas utilisée, il convient d'inscrire la mention « non indiqué » à la place des rubriques laissées en blanc, ou comportant des valeurs nulles ou non valides.

497. Les branches d'activité économique donnent lieu généralement à la définition de codes, dont les différents chiffres correspondent à des codes principaux et secondaires de profession. Les mentions manuscrites, pratiquement inévitables dans le cas de la profession, alourdiront le travail de codage.

498. En cas de retour à la méthode d'imputation dynamique, les variables prises en compte dans la matrice d'imputation comprendront au moins le groupe d'âge et le sexe, mais aussi le niveau de formation ou les principales branches d'activité industrielle, le cas échéant.

5. SITUATION DANS L'EMPLOI (P6E)

499. Par statut on entend la situation d'une personne par rapport à son emploi, c'est-à-dire le type de contrat explicite ou implicite qui la lie professionnellement à d'autres personnes ou organisations. Les principaux critères utilisés pour définir les groupes de la classification sont le type de risque économique, dont un élément est la solidité du lien entre la personne et son emploi, et le type d'autorité qu'elle exerce sur des établissements ou d'autres travailleurs. Il faut veiller à classer les actifs selon le statut dans l'emploi, en se référant aux mêmes emplois que pour les classer par profession, par activité économique et par secteur (Nations Unies, 1998, par. 2.226).

500. La population peut être classée comme suit en fonction du statut dans l'emploi (Nations Unies, 1998, par. 2.227) :

- a) Salariés, en prévoyant la possibilité de classer à part les salariés ayant un contrat stable, (y compris les salariés habituels);
- b) Employeurs;
- c) Travailleurs indépendants;
- d) Famille travailleuse cotisante;
- e) Membres d'une coopérative de production;
- f) Personnes inclassables selon le statut.

501. Les propriétaires/chefs d'entreprises constitués en société, qui sont normalement classés parmi les salariés, mais que certains motifs (de type exposé ou analyse) peuvent porter à classer dans la catégorie des employeurs, doivent être classés séparément.

502. Cette donnée doit être vérifiée uniquement pour les personnes ayant déclaré au titre de l'activité économique « pourvu d'un emploi, au travail » ou « pourvu d'un emploi non salarié, au travail ». Si la méthode d'imputation dynamique n'est pas utilisée, il convient d'inscrire la mention « non indiqué » à la place des rubriques laissées en blanc, ou comportant des valeurs nulles ou non valides. En cas de retour à la méthode d'imputation dynamique, les variables prises en compte dans la matrice d'imputation comprendront au moins le groupe d'âge et le sexe, mais aussi le niveau de formation ou les principales branches d'activité industrielle, le cas échéant.

6. REVENU (P6F)

503. Les sujets de recensement visant les caractéristiques économiques de la population présentés dans la publication intitulée *Principes et recommandations concernant les recensements de la population et de l'habitation*, première révision, portent essentiellement sur la population active telle qu'elle est définie dans les Recommandations de l'Organisation internationale du Travail (OIT), où la notion de production économique s'entend au sens indiqué dans le Système de comptabilité nationale (SCN) [Nations Unies, 1998, par. 2.165]. La population active comprend toutes les personnes des deux sexes qui fournissent ou sont à même de fournir la main-d'œuvre disponible pour la production de biens et de services économiques, tels que défini dans le SCN pendant une période de référence déterminée (Nations Unies, 1998, par. 2.166). Dans ce cadre, le revenu désigne soit : a) le revenu mensuel en espèces ou en nature que chaque personne active tire de son travail; soit b) le revenu annuel total en espèces ou en nature des ménages qu'elle qu'en soit la source. Il est extrêmement difficile de recueillir des informations fiables sur le revenu en particulier celui tiré d'une activité indépendante ou de biens meubles ou immeubles, dans le cadre d'une étude générale sur place, particulièrement d'un recensement de la population. On accroît encore les difficultés en incluant les revenus en nature dans la définition. La collecte de données relatives aux revenus, (même lorsqu'on se limite aux revenus en espèces), dans le cadre d'un recensement de population, soulève des problèmes spécifiques (charge de travail accrue, erreur de réponse, etc.). C'est pourquoi ce sujet est généralement considéré comme à étudier de préférence au moyen d'enquêtes par sondage. Les pays pourront cependant, en fonction de leurs besoins, chercher à obtenir par le recensement quelques informations limitées sur le revenu en espèces. Ainsi circonscrits les renseignements recueillis serviront non seulement certains objectifs immédiats du recensement, mais encore l'étude sur la répartition du revenu, de la consommation et de l'accumulation dans le secteur des ménages (Nations Unies 1998, par. 2.236).

504. La publication *Principes et recommandations* distingue deux types de revenus : le revenu individuel et le revenu des ménages. Ces deux rubriques doivent faire l'objet de vérifications similaires. Pour le revenu individuel et si la méthode d'imputation dynamique n'est pas utilisée, il convient d'attribuer la valeur « non indiqué » ou « in-

connu » en cas de réponse non valide. Sinon, la matrice d'imputation du revenu peut être établie à partir de variables telles que l'âge, le sexe, le niveau de formation, la branche d'activité, la profession et différentes caractéristiques personnelles.

505. Le revenu du ménage est égal à la somme des revenus obtenus par ses membres et doit figurer dans l'enregistrement relatif à l'unité d'habitation. La vérification de cette variable selon la méthode d'imputation dynamique utilise toutefois sensiblement les mêmes éléments que la vérification concernant le chef de ménage, (âge, sexe et niveau de formation), plutôt que les éléments se rapportant à chacune des personnes.

7. SECTEUR INSTITUTION (P6G)

506. Les questions sur le secteur institutionnel d'emploi visent le statut juridique et les principales fonctions, modes de comportement et objectifs de l'entreprise dont relève un emploi (Nations Unies, 1998, par. 2.239).

507. Il existe un lien entre certaines des branches d'activité et professions potentielles et le secteur industriel d'emploi (sociétés, administrations publiques, institutions sans but lucratif et autres). Certains pays peuvent choisir de vérifier l'existence de telles relations entre les variables pour assurer que les tableaux obtenus ne font pas apparaître d'incohérence lors de la tabulation croisée de ces variables.

508. Pour procéder aux opérations de vérification, les pays qui n'utilisent pas la méthode d'imputation dynamique devront attribuer la valeur « inconnu » à la variable secteur institutionnel, lorsque celle-ci n'est pas indiquée. Quant aux pays qui utilisent la méthode d'imputation dynamique, ils devront tenir compte de variables telles que l'âge, le sexe et éventuellement les grandes branches d'activité ou les professions de personnes de la même zone géographique présentant des caractéristiques similaires.

8. LIEU DE TRAVAIL (P6H)

509. Le « lieu de travail » est l'endroit où les personnes pourvues d'un emploi au sens de leur situation du moment, exercent leur activité et/ou celles pourvues d'un emploi au sens de leur situation actuelle, exercent l'activité principale qui a été utilisée pour déterminer d'autres caractéristiques économiques telles que la profession, l'activité économique et le statut. Si l'on peut utiliser les informations relatives au lieu de travail pour établir des profils de régions en fonction des facteurs de la main-d'œuvre (par opposition aux profils démographiques établis en fonction du lieu de résidence), le premier objectif demeure néanmoins de confronter les informations relatives au lieu de travail et celles qui concernent le lieu de résidence (Nations Unies, 1998, par. 2.245).

510. Etant donné que le « lieu de travail » sert à l'établissement de statistiques concernant les déplacements quotidiens domicile-travail, il importe que toute modification introduite dans les informations indiquées rende compte des zones géographiques particulières considérées. Aussi, les équipes de vérification de chaque pays peuvent-elles convenir d'attribuer le cas échéant la mention « inconnu », en pré-

sence de données non valides, et d'étudier uniquement les données « connues » effectivement recueillies.

511. La durée et la complexité des opérations de codage relatives à cette rubrique seront d'autant plus importantes si les mentions manuscrites sont admises et s'il faut les coder. Si les positions numériques du code sont hiérarchisées, par exemple si le premier chiffre représente la province, le second la circonscription administrative, etc. le codage sera vraisemblablement plus efficace et plus précis.

512. Les informaticiens devront vérifier que les matrices d'imputation sont établies uniquement pour des lieux

géographiques vraisemblables. Il est parfois indiqué d'effectuer une nouvelle imputation statique pour chaque circonscription administrative ou autre zone géographique afin de vérifier que les valeurs précédentes ne peuvent être choisies. Quant aux matrices d'imputation proprement dites, il est possible de tenir compte de l'âge, du sexe et sans doute des grandes catégories modifiées de professions ou de branches d'activité. Dans ce cas également, il faudra sans doute utiliser des matrices d'imputation différentes pour les lieux de travail situés sur le territoire national et à l'étranger.

V. — VÉRIFICATIONS CONCERNANT L'HABITATION

513. Les spécifications relatives aux vérifications des données d'habitation tiennent compte de la validité de chacune des informations recueillies comme de leur cohérence. La connaissance de liens particuliers entre les éléments d'information concernant un pays donné permet de concevoir des vérifications de cohérence propres à garantir une meilleure qualité des données à mettre en tableaux. Par exemple, une unité d'habitation ne doit pas avoir une toiture en ciment lorsque les murs sont en bambou. De manière analogue, les unités d'habitation doivent être équipées de canalisations d'eau à l'intérieur de la maison pour pouvoir déclarer une toilette munie de chasse d'eau, ou encore une baignoire ou une douche.

514. Comme avec les données sur la population, l'équipe de vérification doit, pour les données manquantes ou non valides, décider soit de leur attribuer la valeur « non indiqué », soit de procéder par imputation statique (*cold deck*) des valeurs inconnues ou autres ou encore par imputation dynamique (*hot deck*), d'après les caractéristiques d'autres unités d'habitation. Comme précédemment, l'imputation dynamique est souvent préférée puisqu'elle supprime le type d'imputation nécessaire au stade de la mise en tableau, alors que les décisions quant aux valeurs inconnues peuvent se faire uniquement en fonction des informations contenues dans les tableaux proprement dits. Les matrices d'imputation ainsi établies peuvent fournir des valeurs pour les données laissées en blanc ou non valides, ou pour la résolution des incohérences lorsqu'il n'existe aucune rubrique connexe comportant des réponses valides. Certains pays peuvent présenter une certaine variabilité des caractéristiques des habitations au niveau national, mais très peu au sein de la plupart des localités. D'autres pays peuvent présenter pour certaines rubriques particulières des variations très importantes d'une localité à l'autre, notamment entre les zones urbaines et les zones rurales. Il faut donc prendre en compte cette variabilité lors de l'établissement des matrices d'imputation, et tout particulièrement pour les valeurs initiales d'imputation statique. L'équipe de vérification peut décider de préciser les conditions dans lesquelles une valeur doit être attribuée à une rubrique laissée en blanc, d'après les informations concernant une unité d'habitation précédente, dont les caractéristiques sont voisines.

515. Sauf lorsqu'un pays ne dispose pas d'information sur l'habitation concernant des logements collectifs (en groupe), un (et un seul) enregistrement concernant l'habitation doit être attribué à chaque numéro de série (voir Vérifications structurelles, chapitre III). Différentes procédures d'assurance qualité sont présentées au chapitre consacré aux vérifications structurelles. Selon les décisions prises par l'équipe de vérification, le programme conçu à cet effet créera un enregistrement concernant l'habitation si celui-ci

est absent. De manière analogue, le programme de vérification peut supprimer un ou plusieurs enregistrements en présence d'enregistrements dédoublés ou multiples.

516. En principe, chaque enregistrement concernant l'habitation doit faire l'objet d'une vérification sélective limitée aux rubriques appropriées. Les données vérifiées peuvent être différentes selon qu'il s'agit d'un habitat urbain ou rural, et en fonction de plusieurs caractéristiques notamment climatiques. Toutefois, dans la pratique, rares sont les pays qui disposent du temps et des compétences nécessaires, pour concevoir et mettre en œuvre des tableaux de données multiples afin de remplacer les données manquantes ou incohérentes. Quant aux pays qui réalisent effectivement une vérification sélective, ils sont encore moins nombreux.

517. Les renseignements recueillis sur le questionnaire dépendront par ailleurs du type de locaux à usage d'habitation (unité d'habitation ou collectivités) et du caractère occupé ou inoccupé de l'unité considérée. Dans le cas des logements collectifs ou des collectivités, la vérification peut se limiter aux données propres aux collectivités ou porter conjointement sur les données recueillies aussi bien dans les collectivités que dans les autres unités d'habitation.

518. Par définition, il n'existe généralement pas d'enregistrement relatif aux habitations pour les personnes dites sans abri. Si des enregistrements de ce type existent, lorsque le pays décide de leur associer des identificateurs, ils peuvent soit être traités comme ceux relatifs aux locaux collectifs, soit faire l'objet d'une vérification complètement différente (et éventuellement d'aucune vérification).

519. Parfois, la présence d'une mention « non indiqué » est admissible pour une rubrique particulière. Il peut en être ainsi quand l'équipe de vérification d'un pays n'a pas de données de base adéquates permettant d'imputer des réponses pour une caractéristique donnée. La décision de conserver les mentions « non indiqué » doit être prise en respectant l'obligation de produire des tableaux de données appropriés, adaptés aux besoins de planification et d'action des pouvoirs publics. Lorsque les planificateurs ont besoin de certaines informations, dans la mesure où les données effectivement recueillies et les données portant la mention « non indiqué » ont des distributions identiques, l'attribution de valeurs déterminées aux données « non indiqué » ne devrait pas poser de problème. Par contre, lorsque ces mêmes données portant la mention « non indiqué » présentent un certain biais, l'imputation postérieure à la compilation des données risque de poser des problèmes, en particulier dans le cas des zones peu étendues ou dont les caractéristiques sont spéciales. Par exemple, si les personnes interrogées occupent une habitation de qualité médiocre, par rapport à la norme définie par le pays, et refusent d'indiquer certaines des caractéristiques de leur logement, tandis que l'agent de recensement s'abs-

tient de les rapporter, les planificateurs risquent de ne pas être en mesure d'adopter des programmes d'action corrective permettant de remédier à la qualité médiocre de l'habitat.

520. Les vérifications concernant l'habitation sont généralement plus simples par comparaison à celles qui concernent les données démographiques, compte tenu du caractère nettement moins complexe des tabulations croisées. La plupart des pays utilisent seulement différents niveaux d'agrégation géographique pour établir les caractéristiques individuelles en matière d'habitation. Tel qu'indiqué plus haut, les pays qui décident de ne pas utiliser la méthode d'imputation dynamique doivent définir un identificateur pour la mention « inconnu », à employer en présence de réponses non valides, ou incohérentes.

521. Dans les pays qui procèdent par imputation dynamique, l'équipe de vérification doit établir des matrices d'imputation simples dont les dimensions permettent de distinguer différentes caractéristiques de l'habitat. Dans la plupart des cas, la variable primaire la plus indiquée en vue de l'imputation dynamique est le « type de locaux d'habitation » couvrant aussi bien les unités d'habitation proprement dites que les locaux à usage collectif et notamment les types d'unités à l'intérieur de ces catégories.

522. Pour certains pays, la zone géographique peut constituer une dimension de ces matrices d'imputation. La variable « modalités de jouissance » peut également être utilisée à cet effet. Par exemple, si dans un pays la moitié des unités d'habitation sont occupées par des locataires et l'autre moitié par des propriétaires, la variable « modalités de jouissance » peut figurer parmi les dimensions de la matrice d'imputation. Cependant, si 5 % seulement des unités sont occupées en location, une autre caractéristique serait sans doute plus indiquée. La variable « modalités de jouissance » est intéressante à utiliser dans les matrices d'imputation, notamment dans les pays où les principaux régimes d'occupation représentent d'importants pourcentages du parc de logement. Parmi les autres caractéristiques à prendre en compte figurent le type de murs et la présence de l'électricité.

523. Pour chaque pays, les variables utilisées comme dimensions des matrices d'imputation doivent impérativement correspondre aux variables du jeu de données; ainsi, dans le cas des données sur l'habitation, il faut veiller à ce que les données individuelles ainsi que leurs combinaisons permettent de distinguer les caractéristiques recueillies. Dans la publication *Principes et recommandations* une distinction est faite entre données de base et données supplémentaires; les vérifications présentées ci-dessous à la section A mettent l'accent sur les données de base à inclure dans la matrice.

A. — Sujets de base

524. Dans le cadre des recensements de l'habitat, les unités de dénombrement sont les suivantes : a) les immeubles; b) les locaux à usage d'habitation; et c) les habitants des locaux à usage d'habitation. Les Nations Unies ont éta-

bli une liste de 20 sujets de base à vérifier qui présentent une valeur et un intérêt généraux et permettent en outre des comparaisons statistiques internationales détaillées. Pour la commodité des utilisateurs, les codes correspondants proposés ainsi qu'un certain nombre de sujets supplémentaires sont indiqués entre parenthèses ci-après. Les sujets sont regroupés par type d'unités de dénombrement.

1. IMMEUBLE : DESCRIPTION DE L'IMMEUBLE (H01)

525. La classification ci-après est recommandée par les Nations Unies (1998, par. 2. 299) lorsqu'il s'agit d'immeubles dont une partie au moins est utilisée comme habitation :

1. Immeubles constituant une seule unité d'habitation :
 - 1.1. Séparés;
 - 1.2. Non séparés;
2. Immeubles comportant plus d'une unité d'habitation :
 - 2.1. Jusqu'à 2 étages;
 - 2.2. De 3 à 10 étages;
 - 2.3. Onze étages et plus;
3. Immeubles destinés à des personnes vivant en institution.

526. En cas de non-validité de la valeur indiquée, il convient d'attribuer le code « inconnu » lorsqu'il n'est pas fait appel à la méthode d'imputation dynamique. Quant aux pays ayant opté pour la méthode d'imputation dynamique en présence de valeurs non valides, ils devraient utiliser au moins deux caractéristiques, parmi lesquelles pourraient figurer les matériaux de construction des murs extérieurs, la période de construction et/ou le type d'unités d'habitation que comporte l'immeuble (voir Nations Unies, 1998, par. 2.300) pour obtenir des informations « connues » d'après des unités d'habitation similaires de la même zone géographique.

2. MATÉRIAUX DE CONSTRUCTION DES MURS EXTÉRIEURS (H02)

527. Ce sujet vise les matériaux de construction des murs extérieurs de l'immeuble où se trouvent les locaux à usage d'habitation. Si les murs sont construits dans un ou plusieurs matériaux, c'est le matériau prédominant qui doit être indiqué. On recueillera des renseignements sur tel ou tel matériau (brique, béton, bois, adobe) suivant la fréquence de son utilisation dans le pays et son importance du point de vue de la durabilité de la construction. (Nations Unies, 1998, par. 2.304).

528. En cas de non-validité de la valeur indiquée, il convient d'attribuer le code « inconnu » lorsqu'il n'est pas fait appel à la méthode d'imputation dynamique. Quant aux pays ayant opté pour la méthode d'imputation dynamique en présence de valeurs non valides, ils devraient utiliser au moins deux caractéristiques, parmi lesquelles pourraient figurer la période de construction et/ou le type d'unités d'habitation que comporte l'immeuble pour obtenir des informations « connues » d'après des unités d'habitation similaires de la même zone géographique.

3. IMMEUBLES : ANNÉE OU PÉRIODE DE CONSTRUCTION (H03)

529. Cette caractéristique se rapporte à l'âge des immeubles où se trouvent des locaux à usage d'habitation. On recommande d'indiquer l'année exacte pour les bâtiments construits au cours de la période intercensitaire immédiatement précédente, si elle ne dépasse pas 10 ans. Lorsque la période intercensitaire dépasse 10 ans, ou lorsqu'il n'y a pas eu de recensement auparavant, l'année exacte doit être indiquée pour les bâtiments construits au cours des 10 dernières années. Pour les immeubles plus anciens, il convient d'indiquer des périodes permettant d'évaluer l'âge du parc de logement. Il peut s'avérer difficile de recueillir des données sur ce sujet, les occupants risquant dans certains cas de ne pas connaître la date de construction (Nations Unies, 1998, par. 2.307).

530. Certains pays, même ceux qui utilisent la méthode d'imputation dynamique, acceptent la mention « inconnu » en réponse à la question sur l'année ou la période de construction. Si tel est le cas, les pays en question peuvent décider pour cette rubrique de ne pas procéder par imputation dynamique, même en cas d'utilisation de matrices d'imputation pour d'autres variables. En pareille circonstance, lorsque la valeur indiquée est non valide, il convient sans doute d'attribuer néanmoins le code « inconnu » s'il n'y a pas imputation dynamique. Les pays qui font appel à l'imputation dynamique dans le cas de valeurs non valides devraient s'appuyer sur deux caractéristiques au moins, notamment le type d'immeuble, les matériaux de construction des murs extérieurs et/ou le type d'unités d'habitation de l'immeuble, pour obtenir des informations « connues » d'après des unités d'habitation similaires de la même zone géographique.

4. LOCAUX À USAGE D'HABITATION : EMPLACEMENT (H04)

531. L'emplacement des locaux à usage d'habitation est une variable géographique dont l'examen figure au chapitre III consacré aux vérifications structurelles.

5. LOCAUX À USAGE D'HABITATION : TYPE DE LOCAUX (H05)

532. La classification décrite ci-dessous présente un système de numéros de codes à 3 chiffres mis au point par les Nations Unies, (1998, par. 2.328-2.365) servant à grouper en deux grandes catégories les unités d'habitation et les locaux à usage d'habitation présentant des caractéristiques de construction analogues. La répartition des habitants (c'est-à-dire de la population) entre les différentes catégories fournit des indications utiles sur les logements disponibles au moment du recensement. Cette classification fournit également une base de stratification qui sera utile pour les enquêtes par sondage. Les locaux à usage d'habitation peuvent être classés selon les catégories ci-après :

1. Unités d'habitation :
 - 1.1. Logements classiques;
 - 1.2. Habitations de base;
 - 1.3. Unités d'habitation temporaires;
 - 1.4. Unités d'habitation mobiles;
 - 1.5. Unités d'habitation marginales :

1.5.1. Unités d'habitation improvisées;

1.5.2. Unités d'habitation se trouvant dans des bâtiments permanents non destinés à l'habitation;

1.5.3. Autres locaux non destinés à l'habitation;

2. Locaux d'habitation collectifs :

2.1. Hôtels, maisons meublées et autres locaux du même genre;

2.2. Institutions;

2.3. Camps;

2.4. Autres.

533. Les équipes de vérification doivent mettre au point des programmes propres à contrôler la cohérence interne des données concernant l'ensemble des locaux collectifs à usage d'habitation et des unités d'habitation. Si le type de locaux à usage d'habitation est « inconnu » ou non valide, les équipes de vérification décident parfois de mettre au point un contrôle qui examine les autres variables recueillies pour pouvoir attribuer une valeur au type de locaux d'habitation. Sinon, en cas de non-validité de la valeur indiquée, il convient d'attribuer le code « inconnu » en l'absence d'imputation dynamique. Les bureaux nationaux de statistique/recensement qui, par contre, font appel à cette méthode en présence de valeurs non valides, doivent s'appuyer sur deux caractéristiques au moins, telles que le type d'immeuble, les modalités de jouissance, le nombre de pièces, la surface de plancher ou le statut d'occupation pour obtenir des informations « connues » à partir d'unités d'habitation similaires de la même zone géographique.

6. LOCAUX À USAGE D'HABITATION : STATUT D'OCCUPATION (H06)

534. La question de savoir si les locaux à usage d'habitation que leurs occupants n'utilisent que périodiquement, doivent être considérés comme occupés ou non occupés, sera résolue différemment selon qu'on effectue parallèlement un recensement de la population de droit, ou un recensement de la population de fait. Dans les deux cas il semble utile de déterminer autant que possible si les locaux à usage d'habitation sont utilisés comme résidence secondaire. Cela est particulièrement important si la résidence secondaire a des caractéristiques très différentes de la résidence principale, comme c'est le cas par exemple quand les ménages d'agriculteurs quittent leur résidence permanente dans un village pour aller s'installer dans des locaux rudimentaires, pendant certaines saisons de l'année, sur les lieux des exploitations agricoles (Nations Unies, 1998, par. 2.369). La classification recommandée pour les logements classiques ou de base est la suivante :

1. Occupé;
2. Inoccupé :
 - 2.1. Inoccupé de façon saisonnière;
 - 2.2. Inoccupé pour des raisons autres que saisonnières :

- 2.2.1. A louer;
- 2.2.2. A vendre;
- 2.2.3. A démolir;
- 2.2.4. Autres.

535. Si l'unité d'habitation est occupée, le nombre d'occupants (code H17), ainsi que le décompte des enregistrements démographiques, doivent être différents de zéro. Si aucune personne n'est enregistrée, l'unité est inoccupée ou bien les personnes sont manquantes. Tel qu'indiqué plus haut dans le cadre des vérifications structurelles, les spécialistes doivent mettre au point des procédures permettant de déterminer si l'unité est effectivement inoccupée. Si elle est recensée en tant qu'unité occupée, mais se trouve inoccupée pour le moment, il faut définir une méthode permettant de déterminer le type d'inoccupation, soit en la recensant par la mention « inconnu », soit en procédant par imputation dynamique. Si l'unité est recensée comme étant inoccupée, mais s'il est possible d'établir qu'elle est occupée pour le moment sur la base des informations disponibles concernant le nombre d'occupants ou le décompte des enregistrements démographiques, il faut alors modifier la variable « statut d'occupation » et lui affecter la valeur « occupé ».

536. Si la valeur enregistrée n'est pas valide, le nombre d'occupants étant égal à zéro et en l'absence d'enregistrements démographiques, la mention « inconnu, inoccupé » doit être attribuée lorsqu'on ne procède pas par imputation dynamique. Si la valeur de la variable est non valide, avec un nombre d'occupants différent de zéro ou en présence d'enregistrements démographiques, la mention « occupé » doit être attribuée. Les pays ayant opté pour la méthode d'imputation dynamique des valeurs non valides, (pour imputer le type d'inoccupation) doivent s'appuyer sur au moins deux caractéristiques pour obtenir des informations « connues » d'après des unités d'habitation similaires de la même zone géographique; sinon, la mention « inconnu, inoccupé » peut être attribuée.

7. LOCAUX À USAGE D'HABITATION : TYPE DE PROPRIÉTÉ (H07)

537. Il s'agit du régime de propriété des locaux à usage d'habitation proprement dits et non celui du terrain sur lequel ils se trouvent (Nations Unies, 1998, par. 2.370). Le régime de propriété ne doit pas être confondu avec les modalités de jouissance. Les renseignements recueillis doivent permettre de déterminer si les locaux à usage d'habitation appartiennent au secteur public (état, administrations locales, organismes publics) ou s'ils sont propriété privée (appartenant à un ménage, une société privée, une coopérative, une société immobilière, etc.). La question est parfois détaillée afin de savoir si ces locaux sont entièrement payés, achetés à tempérament ou hypothéqués. La classification selon le régime de propriété est la suivante :

- 1. Occupé par le propriétaire;
- 2. Non occupé par le propriétaire :
 - 2.1. Propriété publique;
 - 2.2. Propriété privée;
 - 2.3. Autre.

538. Si le régime de propriété est lié aux modalités de jouissance, ce fait doit être pris en compte lors de la mise au point de la vérification; si ce n'est pas le cas, le type de propriété est vraisemblablement indépendant des autres variables concernant l'habitation. Si la valeur attribuée aux « types de propriété » est non valide, il convient de lui attribuer le code « inconnu », lorsqu'il n'y a pas imputation dynamique. Les pays qui décident d'imputer les valeurs non valides par la méthode d'imputation dynamique, doivent s'appuyer sur au moins deux caractéristiques parmi lesquelles pourraient figurer les matériaux de construction des murs, les modalités de jouissance, le type d'unités d'habitation et le nombre de pièces, afin d'obtenir des informations « connues » à partir d'unités d'habitation similaires de la même zone géographique.

8. LOCAUX À USAGE D'HABITATION : NOMBRE DE PIÈCES (H08)

539. On peut définir la pièce comme un espace d'une unité d'habitation ou autres locaux à usage d'habitation entouré de murs allant du plancher au plafond (ou éventuellement au toit), ou tout au moins jusqu'à une hauteur de 2 m au-dessus du sol, assez grand pour contenir un lit d'adulte (4 mètres carrés au moins). Sont compris dans cette définition les chambres à coucher, les salles à manger, les salles de séjour, les cabinets de travail, les greniers habitables, les chambres de domestiques, les cuisines, les pièces servant à l'exercice d'une profession ou d'un négoce et autres espaces, séparés, utilisés pour l'habitation ou destinés à l'habitation, à condition qu'ils répondent aux critères relatifs à la surface habitable et aux murs. Les couloirs, vérandas, vestibules, etc., les salles de bains et cabinets d'aisance ne doivent pas être comptés comme pièces même s'ils répondent à ces critères. A des fins nationales, on pourra recueillir des renseignements distincts concernant des locaux de moins de 4 mètres carrés répondant par ailleurs à la définition d'une pièce, si l'on considère que leur nombre le justifie (Nations Unies, 1998, par. 2.375).

540. Puisque le nombre de pièces peut être indépendant des autres variables concernant l'habitation, en cas de non-validité de la réponse indiquée, il convient de lui affecter le code « inconnu » lorsqu'il n'y a pas imputation dynamique. Par contre, les pays qui décident de procéder à l'imputation dynamique des valeurs non valides devraient s'appuyer sur au moins deux caractéristiques, telles que le type d'unité d'habitation, les matériaux de construction des murs, les modalités de jouissance et le statut d'occupation pour obtenir des informations « connues » concernant des unités d'habitation similaires de la même zone géographique.

9. LOCAUX À USAGE D'HABITATION : ESPACE PLANCHER (H09)

541. Ce sujet concerne l'espace plancher utile dans les unités d'habitation, c'est-à-dire l'espace plancher mesuré à l'intérieur des murs extérieurs, à l'exclusion des caves et des greniers non habitables. Dans les bâtiments à unités d'habitation multiples, les espaces communs sont à exclure.

La démarche sera différente selon qu'il s'agit d'unités d'habitation ou de locaux collectifs. (Nations Unies, 1998, par. 2.378).

542. L'espace plancher peut être lié au nombre de pièces et/ou au nombre de chambres à coucher; les équipes de vérification nationales peuvent donc décider d'en tenir compte lors de la mise au point des programmes de contrôle. Sinon, l'espace plancher est indépendant des autres vérifications concernant les données sur l'habitation. Il faut parfois spécifier l'unité de mesure, par exemple en mètres carrés. En cas de non-validité de la valeur indiquée, il convient d'attribuer le code « inconnu » lorsqu'il n'y a pas imputation dynamique. Par contre, les pays ayant opté pour l'imputation dynamique des valeurs non valides, il convient de s'appuyer sur au moins deux caractéristiques, notamment le type d'unité d'habitation, les matériaux de construction des murs, les modalités de jouissance et le statut d'occupation, afin de recueillir des données « connues » provenant d'unités d'habitation similaires de la même zone géographique.

10. LOCAUX À USAGE D'HABITATION :
SYSTÈME D'ALIMENTATION EN EAU (H10)⁶

543. D'après les Nations Unies, (1998, par. 2.381), aux fins du recensement l'essentiel est de savoir si les locaux à usage d'habitation ont l'eau sous conduite, c'est-à-dire s'ils sont alimentés en eau par des canalisations reliées soit à un système qui dessert toute la collectivité, soit à des installations privées, telles que réservoirs à pression, pompes, etc. L'unité de dénombrement pour ce sujet est l'unité d'habitation. Il faut indiquer s'il y a un poste d'eau à l'intérieur des locaux à usage d'habitation, ou à l'extérieur, mais à moins d'une certaine distance de la porte. La distance recommandée est de 200 mètres, selon l'hypothèse que l'accès à l'eau sous conduite en deçà de cette distance permet aux occupants de l'unité de logement de se procurer de l'eau pour les besoins du ménage sans déployer d'efforts excessifs. Outre l'emplacement du poste d'eau, la source d'approvisionnement présente aussi un intérêt particulier. La classification recommandée des unités d'habitation en fonction du système d'alimentation en eau est donc la suivante :

1. Poste d'eau sous conduite à l'intérieur de l'unité :
 - 1.1. Provenant du réseau d'adduction collectif;
 - 1.2. Provenant d'une source d'approvisionnement privée;
2. Poste d'eau sous conduite à l'extérieur de l'unité mais à moins de 200 mètres :
 - 2.1. Provenant du réseau d'adduction collectif :
 - 2.1.1. A usage collectif;
 - 2.1.2. A usage commun;
 - 2.2. Provenant d'une source d'approvisionnement privée :

⁶ Pour les variables suivantes l'unité de dénombrement est effectivement le système d'alimentation en eau de l'unité d'habitation : lieux d'aisances et égouts, salles de bains, installations destinées à la préparation des repas, mode d'éclairage et évacuation des déchets solides.

2.2.1. A usage exclusif;

2.2.2. A usage commun;

3. Pas de poste d'eau sous conduite disponible (même à plus de 200 mètres des locaux à usage d'habitation).

544. Par réseau d'adduction collectif on entend un réseau d'adduction soumis à inspection et contrôle des pouvoirs publics. Ces réseaux sont généralement gérés par des organismes publics, mais il arrive qu'ils soient gérés en coopérative ou par des entreprises privées.

545. Il convient vraisemblablement de procéder conjointement à la vérification des rubriques concernant les canalisations d'eau (alimentation en eau, lieux d'aisances et égouts, salles d'eau et approvisionnements en eau chaude). Puisque les données de ce type sont étroitement liées, en cas d'absence ou de non-validité de l'une d'entre elles, les autres données peuvent servir à l'obtention d'une valeur appropriée. Dans les régions sans eau courante, les spécialistes doivent parfois appliquer des vérifications particulières aux unités d'habitation considérées. Sinon, d'autres unités d'habitation de la région étudiée auront probablement des caractéristiques similaires et l'utilisation des rubriques correspondantes est recommandée lorsqu'on procède par imputation dynamique.

546. Si le type indiqué de système d'alimentation en eau est non valide, il convient de lui attribuer le code « inconnu », lorsque ne procède pas par imputation dynamique. En revanche, les pays ayant opté pour cette méthode doivent s'appuyer sur deux caractéristiques au moins. Parmi celles-ci peuvent figurer normalement le type d'unité d'habitation, puis celui de lieu d'aisances, d'égouts et enfin de salles d'eau, de manière à obtenir des informations « connues » sur la base d'unités d'habitation similaires de la même zone géographique.

11. LOCAUX À USAGE D'HABITATION :
LIEUX D'AISANCES ET ÉGOUTS (H11)

547. Dans certains pays on a jugé utile de subdiviser la catégorie des cabinets d'aisances sans chasse d'eau, pour distinguer certains types d'usage courant qui correspondent à un certain niveau d'hygiène. La classification recommandée (Nations Unies, 1998, par. 2.386) pour répertorier les unités d'habitation en fonction des lieux d'aisances est la suivante :

1. Cabinet d'aisances à l'intérieur de l'unité de logement :
 - 1.1. Avec chasse d'eau;
 - 1.2. Sans chasse d'eau;
2. Cabinets d'aisances à l'extérieur :
 - 2.1. Avec chasse d'eau :
 - 2.1.1. A usage exclusif;
 - 2.1.2. A usage partagé;
 - 2.2. Sans chasse d'eau :
 - 2.2.1. A usage exclusif;
 - 2.2.2. A usage partagé;
3. Pas de cabinets d'aisances disponibles.

548. Le type de lieux d'aisances et d'égouts constitue une autre rubrique au titre de l'habitation ayant un rapport avec l'approvisionnement en eau et devrait donc être vérifié

conjointement avec les autres rubriques connexes. Le choix de mentions, telles que « privé », « usage commun », « usage exclusif » etc., pourrait servir à déterminer la cohérence des différentes réponses et si tel n'est pas le cas, les chemins de vérification à suivre pour résoudre le problème. En présence d'une ou plusieurs variables concernant l'eau, il est possible d'établir une estimation concernant les données inconnues ou incohérentes sans faire appel à l'utilisation du code « inconnu » ou à la méthode d'imputation dynamique. Toutefois, faute d'obtenir ainsi une valeur valide, le code « inconnu » doit être attribué lorsqu'il n'y a pas imputation dynamique. Par contre, les pays ayant opté pour l'imputation dynamique des valeurs non valides, doivent s'appuyer sur au moins deux caractéristiques, notamment le type d'unité d'habitation, en règle générale, l'alimentation en eau, les matériaux de construction des murs, les modalités de jouissance et le statut d'occupation, afin de recueillir des données « connues » provenant d'unités d'habitation similaires de la même zone géographique.

12. LOCAUX À USAGE D'HABITATION : SALLES D'EAU (H12)

549. D'après les Nations Unies, (1998, par. 2.390), il convient de déterminer pour chaque local à usage d'habitation, s'il y a une baignoire ou une douche fixe. Pour cette rubrique, l'unité de dénombrement est également l'unité d'habitation. On peut établir en outre si les installations sont à usage exclusif des occupants du local ou non et si la salle d'eau est alimentée en eau chaude pour le bain ou seulement en eau froide. Il est néanmoins possible, que dans certaines régions du monde, la distinction proposée ci-dessus ne corresponde pas le mieux aux besoins du pays. Ainsi, il peut être important de faire la distinction entre une salle d'eau séparée dans les locaux à usage d'habitation, une salle d'eau séparée dans le bâtiment et un établissement de bain public. La classification recommandée des unités d'habitation en fonction des salles d'eau dont elles sont dotées ou non est la suivante :

1. Baignoire ou douche fixe intérieures;
2. Pas de baignoire ou de douche fixe intérieures :
 - 2.1. Baignoire ou douche fixe disponible à l'extérieur de l'unité d'habitation :
 - 2.1.1. A usage exclusif;
 - 2.1.2. A usage commun;
 - 2.2. Ni baignoire, ni douche fixe disponibles.

550. Le type de salles de bain devrait faire l'objet d'une vérification conjointe avec les autres rubriques connexes. Le choix de mentions, telles que « privé », « usage commun », « usage exclusif » etc., pourrait servir à déterminer la cohérence des différentes réponses et si tel n'est pas le cas, les chemins de vérification à suivre pour résoudre le problème. En présence d'une ou plusieurs variables concernant l'eau, il est possible d'établir une estimation concernant les données inconnues ou incohérentes sans faire appel à l'utilisation du code « inconnu » ou à la méthode d'imputation dynamique. Toutefois, si toutes les tentatives échouent et si la valeur est non valide, le code « inconnu » doit être attribué lorsqu'on ne procède pas à une imputation dynamique. Par contre, les pays ayant opté pour l'imputation dynamique des valeurs

non valides, doivent s'appuyer sur au moins deux caractéristiques, notamment le type d'unité d'habitation, en règle générale, l'alimentation en eau, les matériaux de construction des murs, les modalités de jouissance et le statut d'occupation, afin de recueillir des données « connues » provenant d'unités d'habitation similaires de la même zone géographique.

13. LOCAUX À USAGE D'HABITATION :

INSTALLATIONS DESTINÉES À LA PRÉPARATION DES REPAS (H13)

551. D'après la publication intitulée *Principes et recommandations* (Nations Unies, 1998, par. 2.395) la collecte des données sur l'existence d'une cuisine peut être l'occasion de recueillir des renseignements sur le genre d'équipements utilisés pour la préparation des repas (fourneaux, plaques chauffantes, feu nu, etc.), ainsi que sur l'existence d'un évier et d'un endroit prévu pour stocker la nourriture en évitant qu'elle se gâte. La classification recommandée des unités d'habitation en fonction de l'existence d'une cuisine ou de tout autre espace réservé pour la préparation des repas est la suivante :

1. Cuisine intérieure;
2. Autre espace intérieur pour la préparation des repas;
3. Ni cuisine, ni espace réservé à la préparation des repas à l'intérieur de l'unité d'habitation :
 - 3.1. Cuisine ou autre espace pour la préparation des repas à l'extérieur de l'unité d'habitation :
 - 3.1.1. A usage exclusif;
 - 3.1.2. A usage commun;
 - 3.2. Ni cuisine, ni espace pour la préparation des repas.

552. Pour la vérification des données recueillies sur les installations destinées à la préparation des repas, des mentions, telles que « privé », « usage commun », « usage exclusif » etc., pourraient servir à déterminer la cohérence des différentes réponses et, si tel n'est pas le cas, les chemins de vérification à suivre pour résoudre le problème. En présence d'une ou plusieurs variables concernant les installations de préparation des repas, il est possible d'établir une estimation concernant les données inconnues ou incohérentes sans faire appel à l'utilisation du code « inconnu » ou à la méthode d'imputation dynamique. Toutefois, faute d'obtenir ainsi une valeur valide, le code « inconnu » doit être attribué lorsqu'on ne procède pas par imputation dynamique. Par contre, les pays ayant opté pour l'imputation dynamique des valeurs non valides, doivent s'appuyer sur au moins deux caractéristiques, notamment, en règle générale, le type d'unité d'habitation, puis l'alimentation en eau, les matériaux de construction des murs, les modalités de jouissance et le statut d'occupation, afin de recueillir des données « connues » provenant d'unités d'habitation similaires de la même zone géographique.

14. LOCAUX À USAGE D'HABITATION :

MODE D'ÉCLAIRAGE (H14)

553. Il convient de recueillir des renseignements sur le mode d'éclairage des locaux utilisés dans les locaux à usage d'habitation (électricité, gaz, lampes à pétrole, etc.). Si les

locaux sont éclairés à l'électricité, certains pays devront préciser si le courant provient du secteur, de l'utilisation d'une génératrice, ou d'une autre source (usine, mine, etc.). Outre le type d'électricité utilisé, les pays pourront voir si l'électricité est disponible à d'autres fins que l'éclairage (notamment préparation des repas, eau chaude, chauffage, etc.). Si la situation du logement dans le pays fait qu'il est possible de déduire ces renseignements du type d'éclairage utilisé, cela évite une enquête complémentaire (Nations Unies, 1998, par. 2.398).

554. Si la valeur enregistrée n'est pas valide, le code « inconnu » doit être attribué lorsqu'on ne procède pas par imputation dynamique. Par contre, les pays ayant opté pour l'imputation dynamique des valeurs non valides, doivent s'appuyer sur au moins deux caractéristiques, notamment, en règle générale, le type d'unité d'habitation, les matériaux de construction des murs, les modalités de jouissance et le statut d'occupation, afin de recueillir des données « connues » provenant d'unités d'habitation similaires de la même zone géographique.

15. LOCAUX À USAGE D'HABITATION :
TYPE D'ÉVACUATION DES DÉCHETS SOLIDES (H15)

555. D'après la publication *Principes et recommandations* (Nations Unies, 1998, par. 2.401) ce sujet vise la collecte et l'évacuation des déchets solides produits par les occupants d'une unité d'habitation. L'unité de dénombrement est l'unité d'habitation. La classification préconisée des unités de logement en fonction du système d'évacuation des déchets solides est la suivante :

1. Déchets solides collectés périodiquement par des agents agréés;
2. Déchets solides collectés de façon non périodique par des agents agréés;
3. Déchets solides collectés par des agents autonomes;
4. Déchets solides déposés par les occupants dans une décharge locale contrôlée par les autorités;
5. Déchets solides déposés par les occupants dans une décharge locale non contrôlée par les autorités;
6. Autres arrangements, (y compris incinération des déchets solides par les occupants).

556. Le type d'évacuation des déchets solides est indépendant des autres variables concernant l'habitation. Si la valeur enregistrée n'est pas valide, le code « inconnu » doit être attribué lorsqu'on ne procède pas par imputation dynamique. Par contre, les pays ayant opté pour l'imputation dynamique des valeurs non valides, doivent s'appuyer sur au moins deux caractéristiques, notamment, en règle générale, le type d'unité d'habitation, puis les matériaux de construction des murs, les modalités de jouissance et le statut d'occupation ou les installations destinées à la préparation des repas, afin de recueillir des données « connues » provenant d'unités d'habitation similaires de la même zone géographique.

16. LOCAUX À USAGE D'HABITATION :
OCCUPATION PAR UN MÉNAGE OU PLUS (H16)

557. La rubrique occupation par un ménage ou plus est indépendante des autres rubriques concernant l'habitation. Si la valeur attribuée à cette variable n'est pas valide, il convient alors de compter les chefs de ménage et d'utiliser le nombre ainsi obtenu.

17. LOCAUX À USAGE D'HABITATION :
NOMBRE D'OCCUPANTS (H17)

558. Chaque personne qui réside habituellement dans une unité d'habitation ou autres locaux collectifs à usage d'habitation doit être comptée comme un occupant. L'unité de dénombrement pour ce sujet est donc le local d'habitation. Cependant, les recensements de l'habitation étant généralement effectués simultanément avec ceux de la population, cette définition sera plus ou moins applicable suivant que les renseignements recueillis et enregistrés pour chaque personne dans le recensement de la population indiquent où se trouvait cette personne le jour du recensement, ou se réfèrent à son lieu de résidence habituel. Il faut veiller à distinguer, parmi les occupants des unités d'habitation mobiles (bateau, roulotte, caravane, etc.), les personnes qui résident habituellement dans ces unités de celles qui les utilisent en tant que moyen de transport (Nations Unies, 1998, par. 2.407).

559. Le « nombre d'occupants » est lié au nombre d'enregistrements démographiques et ces deux nombres doivent être égaux. S'ils sont différents, il faut alors corriger soit le nombre d'occupants, soit le nombre d'enregistrements concernant la population. D'ordinaire, le nombre d'occupants est ajusté de façon à être égal au nombre de personnes enregistrées dans l'unité d'habitation. Cette rubrique ne doit pas être mise à la valeur « inconnu » et ne doit pas non plus faire l'objet d'une imputation.

18. OCCUPANTS :
CARACTÉRISTIQUES DU CHEF DE MÉNAGE (H18)

560. Les caractéristiques du chef de ménage sont généralement tirées des enregistrements concernant la population, afin de faciliter l'établissement des informations qui seront mises en tableaux croisés à des fins de planification et d'analyse. Ces rubriques, notamment l'origine ethnique, la religion ou le revenu, facilitent la détermination des différences de statut social ou de besoins sociaux. Les caractéristiques en question ayant déjà fait l'objet d'une vérification concernant les rubriques démographiques, elles ne doivent pas faire l'objet d'une vérification supplémentaire, dans le présent contexte.

19. OCCUPANTS : MODALITÉS DE JOUISSANCE (H19)

561. D'après les Nations Unies (1998, par. 2.410), les modalités de jouissance indiquent à quel titre un ménage occupe tout ou partie d'une unité d'habitation. L'unité de dénombrement est le ménage occupant une unité d'habitation. La classification des ménages en fonction des modalités de jouissance est la suivante :

1. Membre du ménage propriétaire de l'unité d'habitation;
2. Membre du ménage louant tout ou partie de l'unité d'habitation :
 - 2.1. Membre du ménage louant tout ou partie de l'unité d'habitation en tant que locataire en titre;
 - 2.2. Membre du ménage louant tout ou partie de l'unité d'habitation en tant que sous-locataire;
3. Autre arrangement.

Il convient de considérer séparément le cas des locaux occupés sans paiement de loyers en espèces ou sans autorisation du propriétaire, particulièrement lorsque cette situation est relativement fréquente.

562. La rubrique modalités de jouissance peut avoir un lien avec le type de propriété (H12); aussi l'équipe de vérification doit-elle parfois envisager le lien entre ces deux rubriques lors de l'élaboration des programmes de contrôle. Sinon, en cas de non-validité de la valeur indiquée pour la variable modalités de jouissance, il convient de lui attribuer le code « inconnu » lorsqu'on ne procède pas par imputation dynamique. Les pays qui optent pour l'imputation dynamique des valeurs non valides doivent s'appuyer sur deux caractéristiques au moins, telles que le type d'unité d'habitation, le loyer et le statut d'occupation pour obtenir des informations « connues », d'après des unités d'habitation analogues de la même zone géographique.

20. OCCUPANTS : LOYER ET COÛT DE LOGEMENT DES PROPRIÉTAIRES OCCUPANTS (H20)

563. La rubrique relative aux loyers et aux coûts de logement des propriétaires occupants est indépendante des autres variables concernant l'habitation, si ce n'est évidemment le fait que les loyers se rapportent uniquement aux unités occupées par des locataires et les coûts de logement exclusivement aux unités occupées par des propriétaires. L'équipe de vérification doit examiner chaque cas et déterminer les liens les plus appropriés entre ces variables. Si la valeur indiquée pour cette rubrique est non valide, il convient de lui attribuer le code « inconnu » lorsqu'on ne procède pas par imputation dynamique. Les pays qui optent pour l'imputation dynamique des valeurs non valides doivent s'appuyer sur deux caractéristiques au moins pour obtenir des informations « connues », d'après les unités d'habitation analogues de la même zone géographique.

B. — Sujets supplémentaires

564. La présente section définit un certain nombre de sujets supplémentaires qui pourront aider nombre de pays à organiser leurs recensements nationaux et leurs enquêtes. Là encore, ils sont classés par type d'unités de dénombrement (immeubles, locaux à usage d'habitation et occupants), les codes de vérification proposés étant indiqués entre parenthèses (Nations Unies, 1998, par. 2.416-432).

1. IMMEUBLES : NOMBRE DE LOGEMENTS (A01)

565. La description des vérifications à effectuer concernant le nombre d'unités d'habitation dans un immeuble fi-

gure aux paragraphes 219-220 du chapitre III, dans le cadre des vérifications structurelles.

2. IMMEUBLE : ASCENSEUR (A02)

566. Ce sujet vise l'existence d'un ascenseur (plateforme qui monte ou descend pour transporter des personnes et du fret) dans un immeuble à plusieurs étages. Les données se rapportent à un ascenseur utilisable la plupart du temps, c'est-à-dire opérationnel la plupart du temps, sous réserve d'un entretien régulier (Nations Unies, 1998, par. 2.419).

567. Si l'immeuble n'a qu'un étage ou s'il s'agit d'une unité d'habitation isolée et indépendante, il ne doit pas y avoir d'ascenseur. S'il est fait état d'un ascenseur, l'équipe de vérification doit décider de l'information à considérer en priorité, soit le nombre d'étages, soit l'existence indiquée d'un ascenseur. Si l'existence de l'ascenseur est retenue en priorité, il faut alors modifier le nombre d'étages, soit en attribuant à cette rubrique le code « inconnu », soit en obtenant une autre valeur en procédant par imputation dynamique. Si le nombre d'étages est considéré en priorité, et si l'immeuble ne comporte qu'un seul niveau, la réponse à la question « existence d'un ascenseur » doit être « non ».

568. Si l'existence d'un ascenseur est indiquée et s'il fonctionne à l'électricité, il convient de faire une vérification pour être sûr que l'immeuble est bien relié au secteur.

569. Enfin, si la valeur indiquée pour la rubrique « existence d'un ascenseur » n'est pas valide, il convient de lui attribuer le code « inconnu » lorsqu'on ne procède pas par imputation dynamique. Les pays qui conviennent de procéder à l'imputation dynamique des valeurs non valides devraient s'appuyer sur deux caractéristiques au moins, telles que type d'immeuble et matériaux de construction des murs extérieurs pour obtenir des informations « connues », d'après les unités d'habitation analogues de la même zone géographique.

3. BÂTIMENTS : BÂTIMENT À USAGE AGRICOLE (H03)

570. Dans plusieurs recensements nationaux, il a été jugé utile d'opérer une distinction entre les immeubles à usage agricole ou les autres. Par immeuble à usage agricole, on entend un immeuble qui fait partie d'une exploitation agricole et est utilisé à des fins agricoles et/ou résidentielles. (Nations Unies, 1998, par. 2.420).

571. La rubrique « immeuble à usage agricole » est indépendante des autres rubriques concernant l'habitation. Les pays peuvent par ailleurs vérifier les liens avec les données sur la population concernant la profession et la branche d'activité. Par ailleurs, si la valeur indiquée n'est pas valide, il convient de lui attribuer le code « inconnu » lorsqu'on ne procède pas par imputation dynamique. Les pays qui conviennent de procéder à l'imputation dynamique des valeurs non valides devraient s'appuyer sur deux caractéristiques au moins, telles que type d'immeuble et matériaux de construction des murs extérieurs pour obtenir des informations « connues », d'après les unités d'habitation analogues de la même zone géographique.

4. MATÉRIAUX UTILISÉS

POUR LA CONSTRUCTION DE LA TOITURE (A04a)

572. Dans certains cas, les matériaux utilisés pour la construction des toitures et des planchers peuvent présenter un intérêt spécial et servir à évaluer plus en détail la qualité des logements dans l'immeuble. Ce sujet vise les matériaux utilisés pour la toiture et/ou les planchers (toutefois, en fonction des besoins particuliers d'un pays, il peut également viser d'autres parties comme la charpente ou les fondations). L'unité de dénombrement est l'immeuble. Seul le matériau prédominant est à prendre en compte et dans le cas de la toiture, il peut s'agir de tuiles, de béton, de tôle, de palmes, de chaume, de bambous ou d'autres matériaux végétaux similaires, de terre, de feuilles de plastique, etc. (Nations Unies, 1998, par. 2.421).

573. Parfois la réponse fournie quant aux matériaux de construction des murs extérieurs ne concorde pas avec les renseignements indiqués quant aux matériaux de construction de la toiture; tel peut être le cas par exemple, si les matériaux de construction indiqués pour les murs ne sont pas suffisamment résistants pour soutenir la toiture. Tel qu'indiqué plus haut, lorsqu'il en est ainsi, les spécialistes doivent choisir soit de modifier la valeur de l'une des deux variables, soit d'attribuer le code « inconnu ». Par ailleurs, si la valeur indiquée n'est pas valide, il convient de lui attribuer le code « inconnu » lorsqu'on ne procède pas par imputation dynamique. Les pays qui conviennent de procéder à l'imputation dynamique des valeurs non valides devraient s'appuyer sur deux caractéristiques au moins, telles que type d'immeuble et matériaux de construction des murs extérieurs, type d'unité d'habitation, matériaux de construction des murs, modalités de jouissance et état d'occupation, pour obtenir des informations « connues », d'après les unités d'habitation analogues de la même zone géographique.

5. IMMEUBLE : MATÉRIAUX DE CONSTRUCTION DES PLANCHERS (A04b)

574. Les matériaux de construction indiqués pour les planchers ne sont pas nécessairement compatibles avec la construction de la toiture et des murs. Lorsque l'équipe de vérification trouve des combinaisons incohérentes ou non valides, elle doit alors décider de modifier une ou plusieurs réponses en attribuant le code « inconnu » ou en faisant appel à des matrices d'imputation. Si la valeur indiquée n'est pas valide, il convient de lui attribuer le code « inconnu » lorsqu'on ne procède pas par imputation dynamique. Les pays qui conviennent de procéder à l'imputation dynamique des valeurs non valides devraient s'appuyer sur deux caractéristiques au moins, telles que type d'immeuble, matériaux de construction des murs extérieurs, type d'unité d'habitation, modalités de jouissance et état d'occupation, pour obtenir des informations « connues », d'après les unités d'habitation analogues de la même zone géographique.

6. IMMEUBLE : ÉTAT D'ENTRETIEN (A05)

575. Ce sujet précise si l'immeuble a besoin de réparations et indique la nature des réparations nécessaires. L'unité

de dénombrement est l'immeuble. La classification des immeubles selon l'état d'entretien peut inclure les catégories suivantes : « aucune réparation nécessaire », « petites réparations nécessaires », « réparations moyennes nécessaires » ou encore « grosses réparations nécessaires » et « état irréparable ». Par « petites réparations » on entend principalement l'entretien ordinaire de l'immeuble et de ses composants (par exemple, vitres cassés). Les réparations moyennes correspondent à la rectification de défauts d'importance moyenne comme l'absence de gouttières sur le toit, plâtres endommagés sur une surface importante, escaliers sans rampes sûres. Les grosses réparations sont nécessitées par un vice de construction de l'immeuble, tel que tuiles ou bardeaux manquants au toit, fissures et trous dans les murs extérieurs, marches manquantes. L'état « irréparable » qualifie des immeubles qui ne peuvent plus être réparés, les vices de construction étant tellement graves qu'on juge préférable de détruire l'immeuble plutôt que d'entreprendre les réparations. Il s'agit le plus souvent d'immeubles qui n'ont plus que la charpente, sans murs extérieurs, ni toit (Nations Unies, 1998, par. 2.422).

576. L'état d'entretien de l'immeuble est indépendant des autres variables concernant l'habitation. Par conséquent, si la valeur indiquée n'est pas valide, il convient de lui attribuer le code « inconnu » lorsqu'on ne procède pas par imputation dynamique. Les pays qui conviennent de procéder à l'imputation dynamique des valeurs non valides devraient s'appuyer sur deux caractéristiques au moins, telles que type d'immeuble, matériaux de construction des murs extérieurs, et type d'unité d'habitation, pour obtenir des informations « connues », d'après les unités d'habitation analogues de la même zone géographique.

7. LOCAUX À USAGE D'HABITATION : NOMBRE DE CHAMBRES (A06)

577. En plus du nombre de pièces, plusieurs pays recensent le nombre de chambres de l'unité d'habitation qui est l'unité de dénombrement pour ce sujet. Par chambre, on entend une pièce équipée d'un lit et utilisée pour dormir (Nations Unies, 1998, par. 2.423).

578. Parfois, les agents de recensement indiquent pour le nombre de chambres une valeur supérieure au nombre de pièces⁷. Si tel est le cas et si la mention « non indiqué » est réservée dans le pays considéré aux réponses non valides ou « incohérentes, elle doit alors remplacer le nombre de chambres. En cas d'imputation dynamique, le nombre de chambres doit être « estimé » au moyen d'une matrice d'imputation dont le nombre de pièces constitue l'un des éléments. De cette façon, le nombre de chambre ne dépassera pas le nombre de pièces, puisque la valeur mentionnée pour le nombre de chambres sera mise à jour uniquement en cas de

⁷ Si des indications figurent conjointement au sujet du nombre de pièces et du nombre de chambres, il convient de les vérifier simultanément, étant entendu que le nombre de chambres ne doit pas dépasser le nombre de pièces. Puisque le nombre de chambres figure au titre de sujet supplémentaire, la vérification est effectuée seulement lorsque ces deux informations sont présentes.

concordance du nombre de chambres et du nombre de pièces. Le cas le plus simple serait celui d'un tableau de données à une seule rangée, le numéro de chaque cas correspondant au nombre de pièces et le contenu au nombre de chambres. Dans une matrice d'imputation plus complexe, pourrait figurer le nombre de personnes de l'unité d'habitation, ainsi que le type de structure.

579. Sinon, en cas de non-validité de la valeur indiquée pour le nombre de chambres, il convient de lui attribuer le code « inconnu » lorsqu'on ne procède pas par imputation dynamique. Les pays qui conviennent de procéder à l'imputation dynamique des valeurs non valides devraient s'appuyer sur deux caractéristiques au moins (dont une serait le nombre de chambres), pour obtenir des informations « connues », d'après les unités d'habitation analogues de la même zone géographique.

8. LOCAUX À USAGE D'HABITATION : TYPE DE COMBUSTIBLE UTILISÉ POUR LA PRÉPARATION DES REPAS (A07)

580. Dans le souci de contrôler de près l'utilisation des ressources naturelles, plusieurs pays incluent dans le recensement de l'habitation le type de combustible utilisé pour la préparation des repas. L'unité de dénombrement est l'unité d'habitation; le type de « combustible utilisé pour la préparation des repas » désigne le combustible le plus utilisé pour la préparation des principaux repas. Si deux combustibles sont utilisés (par exemple électricité et gaz), on citera celui utilisé le plus souvent. La classification du combustible utilisé pour la préparation des repas dépend du contexte national et peut comporter l'électricité, le gaz, le pétrole, le charbon, le bois et les déchets animaux. Il peut également s'avérer utile d'obtenir cette information pour les locaux d'habitation collective, en particulier si le nombre en est important dans le pays considéré (Nations Unies, 1998, par. 2.424).

581. La réponse indiquée quant au type de combustible utilisé pour la préparation des repas doit être vérifiée en même temps que celle indiquée pour les installations de préparation des repas. L'équipe de vérification détermine le lien entre les deux variables en question et met au point un contrôle de cohérence des réponses correspondantes obtenues. L'attribution de valeurs telles que « privé », « à usage partagé », « à usage exclusif » permettra sans doute de déterminer la cohérence des réponses fournies et sinon, de choisir les chemins de vérification permettant de résoudre le problème. En présence de données concernant l'une des deux ou les deux variables relatives à la préparation des repas, il est possible d'établir une estimation des données inconnues ou incohérentes, sans faire appel à la méthode d'imputation dynamique ou à l'attribution du code « inconnu ». Toutefois, si la valeur indiquée n'est pas valide, il convient de lui attribuer le code « inconnu » lorsqu'on ne procède pas par imputation dynamique. Les pays qui décident de procéder à l'imputation dynamique des valeurs non valides doivent s'appuyer sur deux caractéristiques au moins, telles que installations de préparation des repas, le type d'immeuble, matériaux de construction des murs, modalités de jouissance et état d'occupation, pour obtenir des informations

« connues », d'après les unités d'habitation analogues de la même zone géographique.

9. LOCAUX À USAGE D'HABITATION : TYPE DE CHAUFFAGE ET D'ÉNERGIE UTILISÉS À CETTE FIN (A08)

582. Ce sujet fait référence au type de chauffage des locaux à usage d'habitation et à l'énergie utilisée à cette fin. Les unités de dénombrement sont l'ensemble des locaux à usage d'habitation. Ce sujet ne s'applique pas pour plusieurs pays où, du fait de leur position géographique et de leur climat, les locaux à usage d'habitation n'ont pas à être chauffés. Le type de chauffage fait référence à la nature du système utilisé pour chauffer la plus grande partie des locaux; il peut s'agir de chauffage central desservant un ou plusieurs locaux à usage d'habitation, ou il peut ne pas être central, les locaux étant chauffés par divers poêles, cheminées, ou autres dispositifs. Quant à l'« énergie utilisée pour le chauffage », elle est fonction du type de chauffage; il s'agit de la source prédominante d'énergie, combustibles solides (charbon, lignite et produit du charbon et de la lignite, bois), mazout, combustible gazeux (gaz naturel ou liquéfié), électricité (Nations Unies, 1998, par. 2.425).

583. Les types de chauffage et d'énergie utilisée à cet effet sont liés entre eux ainsi qu'à l'alimentation en eau chaude et à la présence de différentes commodités dans l'unité d'habitation telles que l'électricité et le gaz sous conduite. Les équipes de vérification doivent prendre en compte la présence de ces commodités pour mettre au point les spécifications de contrôle du type de chauffage et d'énergie utilisée à cet effet. Le type de chauffage peut être indépendant des autres éléments concernant l'habitation et peut alors faire l'objet d'une vérification séparée. Toutefois, lorsque l'indication recueillie concernant « l'énergie utilisée pour le chauffage » est inconnue ou incohérente, le programme peut alors vérifier le type d'énergie utilisée pour l'éclairage. Enfin, si la valeur indiquée n'est pas valide, il convient de lui attribuer le code « inconnu » lorsqu'on ne procède pas par imputation dynamique. Les pays qui conviennent de procéder à l'imputation dynamique des valeurs non valides devraient s'appuyer sur deux caractéristiques au moins, telles que type d'unité d'habitation, matériaux de construction des murs, modalités de jouissance et état d'occupation, pour obtenir des informations « connues », d'après les unités d'habitation analogues de la même zone géographique.

10. LOCAUX À USAGE D'HABITATION : ALIMENTATION EN EAU CHAUDE (A09)

584. Il s'agit de savoir si les locaux à usage d'habitation sont ou non alimentés en eau chaude, c'est-à-dire en eau chauffée à une certaine température et fournie aux occupants par des conduites et des robinets. On cherchera à savoir s'il y a de l'eau chaude à l'intérieur des locaux à usage d'habitation, en dehors des locaux à usage d'habitation pour l'usage exclusif ou commun, ou pas d'eau chaude (Nations Unies, 1998, par. 2.426).

585. La rubrique alimentation en eau chaude peut être liée aux informations concernant les moyens de chauffage

de l'eau, bien que l'utilisation à cette fin de l'énergie solaire ne soit pas nécessairement liée aux autres rubriques concernant l'habitation. Les équipes de vérification doivent décider des contrôles appropriés à effectuer en fonction des autres données disponibles concernant l'habitation et compte tenu de l'emplacement géographique. En définitive, si la valeur indiquée n'est pas valide, il convient de lui attribuer le code « inconnu » lorsqu'on ne procède pas par imputation dynamique. Les pays qui décident de procéder à l'imputation dynamique des valeurs non valides doivent s'appuyer sur deux caractéristiques au moins, notamment celles concernant l'alimentation d'eau sous conduite, pour obtenir des informations « connues », d'après les unités d'habitation analogues de la même zone géographique.

11. LOCAUX À USAGE D'HABITATION :
GAZ SOUS CONDUITE (A10)

586. Ce sujet fait référence à la desserte des logements en gaz sous conduite, ce dernier étant généralement défini comme gaz naturel ou d'usine, distribué sous conduite, avec enregistrement de la consommation. Ce sujet peut ne pas concerner plusieurs pays qui n'ont pas de source de gaz naturel ou n'ont pas mis en place de réseaux de distribution. (Nations Unies, 1998, par. 2.427).

587. La rubrique « gaz sous conduite » n'est pas liée aux autres rubriques concernant l'habitation, à l'exception du type d'éclairage et de combustible utilisé pour la préparation des aliments : les équipes de vérification doivent déterminer le chemin de vérification approprié, ainsi que les modalités des contrôles de cohérence à effectuer. Si la valeur de cette variable reste non valide ou incohérente, il convient de lui attribuer le code « inconnu » lorsqu'on ne procède pas par imputation dynamique. Les pays qui conviennent de procéder à l'imputation dynamique des valeurs non valides devraient s'appuyer sur deux caractéristiques au moins, telles que type d'énergie utilisée pour le chauffage, type d'immeuble, type d'unité d'habitation, matériaux de construction des murs, modalités de jouissance et état d'occupation, pour obtenir des informations « connues », d'après les unités d'habitation analogues de la même zone géographique.

12. LOCAUX À USAGE D'HABITATION : TÉLÉPHONE (A11)

588. Il s'agit de constater l'existence éventuelle du téléphone à l'intérieur des logements. Par téléphone, on entend une ligne téléphonique et non un poste, puisqu'il est possible d'en brancher plusieurs sur une seule ligne (Nations Unies, 1998, par. 2.428).

589. La rubrique concernant la présence éventuelle du téléphone n'est pas liée aux autres rubriques concernant l'habitation dans le cadre des vérifications effectuées. Toutefois, si certaines zones géographiques sont effectivement dépourvues du téléphone, l'équipe de vérification doit tenir compte de ce fait en élaborant les contrôles. Si les données indiquées sous la rubrique « téléphone » sont non valides, il convient d'attribuer le code « inconnu » lorsqu'on ne procède pas par imputation dynamique. Les pays ayant décidé d'appliquer la méthode d'imputation dynamique aux valeurs non valides, doivent s'appuyer sur au moins deux

caractéristiques, telles que le type d'unité d'habitation, les matériaux de construction des murs et les modalités de jouissance, pour obtenir des informations « connues » provenant d'unités d'habitation similaires de la même zone géographique.

13. LOCAUX À USAGE D'HABITATION :
UTILISATION DE L'UNITÉ D'HABITATION (A12)

590. La variable « utilisation de l'unité d'habitation » permet de savoir si une unité a été ou non utilisée dans sa totalité à des fins d'habitation (résidentielles). Elle peut être utilisée pour l'habitation et à des fins commerciales, de production ou autres (Nations Unies, 1998, par. 2.429).

591. La variable « utilisation de l'unité d'habitation » est indépendante des autres rubriques concernant l'habitation. Si la valeur indiquée n'est pas valide, il convient de lui attribuer le code « inconnu » lorsqu'on ne procède pas par imputation dynamique. Les pays qui optent pour l'imputation dynamique des valeurs non valides doivent s'appuyer sur deux caractéristiques au moins, telles que type d'unité d'habitation, matériaux de construction des murs, modalités de jouissance et type de propriété, pour obtenir des informations « connues », d'après les unités d'habitation analogues de la même zone géographique.

14. NOMBRE DE VOITURES (A13)

592. Le « nombre de voitures » fait référence au nombre de voitures et camionnettes dont disposent normalement les occupants de l'unité d'habitation, c'est-à-dire dont ils sont propriétaires ou qui font l'objet d'un accord plus ou moins permanent comme un contrat de location (Nations Unies, 1998, par. 2.430).

593. La variable « nombre de voitures » est indépendante des autres variables concernant l'habitation. Si dans certaines zones du pays il n'y a aucune voiture, les spécialistes peuvent envisager de mettre au point des vérifications spéciales adaptées à des zones géographiques particulières. Sinon, si la valeur indiquée n'est pas valide, il convient de lui attribuer le code « inconnu » lorsqu'on ne procède pas par imputation dynamique. Les pays qui optent pour l'imputation dynamique des valeurs non valides doivent s'appuyer sur deux caractéristiques au moins, telles que type d'unité d'habitation, matériaux de construction des murs, modalités de jouissance et type de propriété, pour obtenir des informations « connues », d'après les unités d'habitation analogues de la même zone géographique.

15. OCCUPANTS : ÉQUIPEMENTS MÉNAGERS (A14)

594. Les informations recueillies concernent la présence éventuelle d'équipements ménagers tels que lave-linge (A14a), lave-vaisselle (A14b), réfrigérateur (A14c), congélateur (A14d), téléviseur (A14e), ordinateur (A14f), selon le contexte national. (Nations Unies, 1998, par. 2.431).

595. Le fonctionnement de la plupart des appareils en question exige que l'unité d'habitation considérée ait l'électricité. Lorsque le questionnaire fait état de ces équipements, le programme de vérification mis au point doit contrôler la

disponibilité d'électricité dans ces locaux (à l'exception le cas échéant d'un réfrigérateur fonctionnant au gaz ou d'une glacière). Si dans le pays considéré le lave-linge ou le lave-vaisselle doit fonctionner à l'eau courante, cette exigence doit être prise en compte dans le programme de vérification. Il est également possible, au moyen de certains contrôles, de vérifier si un élément particulier doit ou non être présent en fonction de la disponibilité d'électricité et d'eau et de déterminer les corrections à introduire lorsque des incohérences sont constatées. Là encore, des zones géographiques particulières d'un pays sont parfois dépourvues d'électricité ou d'eau courante auquel cas les spécialistes doivent en tenir compte lors de la mise au point de leurs vérifications. Si la valeur indiquée est non valide ou incohérente, il convient de lui attribuer le code « inconnu » lorsqu'on ne procède pas par imputation dynamique. Les pays qui conviennent de procéder à l'imputation dynamique des valeurs non valides devraient s'appuyer sur deux caractéristiques au moins, telles que type d'unité d'habitation, électricité, matériaux de construction des murs, et modalités de jouissance, pour obtenir des informations « connues », d'après les unités d'habitation analogues de la même zone géographique.

16. OCCUPANTS : ESPACES EXTÉRIEURS DISPONIBLES (A15)

596. Ce sujet se réfère à l'existence éventuelle d'un espace extérieur prévu pour les loisirs des membres d'un ménage occupant une unité d'habitation. La classification vise l'espace extérieur faisant partie de l'unité de logement (le jardin par exemple dans le cas d'un pavillon), l'espace extérieur proche du bâtiment (par exemple, une cour ou un terrain de jeu situé près de l'immeuble), un espace commun de loisir à moins de 10 minutes de marche de l'unité de logement (par exemple, jardin public, centre sportif, etc.), ou des

espaces extérieurs situés à plus de 10 minutes à pied (Nations Unies, 1998, par. 2.432).

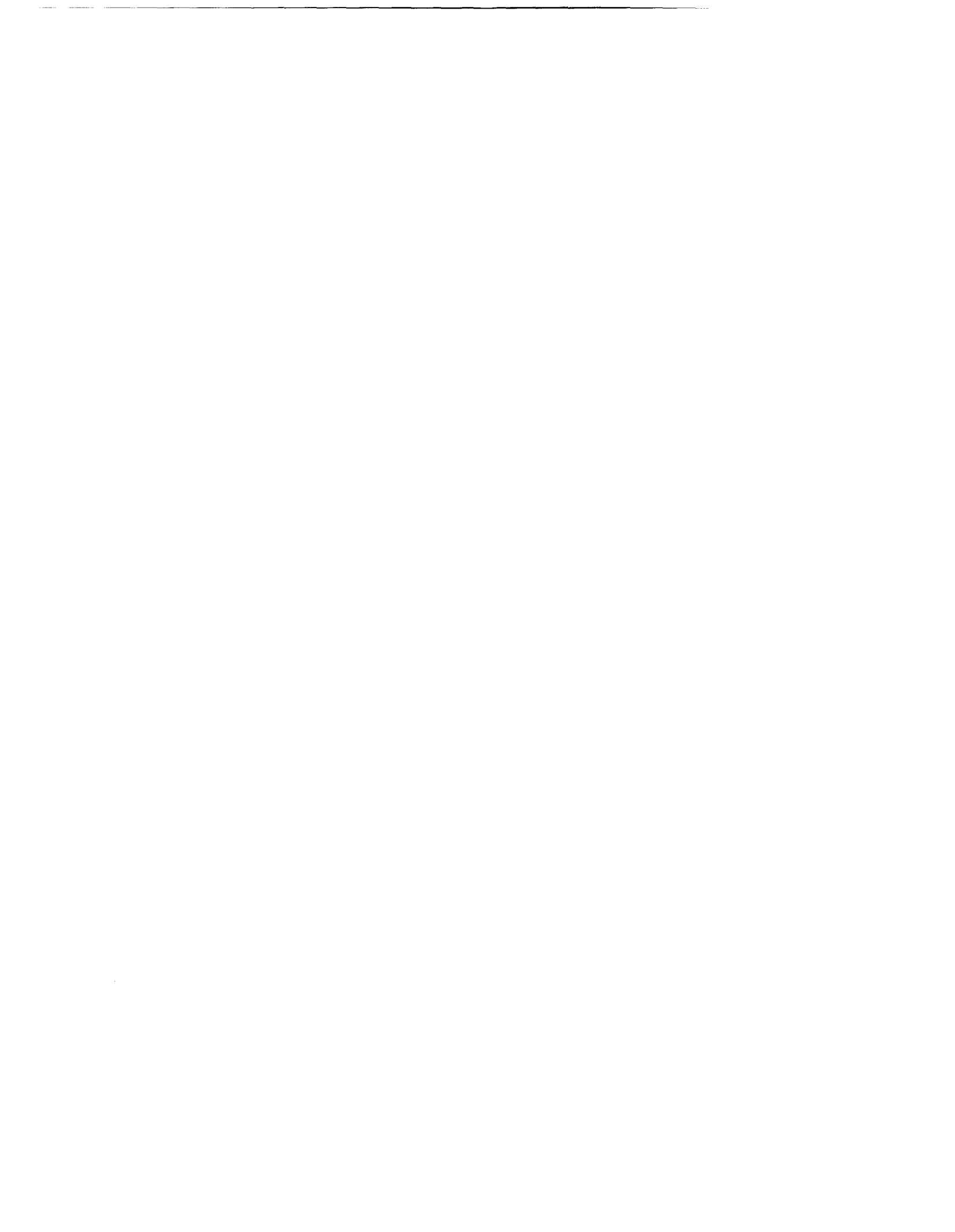
597. L'espace extérieur dont disposent les occupants du ménage est indépendant des autres rubriques concernant l'habitation. Toutefois, dans certaines zones géographiques ou pour certains types d'immeubles, l'espace extérieur disponible est parfois inexistant. Les équipes de vérification peuvent être amenées à envisager le contexte particulier en présence lorsqu'elles définissent leur programme de vérification. Si la valeur indiquée est non valide ou incohérente, il convient de lui attribuer le code « inconnu » lorsqu'on ne procède pas par imputation dynamique. Les pays qui conviennent de procéder à l'imputation dynamique des valeurs non valides devraient s'appuyer sur deux caractéristiques au moins, par exemple le type d'immeuble et le type d'unité d'habitation, pour obtenir des informations « connues » d'après les unités d'habitation analogues de la même zone géographique.

C. — Unités d'habitation occupées et inoccupées

598. Les vérifications décrites ci-dessus se rapportent aux unités d'habitation occupées. Toutefois, les unités d'occupation inoccupées et occupées présentent souvent des caractéristiques différentes et ne feront pas l'objet des mêmes vérifications. Aussi, l'équipe chargée de cette tâche au sein du bureau national de recensement/statistique devra-t-elle mettre au point des vérifications pour chaque type d'unité, lorsque (comme cela est généralement le cas) toutes les données concernant l'habitat ne sont pas recueillies pour les unités d'habitation inoccupées : l'équipe de vérification devra procéder avec le plus grand soin au choix des variables de la matrice d'imputation, celles-ci étant très vraisemblablement différentes.



ANNEXES



ANNEXE I

Variables calculées

599. Afin de tirer le meilleur parti des données recueillies à l'occasion des recensements ou des enquêtes, les pays ont fréquemment besoin d'utiliser des variables qui sont elles-mêmes des combinaisons d'autres variables. Par exemple, les informations concernant la situation d'activité économique (voir chapitre IV section D.1) constituent d'ores et déjà une combinaison de plusieurs variables recueillies lors du recensement. Au lieu de devoir élaborer un programme de recodage, à chaque fois que le bureau national de recensement/statistique veut obtenir un tableau particulier, les informaticiens peuvent rédiger un programme permettant d'effectuer le recodage une fois pour toutes, de stocker les informations recodées dans l'enregistrement personnel considéré, puis de les utiliser pour établir d'autres tableaux. Les bureaux nationaux de recensement/statistique doivent évaluer la fréquence d'utilisation d'informations recodées, ainsi que la pertinence d'une information recodée particulière, lorsqu'ils prennent une décision de produire et d'enregistrer l'information. Il faut, par ailleurs, ne pas oublier que les informations recodées occupent de la place dans les enregistrements personnels. L'espace nécessaire à cet effet sera d'autant plus considérable que l'effectif de la population sera important.

600. Il est possible de créer de cette façon un grand nombre de variables. Par exemple, si la date de naissance est indiquée mais non l'âge, celui-ci peut être obtenu en une seule opération en soustrayant la date de naissance de la date de référence du recensement, et cette information sera stockée dans l'enregistrement. De manière analogue, le revenu du ménage peut être obtenu en additionnant les revenus de chaque individu et en plaçant la somme calculée dans l'enregistrement relatif à l'habitation en vue d'une utilisation ultérieure.

601. Parfois, les variables calculées peuvent provenir de la combinaison d'une ou plusieurs données figurant dans un enregistrement particulier ou parfois dans plusieurs enregistrements. Par exemple, la classification dans la catégorie « population inactive-fréquentant l'école » peut exiger l'examen des réponses fournies à quatre questions. Ainsi, lors de la définition de formats de tableaux ou de la conception de tableaux supplémentaires, l'utilisation de variables calculées facilitera la programmation et augmentera son efficacité, et contribuera à améliorer la comparabilité des données dans le temps. On trouvera ci-dessous plusieurs exemples d'enregistrements calculés.

A. — Variables calculées associées aux données sur l'habitation

1. REVENU DU MÉNAGE

602. La variable calculée concernant le revenu du ménage est égale à la somme des revenus de toutes les catégories de revenus, de chacune des personnes d'un ménage. Parmi les différentes catégories de revenus figurent les salaires, les revenus d'activités d'entreprises, les intérêts et dividendes, les prestations de sécurité sociale, les retraites, les envois d'argent, les rentes et les loyers. Si l'on veut obtenir le revenu total il convient, dans le cadre de la vérification, de vérifier le revenu total de chaque personne en additionnant les catégories individuelles. Ce total est ensuite vérifié par rapport au revenu total indiqué. Si le revenu cumulé obtenu est différent du revenu total déclaré, les équipes de vérification doivent mettre au point une correction. Il faut soit modifier le total, de façon à ce qu'il corresponde à la somme des parties, soit modifier les chiffres concernant une ou plusieurs catégories individuelles. Une fois fixés les revenus totaux pour chacun des membres, la variable correspondant au revenu du ménage est obtenue en additionnant les revenus individuels.

603. L'équipe de vérification doit tenir compte de l'éventualité d'un revenu devenu négatif pour une ou plusieurs personnes du ménage, à la suite d'une faillite économique ou d'une autre cause. Dans ce cas, le revenu total du ménage sera réduit et non augmenté, du « revenu » de cette personne.

2. REVENU FAMILIAL

604. La variable calculée correspondant au revenu familial est égale à la somme des revenus de toutes catégories de chacune des personnes d'une famille. Contrairement aux ménages, les familles sont d'ordinaire uniquement constituées d'individus apparentés, bien que cette définition soit propre à la situation particulière du pays considéré. Pour certains pays, ménages et familles sont identiques, de telle sorte qu'une variable calculée correspondant au revenu familial s'avère inutile. Parmi les catégories de revenu familial peuvent figurer les salaires, les revenus d'activités commerciales, les intérêts et dividendes, les revenus de prestations sociales, les retraites, les envois de fonds, les rentes ou les loyers. Si les informations recueillies concernent également le revenu total, il conviendra lors de la véri-

fication de contrôler le revenu total de chaque individu, en additionnant les catégories individuelles. Ce total est ensuite vérifié par rapport au revenu total déclaré. Si le revenu cumulé obtenu est différent du revenu total déclaré, les équipes de vérification doivent mettre au point une correction. Il faut soit modifier le total, de façon à ce qu'il corresponde à la somme des parties, soit modifier les chiffres concernant une ou plusieurs catégories individuelles. Une fois fixés les revenus totaux pour chacun des membres, la variable correspondant au revenu familial est obtenue en additionnant les revenus individuels.

605. L'équipe de vérification doit tenir compte de l'éventualité d'un revenu devenu négatif pour une ou plusieurs personnes de la famille, à la suite d'une faillite économique ou d'une autre cause. Dans ce cas, le revenu total de la famille sera réduit et non augmenté, du « revenu » de cette personne.

3. TYPE DE FAMILLE

606. Il est parfois intéressant d'identifier le type de famille dans certains tableaux. Par exemple, une variable calculée correspondant au type de famille pourrait prendre des valeurs allant de 1 à 8, et représenter le type et la composition d'une famille. Cette variable calculée pourrait servir à étudier l'incidence de différentes caractéristiques sur la structure familiale.

607. Tel qu'indiqué dans la publication intitulée *Principes et recommandations concernant les recensements de la population et de l'habitat*, première révision (Nations Unies, 1998, par. 2.60-2.66), les définitions de la famille varient d'un pays à l'autre. Aux termes d'une de ces définitions, une famille comprend un chef de ménage ainsi qu'une ou plusieurs autres personnes qui vivent dans le même ménage, apparentées au chef de ménage, par la naissance, par alliance ou par adoption. Toutes les personnes vivant dans le ménage et apparentées au chef de ménage font partie de la famille de ce dernier. Toutefois, certains ménages ne sont pas constitués de famille du fait qu'un ménage peut comprendre un groupe de personnes non apparentées ou être constitué d'une personne vivant seule.

608. Les spécialistes du sujet peuvent classer les familles par type en « familles composées de couples mariés », ou « autres types de familles » en fonction du sexe du chef de ménage et de la présence de parents. Les Etats-Unis par exemple ont utilisé les codes suivants, d'après les données recueillies sur les types de familles, obtenues à partir des réponses aux questions portant sur le sexe et le lien avec le chef de ménage :

a) Codes 1 et 2. **Famille composée d'un couple marié.** Famille dans laquelle le chef de ménage et son conjoint ont été recensés en tant que membres du même ménage; le code 1 est utilisé lorsque le chef de ménage est un homme et le code 2 lorsque le chef de ménage est une femme.

b) Code 3. **Autre type de famille : chef de ménage de sexe masculin, épouse absente.** Famille comportant un chef de ménage de sexe masculin sans conjoint.

c) Code 4. **Autre type de famille : chef de ménage de sexe féminin, sans mari.** Famille comportant un chef de ménage de sexe féminin, sans conjoint.

d) Codes 5 et 6. **Ménage de type non familial.** Ménage ne constituant pas une famille, c'est-à-dire ne comportant pas de conjoint ou d'autres personnes apparentées au chef de ménage; le code 5 est utilisé lorsque le chef de ménage est un homme et le code 6 lorsque le chef de ménage est une femme.

e) Codes 7 et 8. **Ménage constitué de célibataire.** Un célibataire vivant seul est considéré comme un ménage, mais non comme une famille, en raison de l'absence de tout parent. On utilise le code 7 lorsque le chef de ménage est un homme et le code 8 lorsque le chef de ménage est une femme.

609. Une méthode plus simple pour identifier une partie des catégories ci-dessus consiste à définir une variable calculée intitulée « chef de ménage, marié avec conjoint ». L'état matrimonial du chef de ménage peut être enregistré en fonction de la présence éventuelle au foyer du conjoint du chef de ménage. Dans chaque unité d'habitation, les enregistrements concernant la population sont exploités pour explorer la présence d'une personne dont la variable lien avec le chef de ménage est mise à la valeur « conjoint ». Un code unique correspondant à la valeur « Oui » ou « Non » est stocké dans le champ approprié de l'enregistrement relatif à l'habitation. Pour des locaux collectifs à usage d'habitation il est possible soit de laisser cette variable en blanc, soit de lui attribuer un autre code. Les personnes mariées dont le conjoint est présent seront ainsi identifiées comme telles lors de la constitution de tableaux de résultats concernant la population.

4. PERSONNES APPARENTÉES

610. Les personnes dites apparentées sont celles qui sont parentes d'une façon ou d'une autre avec le chef de ménage. La variable calculée concernant les personnes apparentées est constituée par la somme de toutes les personnes parentes avec le chef de ménage. Cette valeur joue un rôle particulièrement important lorsque de nombreuses personnes non apparentées vivent ensemble au sein des unités d'habitation. Lorsqu'un grand nombre de personnes non apparentées vivent ensemble de cette façon, il est fréquent de les classer dans la catégorie des personnes qui vivent dans des « locaux collectifs à usage d'habitation » ou dans des « logements collectifs ».

611. Lors de la création de jeux de données, les bureaux nationaux de statistique définissent souvent des variables calculées pour des jeux différents de personnes apparentées, regroupées par tranche d'âge. Par exemple, des variables calculées peuvent être définies pour des enfants apparentés de 0 à 5 ans, de 5 à 17 ans, de 6 à 17 ans, ou de 0 à 17 ans, ou encore pour des personnes de 65 ans et plus ou de 75 ans et plus.

612. Les « enfants apparentés » pourraient comprendre par exemple les propres enfants du chef de ménage, ainsi que les personnes de moins de 18 ans, indépendamment de

leur état matrimonial, qui sont apparentées au chef de ménage, à l'exception de son conjoint. Parmi les enfants apparentés pourraient ou non figurer les enfants adoptifs, ces derniers n'étant pas apparentés au chef de ménage, cette décision étant toutefois fonction du contexte national particulier en présence.

5. TRAVAILLEURS DE LA FAMILLE

613. Les pays souhaitent parfois comparer les variables concernant le ménage en fonction du nombre de travailleurs, par exemple les distributions du revenu en fonction de la taille du ménage, et les nombres de travailleurs par personne à charge. Le pays peut ainsi déterminer la variable calculée correspondant au nombre de travailleurs de la famille, en additionnant le nombre de personnes ayant travaillé au moins une heure au cours d'une période de référence, par exemple une semaine ou un an (année calendaire ou bien douze derniers mois). Ainsi le nombre de personnes ayant travaillé au cours de la semaine passée pourrait être utilisé si les données recueillies concernent exclusivement cette période.

6. INSTALLATIONS COMPLÈTES DE PLOMBERIE

614. Plusieurs rubriques du questionnaire de recensement permettent de recueillir des données sur les installations de plomberie. Ces rubriques sont généralement associées à la présence de canalisations d'eau, d'un cabinet d'aisances à chasse d'eau, d'une baignoire ou d'une douche et ces données sont généralement recueillies auprès des unités d'habitation tant occupées qu'inoccupées. Une variable calculée associée aux installations complètes de plomberie peut faciliter la comparaison de la situation socio-économique de différents groupes à un moment donné, ou son évolution dans le temps. Ce type de variable calculée pourrait être obtenu par exemple lorsque 3 types d'équipement, conduites d'eau (chaude ou froide), lieu d'aisances à chasse d'eau et baignoire ou douche, sont présents (soit à l'intérieur de l'unité d'habitation, soit à l'extérieur de l'immeuble dans lequel l'unité d'habitation se trouve). L'équipe de vérification devra déterminer le jeu de variables le plus approprié concernant l'installation complète de plomberie.

615. Dans cet exemple la variable calculée peut être obtenue lorsque la collecte de données porte séparément sur les trois rubriques, et pendant la vérification, la présence simultanée de ces trois équipements est établie. Si l'unité d'habitation est munie de canalisations d'eau, d'un lieu d'aisances à chasse d'eau et d'une baignoire et d'une douche, elle est alors équipée d'une installation complète de plomberie. En l'absence de ces trois éléments, elle est dépourvue d'installations complètes de plomberie.

7. INSTALLATION COMPLÈTE DE PRÉPARATION DES ALIMENTS

616. Les recensements sont mis à profit pour recueillir des données sur les installations de préparation des aliments à partir des rubriques du questionnaire concernant ce type d'équipements, ainsi que les réfrigérateurs et les éviers; ces données sont recueillies pour les unités d'habitation tant occupées qu'inoccupées. Ainsi, une unité d'habitation peut

être considérée comme étant équipée d'installations complètes de préparation des aliments, quand des installations de préparation des aliments (fourneau électrique, au kérosène ou au gaz, four micro-ondes et foyer non portatif ou cuisinière), un réfrigérateur et un évier muni de canalisations d'eau se trouvent dans le même immeuble que les locaux d'habitation recensés. Ils ne doivent pas se trouver nécessairement dans la même pièce.

617. L'obtention de la variable calculée comporte dans un premier temps la collecte séparée des trois éléments d'information ci-dessus et dans le cadre des vérifications, la détermination de la présence conjointe de ces trois éléments. « L'absence d'installation complète de préparation des repas » correspond à une situation dans laquelle les trois équipements spécifiés sont des équipements présents, mais ne se trouvent pas dans le même immeuble; une partie mais non la totalité des équipements sont présents ou encore, aucun des trois équipements spécifiés de préparation des aliments n'est présent dans le même immeuble que les locaux d'habitation recensés.

8. LOYER BRUT

618. Les pays recueillent généralement des données concernant les loyers réglés en espèces ou les loyers charges comprises. Les loyers payés en espèces, ne comprennent généralement pas les charges. Parfois, les pays ont besoin en outre de recueillir des informations sur le loyer brut. Le loyer brut est défini comme le loyer net augmenté du coût mensuel moyen estimé des charges (électricité, gaz et eau) et des combustibles utilisés (notamment, le mazout, le charbon, le kérosène et le bois) si les règlements correspondants incombent aux locataires. Le loyer brut est censé éliminer les différences résultant des diverses pratiques quant à la prise en compte des charges et des combustibles au titre du règlement du loyer; les unités d'habitation louées occupées sans paiement de loyer peuvent figurer séparément dans les tableaux sous une rubrique « sans loyer ».

619. La variable calculée relative au loyer brut est obtenue en additionnant le montant du loyer et celui des charges, si ces dernières sont payées séparément.

B. — Variables calculées associées aux données sur la population

1. SITUATION D'ACTIVITÉ ÉCONOMIQUE

620. La disponibilité d'une variable calculée concernant la situation d'activité économique peut être extrêmement précieuse pour la mise en tableau des résultats, mais cela exige des données provenant de différentes variables. Si l'on utilise les catégories définies dans la publication *Principes et recommandations concernant les recensements de la population et de l'habitat*, première révision, une reconfiguration de plusieurs variables est nécessaire. La variable calculée pourrait comporter 9 catégories :

Population active

Personnes pourvues d'un emploi

1. Personnes au travail

- 2. Personnes qui ont un emploi mais ne sont pas au travail
- 3. Personnes engagées dans les forces armées

Personnes sans travail

- 4. A la recherche d'un travail
- 5. Travailleur découragé

Population inactive

- 6. Personnes s'occupant du foyer
- 7. Etudiants
- 8. Personnes incapables
- 9. Autres

621. Les différentes classifications au regard de l'activité économique étant utilisées dans nombre des travaux connexes, l'équipe de vérification doit examiner la possibilité d'introduire une variable calculée dans les enregistrements de données, plutôt que de prévoir un traitement comportant une reclassification de la situation d'activité économique au cours de la mise en tableaux. En effet, cette reclassification risque d'introduire des erreurs puisque les différentes personnes chargées de définir le traitement des données risquent de mettre au point des séquences de reclassification légèrement différentes; ainsi, même un simple programme risque de reclasser différemment les informations en fonction des exigences particulières de la vérification ou de la mise en tableau. Des spécialistes des classifications économiques doivent élaborer les spécifications concernant la variable calculée.

2. NUMÉRO DE SOUS-FAMILLE ET PERSONNES APPARENTÉES À L'INTÉRIEUR D'UNE SOUS-FAMILLE

622. Il existe dans tous les pays des familles étendues et des familles nucléaires. Considérons la famille étendue décrite ci-dessous, les triangles représentant les hommes et les octogones représentant les femmes. Le ménage comporte un chef de ménage et son conjoint (n° 1 et n° 2), ainsi que 2 enfants (n° 4 et n° 7). Leur fils (personne n° 4) est marié à leur belle-fille (personne n° 3) et ils ont deux petits-enfants de leur fils (personne n° 5 et n° 6). Leur fille (personne n° 7) est

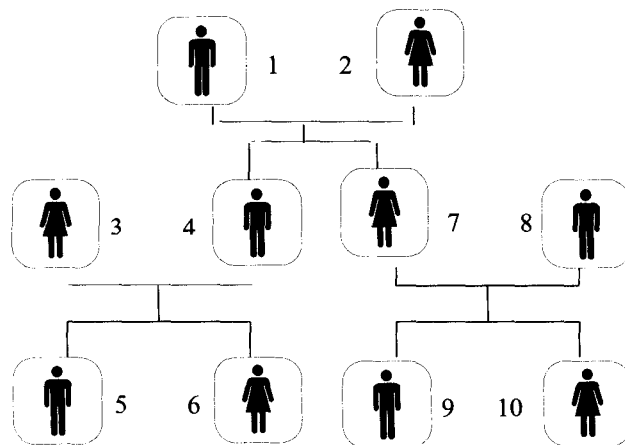
mariée à leur gendre (personne n° 8) et ils ont deux petits-enfants de leur fille (personnes n° 9 et n° 10).

623. Dans la plupart des recensements, si l'équipe d'édition souhaite étudier la structure d'une famille comme celle représentée à la figure A.I.1, la distinction entre les petits-enfants risque de s'avérer délicate, puisque les personnes 5, 6, 9 et 10 auront toutes la mention « petits-enfants » enregistrée au titre de leur lien avec le chef de ménage. Le recodage de la sous-famille et de chacun de ses membres permettra une analyse plus détaillée de la structure familiale.

624. Une sous-famille pourrait être définie comme « un couple marié (mari et femme dénombrés en tant que membre du même ménage) avec ou sans enfant jamais marié de moins de 18 ans ». L'équipe de vérification peut souhaiter étendre cette définition en ajoutant « un parent avec un ou plusieurs enfants jamais mariés, de moins de 18 ans, vivant dans un ménage et apparenté soit au chef de ménage, soit à son conjoint, sans toutefois inclure ces derniers ». Le nombre de sous-familles n'est pas compris dans le décompte des familles, puisque les membres des sous-familles sont décomptés comme faisant partie de la famille du chef de ménage.

625. Les variables calculées relatives aux sous-familles, notamment le nombre et le type de parents, peuvent être définies au cours du traitement des données. Une vérification spéciale peut être mise au point pour faciliter la détermination des sous-familles en fonction des relations à l'intérieur du ménage. Au fur et à mesure de la détermination de chaque sous-famille — un couple constitué d'une personne qui n'est pas chef de ménage et de son conjoint (avec ou sans enfant), ou une personne qui n'est pas chef de ménage ou un parent et des enfants — un numéro séquentiel est attribué à chaque sous-famille. Des codes peuvent ensuite être attribués aux différents liens : le code 1 au « chef de ménage » de la famille, le code 2 au « conjoint » du « chef de ménage » de la famille, et le code 3 à un enfant du « chef de ménage » de la famille. L'existence d'une sous-famille implique celle d'au moins un couple de parents au sein de cette sous-fa-

Figure A.I.1. Illustration d'une famille étendue



mille : un « chef de ménage » et un « conjoint » (codes 1 et 2), ou bien un « chef de ménage » et un « enfant », (codes 1 et 3). Lorsque le chef de ménage, le conjoint et les enfants d'une famille cohabitent, les codes 1, 2 et 3 sont tous attribués au sein de la sous-famille. Les sous-familles sont classées par type : sous-familles constituées de couples mariés, avec ou sans leurs propres enfants; sous-familles mère-enfant, et sous-familles père-enfant. Les parents seuls font partie de familles ou de sous-familles monoparentales. Les couples mariés se composent de maris et de femmes vivant au sein de familles ou de sous-familles constituées de couples mariés.

626. Pour la définition des variables calculées, la relation avec le chef de ménage sert à déterminer les liens avec le chef de la famille; par conséquent, la correspondance avec les sous-familles sera d'autant plus rigoureuse que le codage du lien avec le chef de ménage réalisé dans le cadre du recensement sera plus détaillé, par exemple, si la relation « enfant du conjoint » a son propre code, le programme sera en mesure d'apparier un « fils-fille » à un « fils-fille de conjoint » du sexe opposé pour créer une sous-famille. Sans cette information supplémentaire, l'appariement pourra encore être effectué, mais la relation « autre parent » risque d'être ambiguë après appariement, à moins que l'appariement ne soit erroné. De manière analogue, le programme associera les codes des « frères et sœurs du chef de ménage » aux « conjoints par alliance » et aux « nièces/neveux ».

627. L'exemple de la figure A.I.2, présente un ménage avec deux sous-familles : la sous-famille 1 se compose de la personne n° 3 (chef de ménage de la sous-famille 1), de la personne n° 4 (conjoint) et des personnes n° 5 et n° 6 (enfants de la sous-famille 1). La sous-famille 2 se compose de la personne n° 7 (chef de ménage de la sous-famille 2), de la personne n° 8 (conjoint) et des personnes n° 9 et n° 10 (enfants de la sous-famille 2).

3. ENFANTS DONT LA MÈRE EST VIVANTE

628. Les pays souhaitent parfois obtenir des informations concernant des enfants dont la « mère est vivante », c'est-à-dire les enfants biologiques du chef de ménage ou du conjoint. Un enfant dont la mère est vivante pourrait être par exemple un enfant non marié de moins de 18 ans qui peut être un fils ou une fille par la naissance, un enfant par alliance (de l'un des parents) ou un enfant adoptif du chef de ménage. Les tableaux pourraient indiquer les enfants dont la mère est vivante en précisant s'ils vivent avec les deux parents ou avec seulement l'un d'entre eux. Dans ce système de classification les enfants dont la « mère est vivante » du chef de ménage qui vivent avec les 2 parents existent par définition uniquement dans le cas des familles constituées des couples mariés.

629. Dans une sous-famille un « enfant dont la mère est vivante » peut désigner un enfant non marié de moins de 18 ans (fils, fille ou enfant par alliance); ou encore un enfant adoptif : 1) vivant avec sa mère dans une sous-famille mère-enfant; 2) vivant avec un père dans une sous-famille père-enfant; ou 3) vivant avec l'un ou l'autre conjoint dans une sous-famille constituée d'un couple marié.

630. La variable calculée associée aux « enfants dont la mère est vivante » pourrait être le nombre total d'enfants d'une personne donnée, généralement une femme, conformément aux définitions adoptées par les équipes de vérification. Parfois, les utilisateurs ont besoin d'informations plus détaillées sur la répartition par âge des enfants dont un parent est vivant. Aux Etats-Unis par exemple, des variables calculées sont définies pour le nombre d'enfants dont un parent est vivant, d'une part pour ceux âgés de moins de 6 ans, et d'autre part pour ceux qui ont entre 6 et 17 ans. Ces valeurs sont inscrites dans les enregistrements concernant la population féminine et sont utilisées en particulier pour déterminer les caractéristiques des femmes ayant des enfants au sein de la population active.

Figure A.I.2. Exemple de ménage comprenant deux sous-familles

Numéro de la personne	Lien avec le chef de ménage	Sexe	Sous-famille	
			Numéro	Lien avec le chef de ménage
1	Chef de ménage	M		
2	Conjoint	F		
3	Fils	F	1	1
4	Belle-fille	M	1	2
5	Petit-enfant	M	1	3
6	Petit-enfant	F	1	3
7	Fille	F	2	1
8	Gendre	M	2	2
9	Petit-enfant	M	2	3
10	Petit-enfant	F	2	3

4. PARENTS VIVANT À LA MAISON

631. Ces données présentent les caractéristiques des enfants de familles monoparentales par comparaison aux unités d'habitation dans lesquelles vivent les deux parents. Le programme de vérification établit cette variable calculée en déterminant le nombre de parents d'une personne donnée présents au foyer, sur la base des codes attribués aux liens de parenté. Le programme examine le code attribué aux liens de parenté dans le cas de chaque enfant et utilise conjointement cette information avec celle qui concerne les sous-familles pour déterminer le nombre de parents qui vivent dans l'unité d'habitation.

5. ANNÉE DE SCOLARITÉ DU MOMENT

632. Certains pays posent deux questions concernant l'éducation :

- a) La personne fréquente-t-elle actuellement l'école ?
- b) Quel est le niveau de formation atteint ?

633. Dans ces pays, les équipes de vérification constatent fréquemment un défaut de concordance entre les deux réponses lorsqu'une personne poursuit effectivement des études au moment du recensement. Parfois, le niveau le plus élevé atteint par la personne considérée risque d'être inférieur d'une année au niveau d'étude du moment. Si la personne se trouve à un niveau intermédiaire d'un cursus marqué par des niveaux successifs, les statistiques n'en seront

pas modifiées. Toutefois, si la personne en question suit la première année d'étude d'un cursus en vue de l'obtention d'un niveau déterminé, il ne sera sans doute pas possible d'établir une correspondance avec les données provenant d'autres sources. Par exemple, une personne inscrite en première année d'étude sera enregistrée comme étant scolarisée mais ne sera nullement censée avoir achevé un cycle d'étude quelconque. De manière analogue, quelqu'un qui commence des études secondaires sera enregistré comme étant scolarisé, mais son niveau de formation sera la dernière année d'études primaires.

634. Il est possible de définir une variable calculée intitulée « niveau scolaire du moment » correspondant à cette combinaison de rubriques. Si la personne ne fréquente pas actuellement l'école, le code qui lui sera attribué sera le niveau de formation le plus élevé. Si la personne fréquente actuellement l'école, le programme de vérification ajoutera une unité au niveau de formation (ou au nombre d'années d'étude) et attribuera cette valeur à la variable « niveau scolaire du moment ».

635. Certains pays posent trois questions concernant l'éducation, les deux questions ci-dessus, ainsi qu'une troisième sur le fait de savoir si la dernière année d'étude a été achevée. Si cette information est également recueillie, elle devrait servir en outre à la détermination du « niveau scolaire du moment ».

ANNEXE II

Lien entre présentation du questionnaire et introduction des données

636. Les pages concernant les personnes et celles concernant les ménages constituent les deux présentations les plus courantes des questionnaires destinés à recueillir des informations sur la population, dans le cadre d'un recensement ou d'une enquête.

637. La partie des questionnaires relative aux personnes se compose d'une page simple ou d'une double page, où sont recueillies des informations sur la population, avec une page distincte pour chaque personne.

Cette méthode est utile parce qu'elle permet de faire figurer sur une seule page toutes les informations concernant une personne, ce qui facilite la collecte. De plus, cette présentation permet de mieux contrôler la cohérence interne des données au cours du recensement. Les fiches personnelles peuvent être regroupées sous la forme d'un cahier pour en faciliter la manipulation sur le terrain, tel qu'indiqué à la figure A.II.1.

638. Le codage et la saisie des données à inscrire sur les pages personnelles constituent à toutes fins pratiques une opération mécanique, pour laquelle le codeur/l'opérateur de

saisie n'est pas censé évaluer la validité des informations fournies mais plutôt leur attribuer le code ou la touche de clavier appropriée. La figure A.II.2 représente le flux d'informations concernant une personne donnée, enregistrées sur une fiche particulière. Il est plus facile d'inscrire les données sur une page unique concernant cette personne plutôt que de les introduire au clavier en tournant des pages. Les contrôles de validité sont effectués ultérieurement au cours des vérifications automatiques.

639. Les feuilles de recensement des ménages présentent l'ensemble des informations concernant un ménage si possible sur une seule page, ou sur une série de pages, avec énumération de tous les membres du ménage sur chaque page. Cette façon de répertorier les membres du ménage est intéressante puisque les questions posées n'ont pas à être imprimées pour chaque individu, ce qui économise de la place. En outre, l'agent de recensement peut comparer les réponses fournies aux différents membres du ménage au fur et à mesure de leur obtention.

FIGURE A.II.1. Exemple de questionnaire constitué de pages personnelles

<i>Page personnelle concernant la personne X</i>		<i>Page personnelle concernant la personne X + 1</i>	
Rubrique 1	Rubrique 10	Rubrique 1	Rubrique 10
Rubrique 2	Rubrique 11	Rubrique 2	Rubrique 11
.	.	.	.
.	.	.	.
.	.	.	.

FIGURE A.II.2. Exemple de séquence de collecte de l'information sur les pages personnelles d'un questionnaire

<i>Page personnelle</i>			
Rubrique 1	↓	Rubrique 11	↓
Rubrique 2		Rubrique 12	
Rubrique 3		Rubrique 13	
etc.		etc.	
.		↑	
.			
.			



FIGURE A.II.3. Exemple de questionnaire, page concernant un ménage, sur laquelle figurent toutes les personnes du ménage

Page concernant un ménage

<i>Personne</i>	<i>Rubrique 1</i>	<i>Rubrique 2</i>	<i>Rubrique 3</i>	<i>Rubrique 4</i>	<i>Etc.</i>
1					
2					
3					
4					
5					
.					
.					
.					

FIGURE A.II.4. Exemple de séquence de traitement des informations inscrites sur des pages de questionnaire concernant un ménage, avec plusieurs personnes par page

Page concernant un ménage

<i>Personne</i>	<i>Rubrique 1</i>	<i>Rubrique 2</i>	<i>Rubrique 3</i>	<i>Etc.</i>		
1						
2						
3						
4						
5						
.						
.						
.						

640. Une troisième méthode consiste à prévoir une feuille distincte pour chaque personne, l'agent de recensement les regroupant en un cahier séparé, pendant ou après le dénombrement. Cette méthode est efficace puisque l'agent de recensement réuni le strict nombre de feuilles (pages) nécessaires pour le ménage considéré. Son inconvénient vient du fait que les feuilles peuvent se dissocier au cours des manipulations, ce qui peut créer des problèmes éventuels de vérification et de couverture si le bureau de recensement n'est pas en mesure de réunir les fiches correspondant au ménage approprié.

641. La taille matérielle des pages du questionnaire doit également être examinée, non seulement eu égard au dénombrement, mais aussi du point de vue de la saisie. Au cours du codage et de l'introduction des données, le document doit reposer à plat sur la surface de la table de travail, et les codeurs ou les opérateurs de saisie doivent pouvoir repérer et examiner commodément les informations inscrites sur le formulaire.

642. Lorsque toutes les informations figurent sur une seule page, le personnel peut également introduire aisément

le contenu des pages concernant les ménages, et cette opération sera évidemment plus rapide puisque l'opérateur n'aura pas à tourner les pages. La figure A.II.4 présente la séquence de traitement des informations figurant sur une page concernant un ménage.

643. L'introduction de données sur la population ou l'habitation occupant plusieurs pages peut cependant soulever des difficultés. Afin de les résoudre, le bureau national de statistique est susceptible d'opter pour l'une des deux méthodes ci-dessous :

644. Possibilité d'introduire les données personne par personne. L'opérateur peut introduire la ligne d'informations concernant une personne sur la première page de la série, puis passer à la seconde et aux pages suivantes. Après avoir rempli les pages concernant la première personne, l'opérateur de saisie revient à la première page concernant le ménage en question, puis introduit les données concernant la deuxième personne, la troisième, etc. Ce type de saisie se déroule directement tant que l'opérateur peut rester sur la même ligne de données tout au long de l'introduction. Bien que l'on puisse concevoir des programmes informatisés de

vérification afin de démêler les données enregistrées, lorsque les réponses personnelles sont affectées par erreur à la ligne d'une autre personne, le programme proprement dit est très difficile à élaborer.

645. Possibilité d'introduction des données page par page. L'opérateur de saisie peut introduire une page entière de données avant de passer à la page suivante. Dans ce cas, il introduit toutes les informations figurant sur la première page, indépendamment du nombre de personnes concernées. Ensuite, il tourne la page et introduit les autres informations concernant l'ensemble des personnes. Les sauts de champ peuvent ou non être prises en compte, suivant le type de saisie (avec ou sans vérification automatisée). De toute façon, au cours de la vérification informatisée, les enregistrements des différents jeux de données introduites devront être regroupés et il faudra alors corriger toutes les erreurs éventuelles concernant les numéros affectés aux personnes.

646. Dans l'exemple ci-dessous (figure A.II.5), les données démographiques concernant le ménage ne soulèvent aucun problème particulier de saisie puisque les agents de recensement ont recueilli une réponse pour toutes les rubriques et toutes les personnes.

647. Toutefois, une deuxième feuille concernant le même ménage (figure A.II.6) pourrait présenter certains problèmes d'introduction de données. Par exemple, si le pays convient d'indiquer la langue utilisée, uniquement pour les personnes de 5 ans et plus, cette information sera laissée en blanc pour une huitième personne âgée de 3 ans. L'opérateur de saisie devrait laisser la case en blanc en ce qui concerne cet enfant, sinon la vérification automatique cherchera par la suite à la corriger.

648. De manière analogue, d'autres rubriques devraient être laissées en blanc telles que les personnes dont l'âge est inférieur à l'âge minimum légal pour travailler; les femmes

FIGURE A.II.5. Exemple de feuille concernant un ménage à plusieurs personnes, ne posant pas de problème de saisie

<i>Page concernant un ménage</i>					
<i>Personne</i>	<i>Lien avec le chef de ménage</i>	<i>Sexe</i>	<i>Age</i>	<i>Etc.</i>	
1	Chef de ménage	M	40		
2	Conjoint	F	35		
3	Enfant	F	18		
4	Enfant	M	12		
5	Frère/sœur du conjoint	M	35		
6	Frère/sœur du conjoint	F	30		
7	Frère/sœur	M	5		
8	Frère/sœur	F	3		
Etc.					

FIGURE A.II.6. Exemple de page concernant un ménage à plusieurs personnes, susceptible de poser des problèmes de saisie

<i>Page 2 concernant le ménage</i>				
<i>Personne</i>	<i>Langue</i>	<i>Population active</i>	<i>Nombre d'enfants</i>	<i>Etc.</i>
1	Langue 1	Oui		
2	Langue 1	Non	3	
3	Langue 1	Non	0	
4	Langue 1			
5	Langue 1	Oui		
6	Langue 1	Non	4	
7	Langue 1			
8				
Etc.				

n'ayant pas encore atteint l'âge minimum de procréation et les informations de fécondité concernant l'ensemble des hommes. Sur la figure A.II.6 l'opérateur de saisie risque d'introduire de façon erronée les données de la personne n° 6 concernant le nombre d'enfants (en l'occurrence 4) dans la case de la personne n° 5. La vérification automatique supprimera alors les données sur la fécondité concernant les hommes et imputera les données sur la fécondité concernant les femmes, tout en risquant de ne pas imputer la valeur exacte.

649. Dans nombre de cas un pays est tenu d'utiliser le questionnaire concernant les ménages, pour des contraintes de coût ou de place. Toutefois, lorsque l'effectif de la population est réduit ou si le pays a les moyens d'assumer les frais supplémentaires, le questionnaire constitué de feuilles personnelles contient généralement moins d'erreurs de concordance, suite à des erreurs de saisie, par comparaison aux questionnaires relatifs aux ménages.

ANNEXE III

La saisie manuelle des données

650. De nombreux pays utilisent des scanners, lecteurs optiques de marques (OMR) ou équipements de reconnaissance optique de caractères (ROC). Chacune de ces solutions offre des avantages par rapport à la saisie manuelle, dans la mesure où leur mise en œuvre s'avère souple et rationnelle et ne comporte pas de coûts excessifs. Or, beaucoup de pays, même certains qui se sont engagés sur la voie de la lecture optique, n'ont pas nécessairement les moyens d'assumer les frais initiaux ou les frais d'entretien permanents pendant et après le recensement. Par ailleurs, les scanners obtenus pour le recensement sont souvent employés à différentes fins, notamment le dépouillement d'autres enquêtes et de documents administratifs tels que les fiches d'entrée et de sortie du territoire. Cependant, le transfert des compétences de base nécessaires pour introduire des documents dans un scanner exige l'utilisation d'équipements identiques ou similaires.

651. L'introduction manuelle des données présente notamment l'avantage d'étendre le bénéfice des compétences acquises à cette occasion à d'autres activités menées dans le cadre des bureaux nationaux de recensement/statistique et d'autres administrations publiques. Lorsque le recensement a formé des opérateurs expérimentés de saisie des données, ces mêmes opérateurs sont ensuite en mesure d'effectuer cette tâche pour différentes enquêtes de suivi. Parmi ces enquêtes pourraient figurer des enquêtes postdénombrement (PES), et différents travaux, par exemple concernant la fécondité ou encore les dépenses et les revenus des ménages. Ce personnel peut en outre assurer la saisie des documents administratifs divers, (état civil, commerce, immigration et douane).

652. Lors du choix de l'utilisation d'équipements de lecture optique pour son recensement, un pays doit donc décider s'il continuera à utiliser ce matériel. Or, les machines polyvalentes restent utiles longtemps après le recensement. Ainsi, les bureaux nationaux de recensement/statistique qui procèdent à la saisie manuelle de leur données estimeront que les compétences acquises sont transférables et que les équipements proprement dits restent utiles, au sein du bureau de statistique comme dans d'autres administrations publiques. Le choix de la saisie manuelle ou de la lecture optique des données devrait sans doute dûment tenir compte de la possibilité d'utilisation permanente de ce type d'équipements.

A. — Introduction des données

1. LECTURE OPTIQUE

653. Les pays qui utilisent des scanners ou d'autres dispositifs de lecture optique pour saisir leurs données, ne les corrigent généralement pas au fur et à mesure, bien que les

modifications éventuelles puissent dépendre des instructions de saut de champ intégrées au système. Le choix de la saisie manuelle des données comporte cependant différentes options suivant la rapidité d'introduction requise et l'importance des contrôles manuels nécessaires : chacune de ces options est subordonnée aux exigences des études de vérification, aux compétences des opérateurs et à la complexité du programme de vérification. Pour les recensements de grande envergure, le problème majeur est l'introduction effective des données. La rapidité d'obtention des résultats est le critère déterminant de choix de la méthode. La publication *Principes et recommandations* fait observer que des équipements adéquats, une bonne formation et des opérateurs expérimentés permettent aux pays d'obtenir des améliorations plus rapides. Or, ces différentes conditions exigent néanmoins un financement adéquat, lequel n'est pas toujours possible.

654. La quantité et le type de matériel nécessaire pour la saisie dépendent de la méthode choisie, des délais impartis pour cette étape des opérations, de la taille du pays, du degré de décentralisation de la saisie et de plusieurs autres facteurs. En saisie au clavier, la cadence moyenne de frappe va de 5000 à 10 000 caractères à l'heure, certains opérateurs étant bien au-dessous de cette moyenne, tandis que d'autres sont bien au-dessus. Les facteurs qui influent sur la cadence sont notamment : a) le programme et le logiciel choisis; b) la complexité des tâches confiées aux opérateurs; c) les caractéristiques ergonomiques, la fiabilité et la rapidité du matériel; d) l'arrivée régulière du travail; e) la formation et les compétences du personnel; et f) la diligence des opérateurs (Nations Unies, 1998, par. 1.193).

2. SAISIE MANUELLE CONTINUE

655. L'introduction continue s'effectue sous deux formes. La première consiste à introduire toutes les données au fur et à mesure de leur apparition, sans sauter de champ. La saisie est alors plus rapide, puisque les opérateurs n'ont pas à s'arrêter quand ils trouvent des informations non valides ou incohérentes; elle peut en outre s'avérer plus précise, parce que plus mécanique. Le deuxième type d'introduction continue implique des interruptions pour vérifier la validité ou la cohérence des réponses indiquées dans le questionnaire, de telle sorte que les opérations se déroulent plus lentement, et exigent beaucoup plus de compétences de la part du personnel. Il faut absolument réfléchir soigneusement au prix élevé de cette option en termes de rapidité. Paradoxalement, cette méthode peut également améliorer la qualité lorsque les opérateurs constatent que les données ont été correctement enregistrées, mais mal codées. Quant aux erreurs d'introduction proprement dite des données, elles sont

parfois immédiatement corrigées si le progiciel de vérification assure un contrôle automatique.

a) *Introduction continue sans sauts de champ*

656. Lorsque toutes les données sont introduites ou sautées manuellement, il est possible de maintenir un certain rythme de travail et certaines instructions de saut de champ n'élimineront pas des données valides mais momentanément incohérentes. Par exemple, si une personne est déclarée comme étant de sexe masculin, les vérifications exigeront généralement de sauter toute la section de données concernant la fécondité. En l'occurrence, l'opérateur sautera les champs correspondants (au moyen de la barre d'espace ou de la flèche pour sauter l'enregistrement correspondant à un homme ou à une femme suffisamment jeune) tous les champs devant être laissés en blanc. Toutefois, cela prend du temps et l'espacement risque d'être inexact. Par exemple, l'opérateur peut introduire un espace excessif ou insuffisant, et les autres données sont alors mal introduites dans une erreur d'alignement. Si tous les champs sont saisis de cette façon, alors ces informations peuvent être saisies sans aucune instruction de saut. Ainsi, lorsqu'un opérateur doit introduire les informations concernant une femme adulte avec des données sur la fécondité (c'est-à-dire ayant indiqué un nombre d'enfants, un nombre d'enfants vivants ou un nombre d'enfants nés au cours de l'année écoulée, ces réponses ayant été codées), tous ces éléments sont introduits au clavier. Si les informations concernant la fécondité sont introduites, le programme de vérification informatisé peut déterminer quelles données ou série de données sont valides et doivent être modifiées. Si le programme de vérification établit que la personne considérée est une femme adulte, mais que les informations correspondantes de fécondité sont laissées en blanc, il faut alors obtenir les données sur la fécondité nécessaires à la mise en tableau des résultats, en procédant par imputation dynamique ou par tout autre moyen approprié. En cas de perte de l'information effectivement recueillie, en raison des instructions de saut de champ, l'équipe de vérification doit décider si le gain d'efficacité et de vitesse justifie cette perte. En présence d'instructions de saut de champ, les opérateurs peuvent néanmoins remonter aux écrans précédents pour retrouver l'emplacement des corrections introduites. Par ailleurs, ils perdront certes du temps pour franchir les champs correspondants aux éléments non introduits, mais les incohérences entre sexe, âge et données sur la fécondité peuvent être résolues au cours des vérifications, plutôt qu'au stade de la saisie..

b) *Introduction continue avec sauts de champ*

657. Une deuxième méthode d'introduction continue des données implique une saisie au clavier, en observant des instructions de saut. Là encore, si l'équipe de vérification exige l'observation de ces instructions, généralement pour reproduire le mode de collecte des données par les recenseurs, la saisie au clavier est facilitée et accélérée lorsque les instructions sont faciles à suivre et si elles sont assimilées rapidement par les opérateurs; si lesdites instructions sont particulièrement complexes, ils risquent d'être induits en erreur

et d'introduire systématiquement les données au mauvais endroit. En présence de ce type d'instruction, l'efficacité maximale de saisie est obtenue avec des instructions limitées couvrant des parties importantes de l'enregistrement.

658. L'équipe de vérification devra définir les instructions de saut de champ adaptées à l'enquête ou au recensement entrepris dans le pays. Par exemple, il est logique de sauter toutes les rubriques sur l'emploi dans le cas des enfants, c'est-à-dire des personnes n'ayant pas encore atteint l'âge minimum pour travailler fixé dans le pays considéré. Souvent, ces rubriques représentent la moitié des rubriques concernant la population et il est donc parfaitement indiqué de les sauter dans le cas des enfants, sauf situation spéciale comme celle des enfants dont l'âge est limite, ou si l'on étudie la main-d'œuvre enfantine.

659. L'équipe de vérification examine les rubriques une par une afin de décider lesquelles seront prises en compte pour chaque groupe d'âge. Les opérateurs peuvent regrouper les rubriques de façon à faciliter les sauts de champ.

660. Les décisions concernant les sauts de champ ne sont pas toujours faciles à prendre. Par exemple, si l'on considère la séquence ci-dessous :

-
1. Quelle est la citoyenneté de cette personne ?
 - Née dans ce pays (passer à la rubrique 3)
 - Naturalisée
 - Non citoyen
 2. Quelle est l'année d'arrivée de cette personne ?
 3. RUBRIQUE SUIVANTE
-

661. Une instruction de saut pourrait être définie afin de passer de la rubrique 1 à la rubrique 3, c'est-à-dire de sauter la rubrique concernant l'année d'arrivée pour les personnes nées dans le pays considéré. Toutefois, les opérateurs de saisie ne respectent pas toujours les instructions de saut, soit parce que l'agent de recensement ou le codeur a fait une erreur, soit parce qu'ils commettent eux-mêmes une erreur. Parmi les nombreux facteurs en cause figurent le niveau de compétence des opérateurs, le contexte culturel, la disposition du questionnaire et la disposition des écrans de saisie. L'équipe de vérification cherche souvent à déterminer si la définition d'instructions de saut de champ s'avère justifiée dans un cas de ce type.

3. SAISIE MANUELLE DISCONTINUE

662. Bien que l'introduction discontinue de données soit une méthode applicable dans le cadre d'un recensement, elle convient davantage aux enquêtes, en particulier aux petites enquêtes, dont les résultats risquent d'être affectés par les réponses imputées. En fonction des informations disponibles pour apporter des modifications ou introduire des corrections, la saisie discontinue peut impliquer des corrections manuelles ou automatisées.

663. Considérons le cas d'une petite enquête. En l'occurrence chacune des réponses obtenues est importante. Si un pays prélève un échantillon de 1 % de la population, chaque réponse représente par exemple 100 personnes, ou 100 unités de logements ou exploitations agricoles. Ainsi,

un petit nombre de réponses non valides ou incohérentes, pourrait avoir une incidence considérable sur les résultats. Tant les démographes que les autres spécialistes des sciences sociales souhaitent alors contrôler rigoureusement les traitements effectués.

664. Ce contrôle peut être exercé de différentes façons. Les démographes ainsi que les autres spécialistes peuvent introduire les données eux-mêmes et vérifier au fur et à mesure l'absence de réponses aberrantes, non valides ou incohérentes, d'après les données enregistrées sur les formulaires de collecte. Ils sont souvent en mesure de résoudre immédiatement des incompatibilités, des erreurs de codage ou autres incohérences, tout en examinant directement les informations recueillies. Ils peuvent parfois choisir de renvoyer sur le terrain des questionnaires incomplets ou non valides. Ce type de saisie discontinue donne les meilleurs résultats, puisque le démographe fait alors office simultanément d'opérateur de saisie, mais il s'avère de loin le plus coûteux et nombre de pays n'ont pas les moyens de recourir à cette méthode.

665. Les équipes de vérification peuvent mettre au point des règles très précises pour définir ce que les opérateurs de saisie doivent faire en présence de cas particuliers. Pour chaque problème non résolu de code non valide, ils peuvent décider de ce que l'opérateur doit introduire. L'équipe de vérification peut résoudre les cas non couverts par les règles détaillées et modifier le cas échéant lesdites règles (mais au risque d'incohérences parmi les données (selon que leur introduction a précédé ou suivi la modification des règles).

666. Les instructions de saut de champ jouent un rôle important aussi bien dans le cadre d'une introduction continue que dans celui d'une saisie discontinue. En effet, les opérateurs de saisie doivent être au courant et connaître parfaitement toutes les instructions de saut en vigueur. Tel qu'indiqué plus haut, celles-ci peuvent accélérer l'introduction de données mais généralement au prix d'une certaine perte de qualité. En ce qui concerne la saisie discontinue, une règle empirique généralement observée veut que la qualité soit d'autant plus élevée que les sauts sont moins nombreux.

B. — Expérimentation des instructions de saisie

667. Une fois les instructions de saisie définies, les bureaux nationaux de recensement/statistique doivent demander aux opérateurs de saisie d'expérimenter les instructions

avant de décider des modalités d'exécution des tâches, notamment d'utiliser ou non la méthode de saisie dite interrompue. L'expérimentation des instructions de saisie permet d'éliminer les erreurs du système et d'optimiser la procédure de saisie.

C. — Contrôle

668. Le bureau national de recensement/statistique doit décider en outre du niveau de contrôle approprié. Pour beaucoup d'experts, il convient de procéder à un contrôle systématique; en pareil cas, les réponses sont toutes réintroduites afin de vérifier que les données recueillies sont celles qui sont introduites dans la machine en vue de leur traitement informatique. Souvent toutefois, la vérification intégrale n'est pas réalisable, soit parce que le pays ne dispose pas du temps nécessaire pour réintroduire toutes les données, soit parce qu'il ne dispose pas des ressources financières ou humaines requises. De plus, le test fait apparaître un taux d'erreur d'introduction des données très faible, les erreurs imputables aux opérateurs étant très rares, un contrôle complet est vraisemblablement inutile.

1. CONTRÔLE PAR COMPARAISON

669. Les techniques de vérification procèdent par comparaison de façon indépendante. Dans le premier cas, les opérateurs introduisent les données sur celles précédemment introduites par d'autres agents. Lorsque les caractères introduits au clavier sont différents le progiciel informe l'opérateur et selon le programme, soit les données précédentes sont écrasées, soit la différence est notifiée à l'opérateur. Puisque les données sont introduites sur la base des questionnaires originaux, en règle générale l'opérateur peut lui-même prendre une décision dûment informée quant à l'exactitude éventuelle des données initialement introduites.

2. CONTRÔLE INDÉPENDANT

670. Suivant la technique de vérification dite indépendante, les opérateurs reprennent depuis le début l'introduction de données; ils créent un fichier complètement indépendant sur la base des questionnaires originaux. Les deux fichiers ainsi obtenus, d'une part le jeu de données initialement introduites et le jeu de données de contrôle, sont alors comparés au moyen d'un programme informatique afin de vérifier les différences. On peut imaginer que la rectification des caractères non valides ou incohérents exige quelques interventions manuelles.

ANNEXE IV

Exemple d'organigrammes

671. L'une des tâches de l'équipe de vérification consiste à structurer les relations entre les variables utilisées lors du processus de vérification. La mise au point d'organigrammes facilite l'identification des différents liens entre les variables, ainsi que l'élaboration de spécifications de vérification claires et précises. Ces spécifications quant au lien entre les variables aident aussi bien les spécialistes du sujet que les informaticiens à visualiser le processus de vérification, et facilitent les échanges entre les deux groupes.

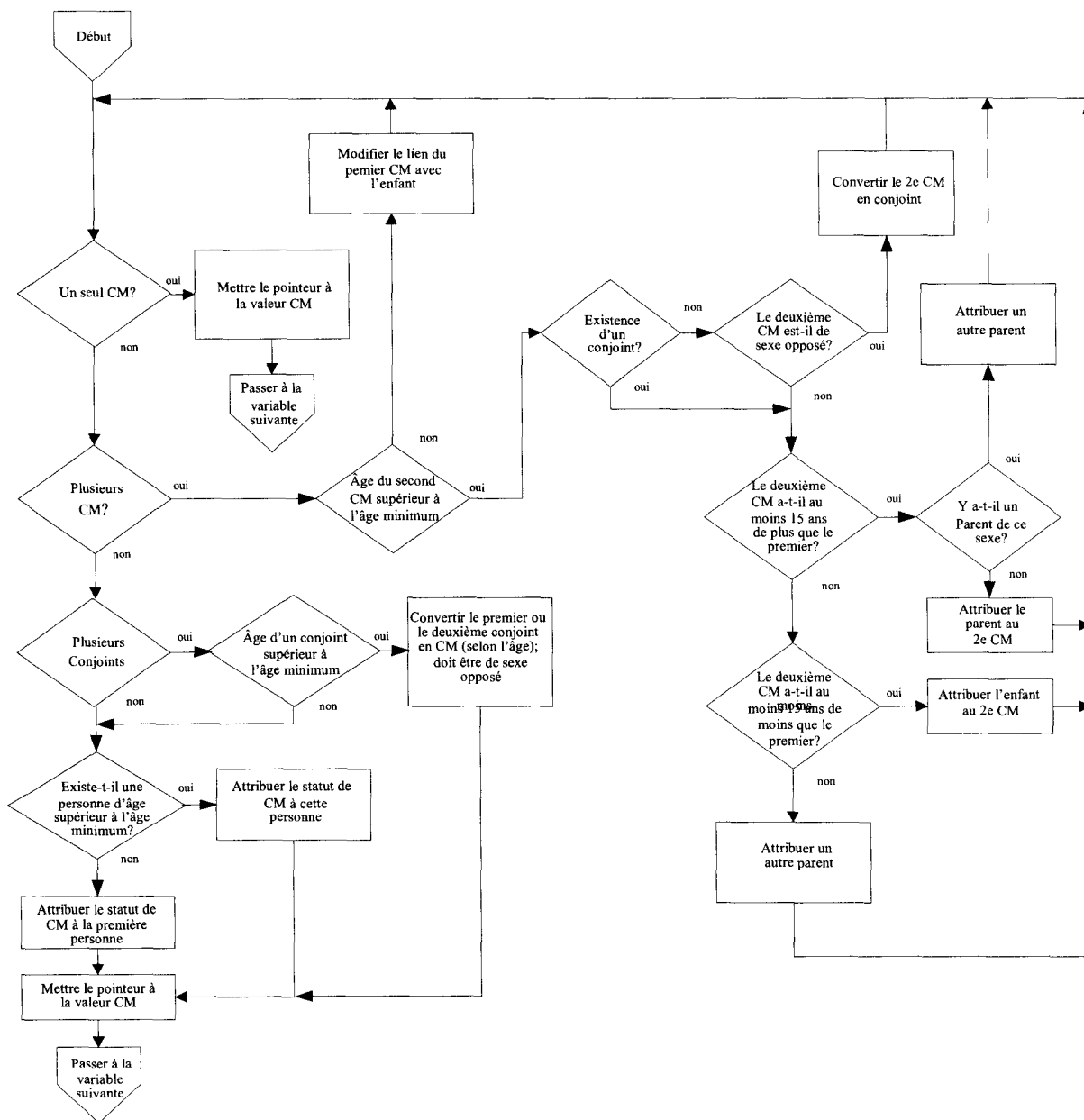
672. On trouvera ci-dessous trois exemples d'organigrammes :

- a) Organigramme de détermination du chef de ménage;
- b) Organigramme de détermination de la présence d'un conjoint dans le ménage;
- c) Organigramme de vérification de la variable « sexe » du chef de ménage et de son conjoint.

Ces organigrammes sont présentés uniquement à titre d'exemples et doivent être considérés comme tels. L'équipe de vérification peut leur apporter les modifications nécessaires en fonction du contexte national.

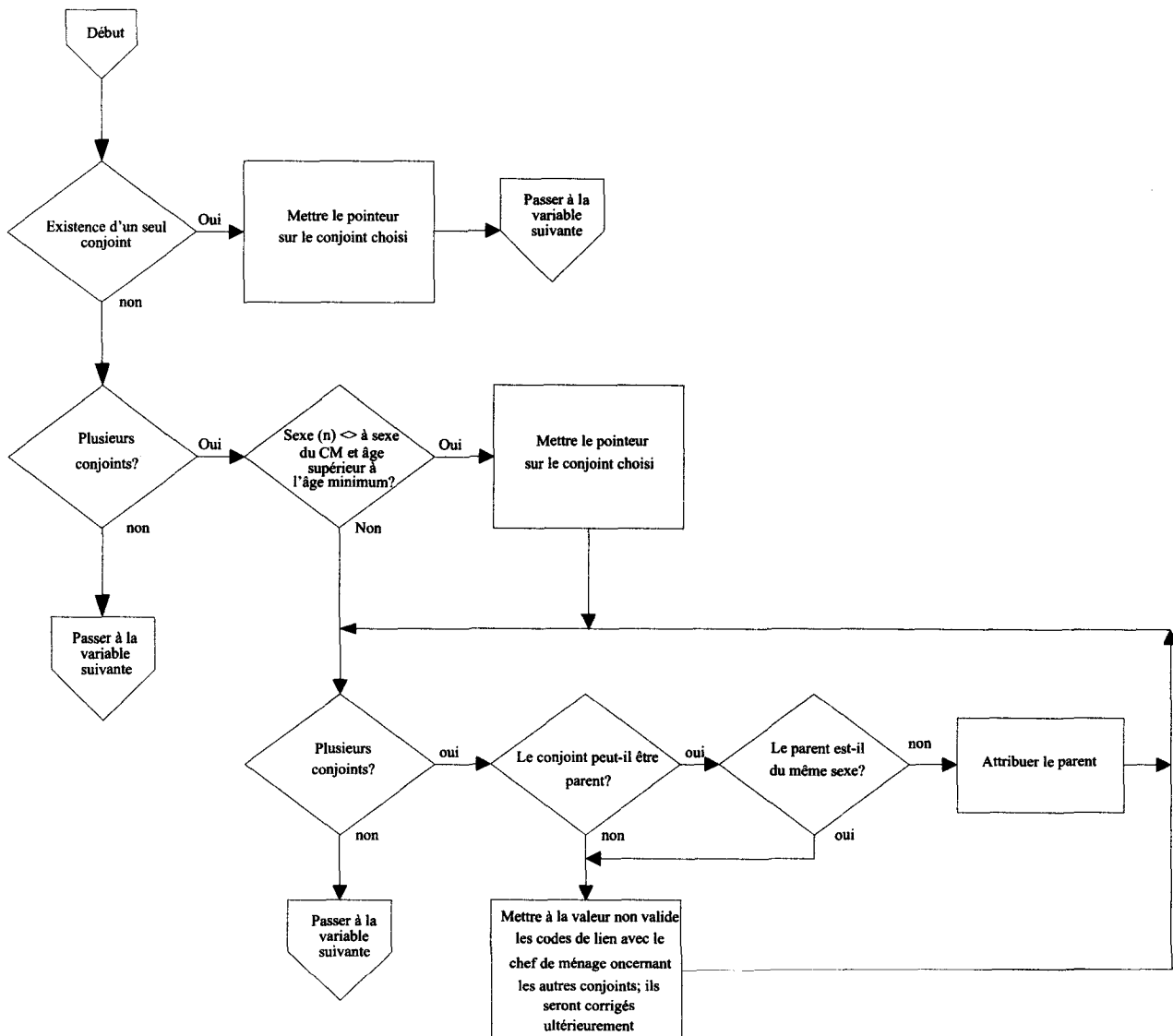
673. Chacune des variables prises en compte dans un recensement doit faire l'objet d'un organigramme de vérification. L'équipe de vérification doit s'employer à mettre au point les organigrammes et les informaticiens doivent les utiliser avec les spécifications pour élaborer les programmes de vérification des données de recensement. Organigrammes et spécifications doivent faire l'objet d'une documentation adéquate pour pouvoir être utilisés lors du traitement de données des enquêtes et recensements futurs.

FIGURE A.IV.1. Exemple d'organigramme de détermination du chef de ménage (CM)



*Âge minimum à spécifier par l'équipe de vérification.

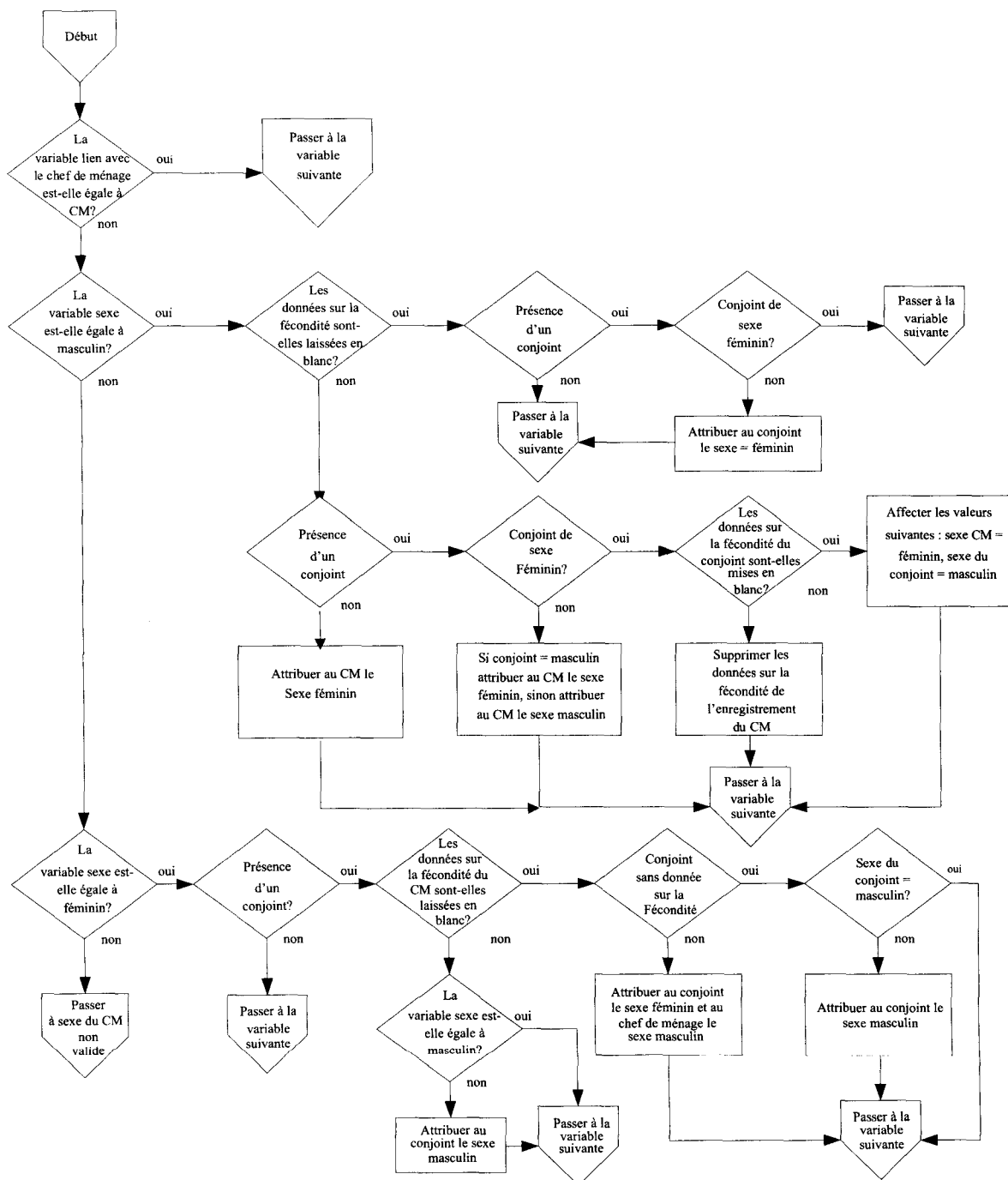
FIGURE A.IV.2. Exemple d'organigramme de détermination de la présence d'un conjoint dans le ménage



Note: CM = Chef de ménage

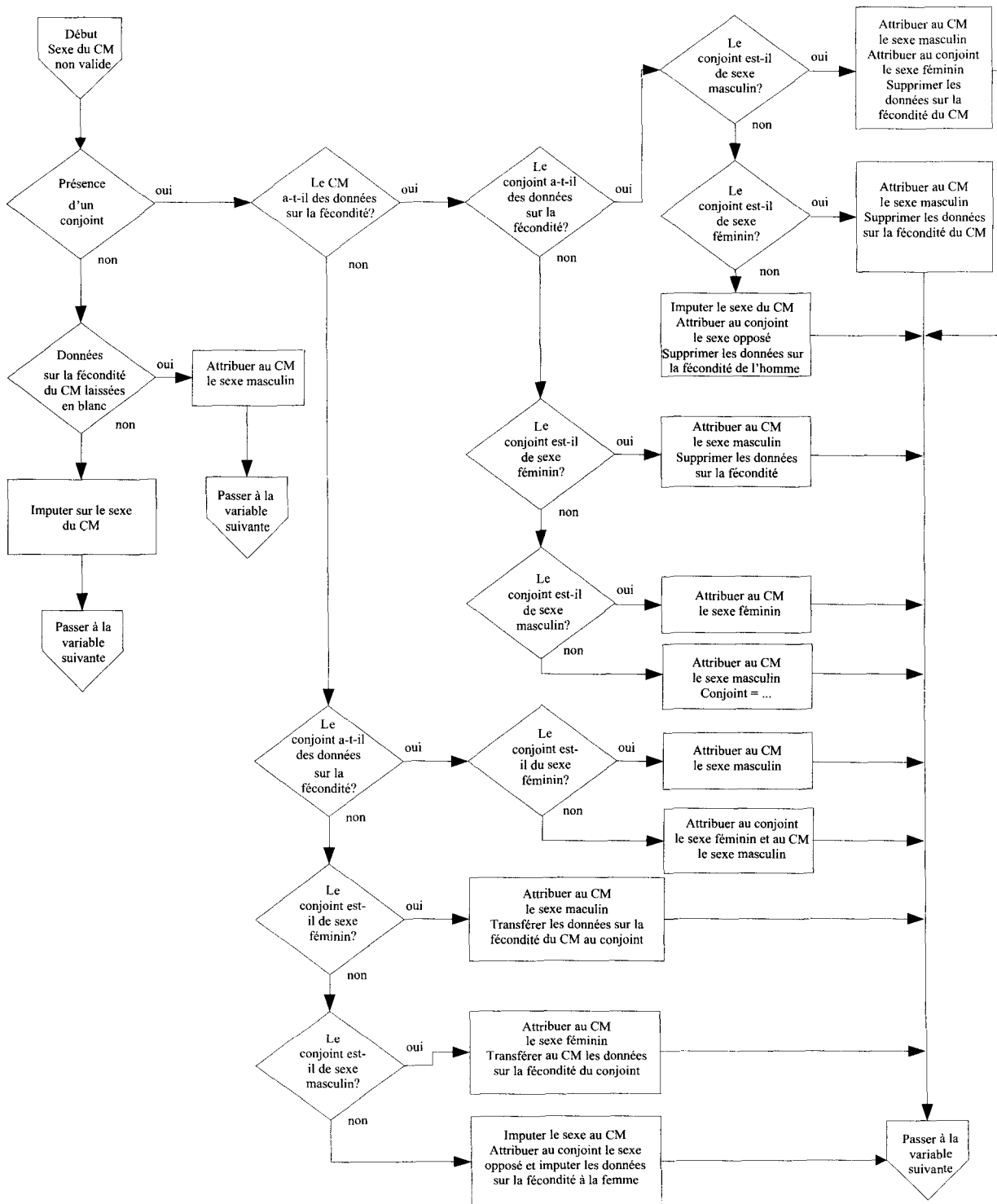
* Âge minimum à spécifier par l'équipe de vérification.

FIGURE A.IV.3. Exemple d'organigramme de vérification de la variable sexe concernant le chef de ménage et le conjoint



Note : CM = chef de ménage.

FIGURE A.IV.3 (suite)





ANNEXE V

Méthodes d'imputation

674. Un certain nombre de méthodes d'imputation ont été mises au point. La plupart des méthodes décrites ci-dessous sont analysées dans les publications de Kalton et Kasprzyk (1982, 1986); Sande (1982) et Särndal, Swensson et Wretman (1992).

675. Les méthodes d'imputation peuvent être qualifiées de stochastiques ou de déterministes, en fonction du caractère aléatoire des données imputées.

676. **Les méthodes d'imputation déterministes** comprennent l'imputation déductive, les méthodes d'imputation par modélisation, par exemple par la moyenne et par régression et (le cas échéant) par voisin le plus proche.

677. **L'imputation déductive** est une méthode qui permet de déduire avec certitude une donnée dont la valeur est manquante, ou incohérente. Ce type d'imputation sera souvent réalisé en fonction de l'ensemble des réponses fournies au titre des autres rubriques du questionnaire.

678. Plus généralement, la technique d'imputation doit remplacer la valeur entachée d'incertitude par la valeur exacte. Les paragraphes suivants présentent un certain nombre de procédures d'imputation d'usage courant.

679. A l'exception des algorithmes d'imputation dynamique à partir d'un donneur unique, les méthodes décrites ci-après concernent les procédures d'imputation rubrique par rubrique. Ainsi, à l'intérieur de chaque classe d'imputation, les rubriques de l'enregistrement sont considérées l'une après l'autre, de manière séquentielle. En général, cela consiste à considérer uniquement les vérifications concernant explicitement la rubrique en question ou un ensemble restreint de variables étroitement liées. Compte tenu de l'existence possible de vérifications explicites ou implicites associant la (les) rubrique(s) en question à d'autres rubriques devant être examinées ultérieurement au cours du processus, une valeur imputée peut éventuellement (tout en satisfaisant aux vérifications en cours), entraîner des rejets à l'occasion d'autres vérifications ultérieures. Les valeurs imputées ne pourront de façon certaine franchir avec succès la totalité des vérifications que si l'on envisage la réalisation d'une série complète de vérifications, y compris de l'ensemble des vérifications implicites. On entend par vérification implicite une vérification pouvant être déduite par combinaison logique d'au moins deux vérifications explicites.

680. Dans les descriptions ci-dessous l'expression « enregistrements acceptés » désigne les enregistrements qui ont subi avec succès toutes les vérifications concernant la (les) rubrique (s) en question. Les « enregistrements rejetés » désignent ceux qui ont été rejetés après au moins une vérification concernant la ou les rubriques en question.

681. **L'imputation de la moyenne générale** a pour effet d'attribuer la valeur moyenne de la rubrique (calculée pour les enregistrements acceptés), à la réponse manquante ou incohérente de tous les enregistrements rejetés. Cette méthode est susceptible de produire des estimations ponctuelles fondées, mais s'avère moins intéressante lorsque le calcul de la variance doit utiliser un estimateur standard. Les estimations de variance risquent d'être fortement sous-estimées sauf lorsque le taux d'imputation est très faible ou en cas d'utilisation d'un estimateur de variance modifié tenant compte de l'imputation.

682. **Imputation de la moyenne de classe** : cette méthode utilise les classes d'imputation définies pour créer des groupes d'enregistrement présentant un certain degré de similitude. A l'intérieur de chaque classe, la moyenne de la rubrique relative aux enregistrements acceptés est imputée aux réponses manquantes ou incohérentes de chacun des enregistrements rejetés. Bien que cette technique soit très proche de celle de l'imputation de la moyenne générale, l'incidence sur la distribution et les problèmes liés à l'estimation de la variance sont alors généralement nettement moindres.

683. **L'imputation par régression** ou plus généralement par **modèle** utilise les données des enregistrements acceptés, pour calculer selon une loi de régression la variable devant faire l'objet d'une imputation, d'après une série de variables prédictives. Les variables prédictives de la régression peuvent figurer parmi les rubriques du questionnaire ou peuvent être des variables auxiliaires. La loi de régression permet ensuite d'imputer les valeurs manquantes ou incohérentes. Il s'agit d'un cas particulier de la technique d'imputation par modélisation. Cette méthode est souvent utilisée pour les variables continues des enquêtes économiques, caractérisées par la possibilité souvent observée de prévoir de façon satisfaisante les valeurs occasionnelles du moment d'après les valeurs occasionnelles.

684. La **méthode d'imputation par voisin le plus proche** ou **d'appariement par une fonction de distance** a pour effet d'attribuer une valeur rejetée par une vérification, d'après celle de l'enregistrement accepté « le plus proche », cette caractéristique de proximité étant définie au moyen d'une fonction de distance, sur la base d'autres variables connues. Il s'agit d'une méthode applicable à l'intérieur de classes d'imputation. Elle est généralement considérée comme adaptée aux variables continues, mais peut également être appliquée aux variables non numériques.

685. **Les méthodes d'imputation stochastiques** incluent les méthodes fondées sur une loi de régression ou toute autre méthode déterministe avec addition de résidus

aléatoires, ainsi que les méthodes d'imputation dites dynamiques (*hot deck*) ou statiques (*cold deck*).

686. A chaque méthode déterministe correspond une méthode stochastique. Il suffit pour cela d'ajouter à la valeur obtenue par imputation déterministe un résidu aléatoire provenant d'une distribution appropriée. Cette façon de procéder contribuera à mieux conserver la distribution de fréquences des données du fichier. Kalton et Kasprzyk (1986) passent en revue différentes versions de cette technique.

687. **Les méthodes d'imputation dynamique et statique [dites à « donne » variable (*hot deck*) et à « donne » fixe (*cold deck*)]** s'emploient à créer une variabilité plus réaliste des valeurs imputées par comparaison aux résultats obtenus avec les méthodes déterministes. Les méthodes d'imputation dynamiques remplacent les valeurs manquantes ou incohérentes par des valeurs sélectionnées (de façon aléatoire) dans les enregistrements de l'enquête ou du recensement en cours qui ont subi avec succès les vérifications. Quant aux méthodes d'imputation statique, elles imputent des valeurs obtenues à partir de sources différentes, souvent des données historiques recueillies dans le cadre d'enquêtes ou de recensement antérieurs, du même type. Il existe différentes formes d'imputation dynamique ou statique.

688. **L'imputation globale aléatoire** est la forme la plus simple d'imputation dynamique (*hot deck*). Pour chaque enregistrement rejeté, un enregistrement accepté est choisi au hasard dans l'ensemble des enregistrements acceptés et la valeur indiquée correspondant à la rubrique en question est imputée à l'enregistrement rejeté.

689. **L'imputation aléatoire à l'intérieur de classes** utilise également des classes d'imputation afin de limiter la sélection aléatoire de l'enregistrement donneur à un ensemble censé présenter des similitudes par rapport à l'enregistrement devant faire l'objet d'une imputation.

690. **La technique d'imputation dynamique séquentielle** fait également appel aux classes d'imputation et présente l'avantage d'exiger seulement une seule passe dans le fichier de données pour achever le processus d'imputation. La procédure commence par l'attribution statique d'une valeur à chaque classe d'imputation, les enregistrements du fichier de données étant examinés successivement. Lorsqu'un enregistrement vérifié accepté est détecté, sa valeur pour la rubrique en question remplace la valeur enregistrée pour la classe d'imputation. Lorsqu'un enregistrement rejeté par la vérification est détecté, sa valeur manquante ou incohérente est remplacée par la valeur enregistrée. Le nombre de classes d'imputation ne peut être très élevé, puisqu'il faut être assuré de la disponibilité de donneurs, dans chaque classe d'imputation. Si l'ordre d'enregistrement dans le fichier de données est aléatoire, cette méthode sera sensiblement équivalente à une imputation aléatoire à l'intérieur de classes. Elle a par ailleurs pour inconvénient d'impliquer dans nombre de cas une utilisation répétée des donneurs, ce qui risque d'affecter la distribution de la variable considérée et les estimations de variance.

691. **Imputation dynamique hiérarchique** : il s'agit d'une amélioration de l'imputation dynamique séquentielle,

qui consiste à utiliser un nombre élevé de classes d'imputation. Lorsque l'identification d'un donneur dans la classe d'imputation initiale s'avère impossible, les classes sont réduites dans un ordre hiérarchique, jusqu'à ce qu'un donneur soit trouvé.

692. **Les algorithmes d'imputation dynamique à donneur unique** ont pour objectif d'imputer des données à un enregistrement rejeté, à partir d'un donneur unique. Ils permettent donc l'imputation conjointe de toutes les données d'un enregistrement identifié comme problématique par les vérifications. En pratique, l'objectif consiste souvent à utiliser un donneur unique pour chaque ensemble de variables étroitement liées de l'enregistrement. Cette approche présente l'avantage notable d'améliorer non seulement les distributions marginales à l'instar des méthodes ci-dessus d'imputation dynamique, mais également les distributions de fréquence conjointes (ou mixtes). Autre avantage des méthodes d'imputation dynamique à donneur unique, elles limitent le problème des valeurs d'imputation qui seront rejetées lors des vérifications envisagées pour des groupes de variables ultérieurs. Dans le cadre des méthodes d'imputation dynamique à donneur unique, on entend par « enregistrement accepté », un enregistrement qui a subi avec succès toutes les vérifications appliquées au groupe de variables. Par contre, un « enregistrement rejeté » est un enregistrement qui a été rejeté par au moins une de ces vérifications.

693. **La méthode de vérification et d'imputation de Fellegi-Holt** (Fellegi et Holt, 1976) considère simultanément toutes les vérifications. Elle se caractérise essentiellement par le fait que les règles d'imputation sont tirées des vérifications correspondantes, sans spécification explicite. Pour chaque enregistrement rejeté, la méthode prévoit initialement une phase de localisation de l'erreur, au cours de laquelle elle détermine l'ensemble minimum de variables à imputer, ainsi que les domaines de variation acceptables pour les valeurs à imputer et procède ensuite à l'imputation proprement dite. Dans la plupart des cas, un donneur unique est choisi parmi les enregistrements acceptés, en procédant par appariement, d'après les autres variables impliquées par les vérifications, mais ne devant pas faire l'objet d'une imputation. La méthode cherche à établir une correspondance exacte unique ; elle peut être étendue de façon à tenir compte des autres variables non explicitement liées aux vérifications. Parfois, aucun donneur adéquat ne peut être identifié, ce qui exige alors le recours à une méthode d'imputation par défaut.

694. **La nouvelle méthodologie d'imputation (NMI)** (Bankier, Nadeau et Newcombe, 1996; Bankier, Lachance et Poirier, 1999) s'apparente à la méthode de Fellegi-Holt dans la mesure où elle considère simultanément toutes les vérifications, ne spécifie pas explicitement les opérations d'imputation et réalise celles-ci d'après un donneur unique. Pour chaque enregistrement rejeté, la méthode identifie les imputations impliquant le moins de changements possibles, sous réserve de la disponibilité de donneurs potentiels. Ceci garantit la disponibilité d'un donneur. Contrairement à la méthode de Fellegi-Holt, la méthode NMI cherche dans un premier temps des donneurs, puis détermine les imputations

qui impliquent le moins de changements. La méthode NMI recherche des donneurs en procédant par appariement, en utilisant toutes les variables (notamment celles susceptibles d'être imputées) impliquées par les vérifications, et peut se contenter d'appariements approximatifs pour les variables numériques, complétés par des appariements exacts pour la plupart des autres variables, mais pas nécessairement pour toutes. Les imputations établies en fonction de chaque donneur potentiel sont définies, et celles qui impliquent le moins possible de modifications sont identifiées. La méthode prend également en considération les imputations impliquant des modifications à peu près minimales; celles-ci permettent parfois de réaliser des enregistrements imputés plausibles. Enfin, une imputation est choisie de façon aléatoire parmi les imputations à modifications minimales et à peu près minimales; une fois choisie, cette imputation est effectuée.

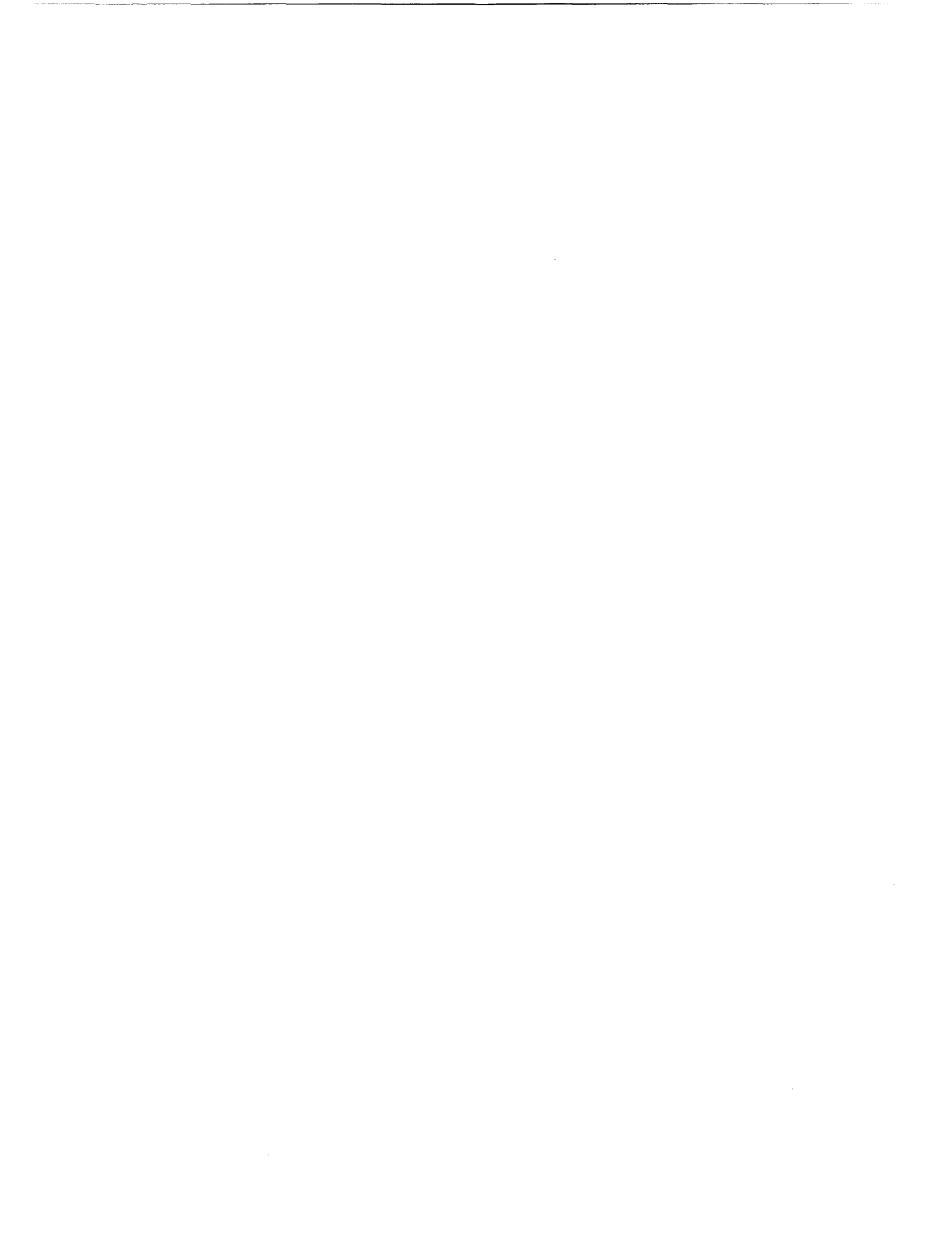
695. Bien que ces deux méthodes (Fellegi-Holt et NIM) exigent des moyens de calcul importants, il existe des algorithmes efficaces autorisant leur mise en œuvre au moyen d'ordinateurs modernes. Tel est particulièrement le cas pour la nouvelle méthodologie d'imputation qui est en mesure de traiter des problèmes de vérification et d'imputation sensiblement plus importants que la méthode Fellegi-Holt.

696. Toutes les méthodes d'imputation ci-dessus permettent d'obtenir une valeur imputée et une seule pour chaque valeur manquante ou incohérente. Elles introduisent toutes une certaine disposition dans la distribution ordinaire des variables concernées et risquent d'aboutir à des estimations de variance inadéquates en cas d'utilisation d'estima-

teurs de variance standard. Le biais introduit dépend dans une large mesure de l'importance des imputations effectuées et de la méthode employée.

697. **L'imputation multiple** est une méthode proposée par Rubin (1987) qui implique plusieurs imputations (en nombre m) pour chaque valeur à imputer. Ensuite, sur la base du jeu de données complet, m estimations peuvent être obtenues pour la variable. À partir de ces estimations, une estimation combinée unique est calculée, ainsi qu'une estimation de la variance groupée exprimant le degré d'incertitude quant au choix de la valeur à imputer. La méthode d'imputation multiple présente l'inconvénient d'exiger un travail accru de traitement des données et de calcul des estimations.

698. La plupart des systèmes d'imputation font appel à un ensemble de méthodes d'imputation; généralement, l'imputation déductive est utilisée quand cela est possible avant de faire appel à une ou plusieurs autres procédures. La plupart des bureaux nationaux de statistique emploient une méthode ou une autre d'imputation dynamique pour les vérifications et les imputations de données de recensement. La méthode d'imputation dynamique séquentielle et celle de Fellegi-Holt sont celles dont l'usage est le plus courant. Parmi les bureaux nationaux de statistiques qui utilisent actuellement la méthode de Fellegi-Holt, l'un d'entre eux lui substitue à présent la nouvelle méthode d'imputation (NMI) et plusieurs autres envisagent d'en faire autant. Toutefois, compte tenu de l'éventail escompté des lecteurs du présent manuel, cet ouvrage intéresse plus particulièrement une forme d'imputation dynamique séquentielle.



ANNEXE VI

Progiciels de vérification

699. Grâce aux micro-ordinateurs relativement bon marché actuellement disponibles, les pays devraient être en mesure de vérifier les données de leurs recensements et de leurs enquêtes en procédant de façon rigoureuse et en respectant les délais impartis. Jusqu'à une date récente, chaque pays devait rédiger son propre programme de vérification adapté à ses besoins, ce qui comportait un coûteux investissement en temps, aussi bien pour l'élimination des erreurs que pour le traitement. A la faveur des nouveaux progiciels de vérification standardisés, il est actuellement beaucoup plus facile de répondre aux besoins d'un pays en matière de vérification, et les compétences informatiques requises sont nettement moindres.

700. L'utilisation d'un progiciel de vérification comporte l'avantage suivant : à condition d'être correctement mis en œuvre, les données obtenues seront cohérentes et débarrassées des erreurs, de telle sorte que les tableaux pourront être produits rapidement. Nombre de progiciels, par exemple, les systèmes d'analyse statistique (SAS) et les progiciels SPSS de statistiques pour les sciences sociales, ou divers langages évolués peuvent servir à l'élaboration de programmes de vérification. Sinon, un pays peut utiliser un des progiciels rédigés expressément en vue de la vérification des données des recensements et des enquêtes. Pour la plupart des pays, la vérification au moyen d'un progiciel général est une solution beaucoup plus rapide, par comparaison aux programmes sur mesure, puisque le progiciel n'exigera pas le même niveau de connaissances informatiques que les programmes sur mesure.

701. Un bon progiciel de vérification garantira la communication entre les spécialistes du sujet et les programmeurs. Il doit en effet prévoir l'introduction de commentaires dans le programme proprement dit, à moins que les spécialistes du sujet aient accès ou comprennent le programme objet. Les démographes et les autres spécialistes concernés devraient être en mesure d'examiner le programme ligne par ligne et de savoir exactement quelles sont les opérations effectuées.

702. Tout progiciel de vérification existant dont un pays envisage l'utilisation doit pouvoir exécuter et produire des

comptes-rendus relatifs aux différents contrôles, tests et imputations que comporte la vérification des données du recensement. Ces exigences doivent prévaloir même lorsque les informaticiens élaborent des programmes de vérification sur mesure. Le progiciel doit répondre aux exigences suivantes :

a) Capacité de saisie et/ou de vérification des données introduites. Ce progiciel doit permettre l'introduction d'instructions de saut de champ. Par exemple, l'équipe de vérification peut décider de sauter les données sur la fécondité concernant les hommes;

b) Réalisation de vérifications structurelles, ce qui permettra d'établir la présence effective des types d'enregistrement requis, par exemple d'un enregistrement concernant l'habitation pour chaque numéro séquentiel;

c) Créer des enregistrements lorsque ces derniers sont manquants et/ou pondérer les enregistrements existants;

d) Etablir la validité de la valeur de chaque variable;

e) Enregistrer tout ou partie des enregistrements déjà vérifiés;

f) Contrôler la cohérence entre plusieurs caractéristiques du même enregistrement et d'un enregistrement à l'autre. Un sous-ensemble de ces contrôles consiste à tester la cohérence au sein des ménages, en comparant les réponses à celles des membres du ménage précédent. Imputation de valeurs par la technique d'imputation dynamique (*hot deck*), si le pays choisit d'utiliser cette méthode;

g) Utiliser plusieurs valeurs au sein d'un enregistrement ou provenant de plusieurs enregistrements, afin d'établir la valeur d'une variable calculée et introduire celle-ci dans l'enregistrement approprié;

h) Identifier et éliminer les enregistrements en double;

i) Produire un journal des erreurs et des modifications, classées par petites zones géographiques.

D'ordinaire, les progiciels ou les programmes vérifient les enregistrements les uns à la suite des autres, bien que les progiciels actuels permettent d'effectuer des contrôles d'un enregistrement à l'autre, notamment parmi les unités d'habitation.

GLOSSAIRE

Analyse rétrospective (<i>Audit trail</i>)	Méthode permettant de suivre les modifications successives des valeurs attribuées à un champ de données, ainsi que la cause de chaque modification. Les analyses rétrospectives sont généralement effectuées une fois l'interview initiale terminée.
Appariement (association) (<i>machine</i>)	Selon la procédure d'imputation dynamique (<i>hot deck</i>), opération consistant à appairer un enregistrement donneur et un enregistrement destinataire (candidat) [Pierzchala, 1995].
Appariement par fonction de distance (<i>distance function machine</i>)	A pour effet d'attribuer la valeur d'une rubrique lorsqu'un enregistrement a été rejeté par une vérification, d'après un enregistrement dûment vérifié de l'unité voisine la plus proche, la proximité étant définie au moyen d'une fonction de distance, d'après les autres variables connues.
Appariement statistique [<i>statistical machine (in hot deck)</i>]	Opération qui consiste à appairer un enregistrement donneur et un enregistrement destinataire (candidat), selon un critère statistique déterminé, afin de pouvoir transférer des données du premier au second (Pierzchala, 1995).
Cohérence interne (<i>Internal consistency</i>)	Ce terme s'applique aux relations entre les variables d'une unité donnée de l'échantillon; c'est la cohésion interne que visent les vérifications dans la plupart des méthodes d'enquête (Ford, 1983; Pierzchala).
Correction automatique (<i>Automated correction</i>)	Corrections d'erreurs sur les données sans intervention humaine. Un des éléments de la vérification automatique des données (Pierzchala, 1995).
Donne fixe (<i>cold deck</i>)	Matrice initiale d'imputation statique. Base de référence des corrections effectuées, dont les éléments sont déterminés avant le début de la correction, et qui restent inchangés tout au long de son déroulement. Ainsi la base de référence pourrait être constituée des données relatives à une année antérieure. Une méthode de « donne » fixe modifiée peut consister à ajuster les valeurs de la donne fixe en fonction des données du moment (données agrégées).
Enregistrement (<i>record</i>)	Représentation des données d'enquête stockée sur support magnétique et lisible par l'ordinateur. Généralement il existe un enregistrement par questionnaire, bien qu'il soit possible de scinder les données d'un questionnaire en plusieurs enregistrements, par exemple de population et d'habitation. (Pierzchala, 1995).
Enregistrement « pur » (dont les erreurs ont été éliminées) [<i>clean record</i>]	Enregistrement qui ne comporte aucune valeur manquante et qui a subi avec succès toutes les vérifications (Pierzchala, 1995).
Enregistrements acceptés (<i>passed edit records</i>)	Enregistrements qui ont été acceptés par toutes les vérifications concernant la rubrique en question.
Enregistrements rejetés (par la vérification) [failed edit records]	Lors des opérations de vérification et d'imputation, enregistrements rejetés par au moins une vérification concernant la (les) rubrique (s) en question.
Erreurs de qualité (<i>quality errors</i>)	Erreurs susceptibles d'affecter la qualité des données : par exemple, erreur systématique introduisant un biais (Granquist 1984; Pierzchala, 1995).
Fonction de distance (<i>distance function</i>)	Dans le cas de données numériques, fonction définie sur les variables homologues des enregistrements destinataires (candidats) et donneurs permettant de quantifier la notion de similitude. Cette fonction permet

	d'identifier des enregistrements appariés dans le cadre de la méthode d'imputation dynamique (<i>hot deck</i>) [Pierzchala, 1995].
Imputation (<i>imputation</i>)	Attribution d'une valeur à un champ de données, soit en cas de non réponse, soit pour remplacer une valeur enregistrée dont l'incompatibilité avec une série de vérifications a été établie (Pierzchala, 1995).
Imputation « hot deck (à donne variable) [<i>hot deck imputation</i>]	Méthode d'imputation selon laquelle les enregistrements donneurs proviennent de la donne momentanée de donnée de l'échantillon (la méthode <i>cold deck</i> (à donne fixe) désigne par contre la méthode d'imputation selon laquelle l'enregistrement donneur provient de données d'enquête antérieures) [Pierzchala, 1995].
Imputation aléatoire à l'intérieur de classes (<i>random imputation within classes</i>)	Usage de classes d'imputation afin de limiter la sélection aléatoire de l'enregistrement donneur à un ensemble censé présenter une certaine similitude par rapport à l'enregistrement devant faire l'objet d'une imputation.
Imputation déductive (<i>deductive imputation</i>)	Méthode selon laquelle une valeur manquante ou incohérente peut être établie avec certitude, souvent d'après les caractéristiques des réponses fournies à d'autres rubriques du questionnaire.
Imputation déterministe (<i>deterministic imputation</i>)	Cette situation est observée lorsque la valeur d'un seul champ permet à l'enregistrement de subir avec succès toutes les vérifications. Ce produit dans certains cas (par exemple, lorsque la somme calculée des éléments d'un total n'est pas égale au total proprement dit). Il s'agit de la première solution à explorer en cas de vérification automatique et d'imputation des données bancaires (Pierzchala, 1995).
Imputation dynamique par donneur unique (<i>single donor hot deck imputation</i>)	Procédé ayant pour effet d'imputer des données à partir d'un donneur unique, pour un enregistrement rejeté, permettant d'imputer conjointement les valeurs de toutes les rubriques dans le cas d'un enregistrement identifié comme problématique par les vérifications.
Imputation dynamique séquentielle (<i>sequential hot-deck imputation</i>)	Procédé d'imputation utilisé en cas de vérification séquentielle d'une série de variables, lorsque seules les valeurs vérifiées sont utilisées par la suite comme variables d'imputation dynamique.
Imputation globale aléatoire (<i>random overall imputation</i>)	Pour chaque enregistrement rejeté par la vérification, un enregistrement accepté et choisi au hasard dans l'ensemble d'enregistrements acceptés et la valeur déclarée qu'il contient pour la rubrique en question est imputée à l'enregistrement rejeté par la vérification.
Imputation multiple (<i>multiple imputation</i>)	Effectuer à plusieurs reprises une imputation pour chaque valeur devant être calculée de cette façon et déterminer ensuite une estimation de la valeur à affecter à la rubrique.
Imputation par donneur (<i>donor imputation</i>)	Méthode ayant pour effet d'apparier chacun des enregistrements devant faire l'objet d'une imputation, qu'il s'agisse de l'enregistrement destinataire ou candidat, avec un enregistrement issu d'une population « donneuse » définie, par exemple dans le cas de la méthode d'imputation dynamique (<i>hot deck</i>) [Pierzchala, 1995].
Imputation par la moyenne de classe (<i>class mean imputation</i>)	Méthode qui utilise les classes d'imputation définies pour créer des groupes d'enregistrements présentant une certaine similitude.
Imputation par la moyenne générale (<i>overall mean imputation</i>)	Attribue la moyenne de la rubrique concernant les enregistrements non rejetés par les vérifications, à la valeur manquante ou incohérente de tous les enregistrements rejetés par la vérification.
Imputation par le voisin le plus proche (<i>nearest neighbour imputation</i>)	Attribue la valeur d'une rubrique dans un enregistrement rejeté à partir de l'enregistrement accepté du « voisin le plus proche », la notion de proximité étant définie au moyen d'une fonction de distance calculée d'après les autres variables connues.

Imputation par modèle (<i>model-based imputation</i>)	Utilisation des données des enregistrements acceptés par la vérification de façon à calculer par régression la variable devant faire l'objet d'une imputation sur la base d'un ensemble de variables prédictives.
Imputation par régression (<i>regression imputation</i>)	Technique utilisant les données des enregistrements acceptés par la vérification afin de calculer par régression la variable devant faire l'objet d'une imputation à partir d'un ensemble de variables prédictives.
Imputation statistique (<i>statistical imputation</i>)	On pourrait choisir comme exemple d'imputation statistique l'utilisation d'une loi de régression selon laquelle la variable dépendante doit être attribuée aux variables indépendantes obtenues à partir des réponses présumées valides (Pierzchala, 1995).
Imputation structurelle (<i>structural imputation</i>)	On procède par imputation structurelle lorsqu'une relation structurelle reste valable entre plusieurs variables. Par exemple, une valeur totale doit être égale à la somme des valeurs partielles (des éléments). Par conséquent, dans le cas d'une mère le nombre d'enfants doit être égal au nombre d'enfants vivants augmenté du nombre d'enfants décédés (Pierzchala 1995).
Indicateur (<i>flag</i>)	Un indicateur est une variable servant à noter des informations utiles concernant une ou plusieurs autres variables. Par exemple, en cas de modification d'une rubrique devenue valide après avoir été non valide, l'indicateur peut servir soit à enregistrer l'information d'origine soit à noter simplement la modification de la valeur de la rubrique.
Introduction de données discontinue (<i>heads up data entry</i>)	Mode d'introduction des données selon lequel l'équipement d'introduction des données détecte les erreurs qu'elles contiennent au fur et à mesure, ce qui permet à l'opérateur de les corriger immédiatement (Pierzchala, 1995). Voir « saisie interactive ».
Introduction de données ininterrompue (<i>heads down data entry</i>)	Mode d'introduction des données selon lequel la machine dans laquelle les données sont introduites ne détecte pas les erreurs au fur et à mesure de l'introduction des dites données, ce qui permet à l'opérateur de les introduire rapidement et sans difficulté.
Jeu minimal (<i>minimal set</i>)	Champ de données devant faire l'objet d'une imputation de façon à garantir que toutes les vérifications ont été appliquées avec succès (Pierzchala, 1995).
Langage symbolique (<i>pseudocode</i>)	Instructions ou spécifications de vérification écrites.
Liste de codage (<i>code list</i>)	Liste de toutes les valeurs admises (acceptables) susceptibles d'être attribuées à une rubrique.
Macro-vérification(<i>macro-edit</i>)	Détection d'erreurs particulières à l'issu (1) de contrôles effectués sur des données agrégées ou (2) de contrôles effectués sur l'ensemble des enregistrements. Les vérifications sont fondées sur les estimations (Granquist, 1987; Pierzchala, 1995).
Méthode Fellegi-Holt de correction automatique (<i>Fellegi-Holt methode for automatic correction</i>)	Méthode de correction automatique ayant pour effet de modifier le plus petit nombre possible de rubriques de données; le modèle de Fellegi-Holt sert à déterminer des jeux acceptables de valeurs ou des domaines de variation pour les rubriques à imputer. Il est possible de procéder par imputation séquentielle ou simultanée, par la méthode <i>cold deck</i> ou <i>hot deck</i> .
Micro-vérification (<i>micro-edit</i>)	Vérification traditionnelle effectuée sur des données effectuées au niveau de l'enregistrement. Antonyme logique des macro-vérifications (Pierzchala, 1995).
Micro-macro vérification (<i>micro-macro edit</i>)	Procédure de vérification selon laquelle des micro-vérifications détaillées sont remplacées par une combinaison de micro-vérifications et de vérifications statistiques/macro-vérifications. Les micro-vérifications effectuées selon la procédure mixte sont moins détaillées que dans le premier cas. Le principe consiste à « définir des vérifications d'enquête fon-

Nouvelle méthodologie d'imputation (<i>new imputation methodology</i>)	dées sur un principe d'« incidence sur les estimations » et non sur une formule attrape-tout de détection des incohérences de données (Granquist, différentes dates, Pierzchala, 1995).
Organigramme (<i>flow chart</i>)	Méthode semblable à la méthode de Fellegi-Holt dans la mesure où elle envisage simultanément toutes les vérifications, ne spécifie pas explicitement les opérations d'imputation et les effectue à partir d'un donneur unique. La méthode NMI examine chaque enregistrement rejeté de façon à déterminer les opérations d'imputation comportant le moins possible de modifications, sous réserve de la disponibilité de donneurs potentiels.
Passage machine ou phase d'exécution (<i>production run</i>)	Description schématique de toutes les fonctions qui doivent être exécutées.
Pointeur (<i>Pointer</i>)	Opération consistant à traiter d'importantes quantités de données après élimination des erreurs initiales contenues dans le programme de vérification ou d'exploitation.
Points aberrants (<i>outliers</i>)	Un pointeur est une variable utilisée pour repérer une rubrique ou une autre variable en vue de références ultérieures. Par exemple, on utilise des pointeurs pour noter les numéros de ligne du « chef de ménage » et du « conjoint » pour utilisation ultérieure afin de vérifier que les conjoints sont de sexe opposé et que l'un et l'autre sont mariés.
Pondération (<i>weights</i>)	Valeurs de rubriques situées en dehors de certaines limites en fonction de certaines règles adoptées pour les définir (Pierzchala, 1995).
Recherche (<i>searching</i>)	Suivant la méthode de vérification et d'imputation proposée par Fellegi-Holt, des pondérations sont attribuées aux différents champs en fonction de la fiabilité des données. Plus le coefficient de pondération est élevé et plus l'imputation de la valeur de ce champ sera probable (toutes choses étant égales par ailleurs). Des coefficients de pondération peuvent également être attribués aux vérifications (Pierzchala, 1995).
Saisie de données (<i>data capture</i>)	Selon la procédure d'imputation dynamique (<i>hot deck</i>), opération consistant à rechercher un enregistrement donneur (Pierzchala, 1995).
Saisie interactive (<i>Interactive keying</i>)	Processus consistant à introduire les données recueillies sous une forme lisible par l'ordinateur. Les contrôles de vérification élémentaires sont souvent effectués dans le cadre des sous-modules du logiciel de saisie.
Sans réponse (<i>non-response</i>)	Mode d'introduction des données selon lequel l'équipement d'introduction détecte des erreurs dans les données au fur et à mesure, permettant à l'opérateur de corriger immédiatement les erreurs (voir saisie discontinue).
Série complète de vérifications (<i>complete set of edits</i>)	Questionnaire incomplet ou questionnaire manquant (Pierzchala, 1995).
Similitude (<i>similarity</i>)	Réunion des vérifications explicites et implicites. Nécessaire à la détermination des domaines possibles de valeurs d'imputation (si l'on veut que les valeurs imputées ne soient pas rejetées par les vérifications) [Pierzchala, 1995].
Système de Fellegi-Holt (<i>Fellegi-Holt system</i>)	Dans le cas de données numériques, notion de proximité de deux enregistrements en fonction des variables d'appariement prescrites. Une fonction de distance permet de quantifier cette notion d'après certains critères (Pierzchala, 1995).
	Se rapporte aux hypothèses et aux objectifs de vérification et d'imputation définis par Fellegi et Holt dans leur communication de 1976 au <i>Journal of the American Statistical Association</i> . Le modèle de Fellegi-Holt a pour caractéristique essentielle de mettre en évidence la nécessité de recourir aux vérifications implicites pour garantir que la présence d'une série de valeurs dans les champs de données qui ne sont pas imputés conduit toujours à l'obtention d'enregistrements définitifs (imputés) qui

	subissent avec succès toutes les vérifications.
Tableau ou rangée (<i>array</i>)	Ensemble de valeurs numériques. Appelé parfois « matrice », un tableau peut servir à enregistrer des données numériques à caractère répétitif.
Variables d'appariement (<i>machine variables</i>)	Variables utilisées pour établir un appariement entre un enregistrement destinataire (candidat) et un enregistrement donneur (Pierzchala, 1995).
Vérification (1 ^{re} définition) [<i>edit (definition 1)</i>]	Relations logiques auxquelles sont assujetties les valeurs que peuvent prendre chaque variable (Pierzchala, 1995).
Vérification (2 ^e définition) [<i>edit (definition 2)</i>]	Application de règles mettant en évidence des combinaisons interdites de réponses (Pierzchala, 1995).
Vérification à plusieurs variables (<i>multivariate edit</i>)	Type de vérifications statistiques ayant pour effet d'évaluer les données et d'identifier les points aberrants au moyen de distribution à plusieurs variables (Pierzchala, 1995).
Vérification de cohérence (<i>consistency edit</i>)	Vérification de relations fondamentales, par exemple l'égalité entre un total et la somme des parties ou encore le fait que les « superficies récoltées » soient toujours inférieures aux « superficies plantées » (Pierzchala, 1995).
Vérification de validation (<i>validation edit</i>)	Contrôle des vérifications effectué entre les champs d'un enregistrement particulier. Cela inclut le contrôle de chaque champ, de chaque enregistrement afin de vérifier qu'il contient une information ou une donnée valide et le contrôle des informations figurant dans une combinaison prédéterminée de champs afin d'établir la cohérence des données ainsi enregistrées (Pierzchala 1995).
Vérification déterministe (<i>deterministic edit</i>)	En cas de rejet cette vérification met en évidence une erreur dont la probabilité est égale à un. Exemple : âge = 5 et état matrimonial = mère. Par opposition aux vérifications stochastique (Pierzchala, 1995).
Vérifications explicites (<i>explicit edits</i>)	Vérification explicitement définie par un spécialiste du sujet (on définit des vérifications explicites par opposition aux vérifications implicites) (Pierzchala, 1995).
Vérification implicite (<i>implied edit</i>)	Vérification non formulée déduite logiquement des vérifications explicites élaborées par un spécialiste du sujet (Pierzchala, 1995).
Vérification inter-enregistrements (<i>between-record edit</i>)	Vérification effectuée sur des champs impliquant plusieurs enregistrements du jeu de données de l'enquête. Les vérifications statistiques sont un exemple de vérification inter-enregistrements, parce que les distributions sont établies à partir d'ensembles de champs de données issus de tous les enregistrements du jeu de données (Pierzchala, 1995).
Vérification intra-enregistrement (<i>within-record edit</i>)	Autre désignation des vérifications de validation (Pierzchala, 1995).
Vérification manuelle (<i>hand or manual edit</i>)	Vérification effectuée par des agents avant introduction des données dans l'ordinateur (voir vérification manuelle) (Pierzchala, 1995). Vérification rétrospective (<i>edit trail</i>).
Voir « analyse rétrospective ».	Vérification statistique (<i>statistical edit</i>). Série de contrôles fondés sur une analyse statistique des données de la personne interrogée : par exemple, le rapport des valeurs attribuées à deux champs est situé entre des limites déterminées par une analyse statistique de ce rapport obtenue pour des déclarants présumés valides (Greenberg et Surdi 1984; Pierzchala, 1995).
Vérification stochastique (<i>stochastic edit</i>)	Dans un enregistrement rejeté, cette vérification met en évidence une erreur dont la probabilité est inférieure à 1 (Pierzchala, 1995).
Vérification structurelle (<i>structural edit</i>)	Contrôle fondé sur l'existence d'un lien logique entre plusieurs champs vérifiés. Par exemple, une valeur totale doit être égale à la somme des élé-

ments : ou en raison d'instructions de saut de champ inhérent à un questionnaire, deux variables situées sur deux « chemins » séparés du questionnaire ne peuvent être simultanément non nuls. Une vérification structurelle garantit la préservation de la structure du questionnaire dans l'enregistrement de données (Pierzchala, 1995).

Vérifications quantitatives
(*quantitative edits*)

Vérifications appliquées à des champs de données mesurés sur une échelle continue (Pierzchala, 1995).

RÉFÉRENCES

- Banister, J. (1980). Use and abuse of census editing and imputation. *Asian and Pacific Census Forum*, vol.6, n° 3, pp.1-20.
- Bankier, M., M. Lachance et P. Poirier, (1999). A generic implementation of the new imputation methodology. Dans *Proceedings of the Section on Survey Research Methods*. Alexandria, Virginia, American Statistical Association forthcoming.
- Bankier, M., A.-M. Houle et M. Luc (n. d.). Canadian census demographic variables imputation. Manuscrit.
- Bankier, M., M. Luc, C. Nadeau, et P. Newcombe (1996). Imputing numeric and qualitative census variables simultaneously. Dans *proceedings of the Section on Survey Research Methods*. Alexandria, Virginia, American Statistical Association, pp. 287-292.
- Boucher, L. (1991). Micro-editing for the annual salary of manufacturers: what is the value added? Dans *Proceedings of the Annual Research Conference*. Washington D.C., United States Bureau of the Census, pp. 765-781.
- Commission de statistique des Nations Unies et Commission économique pour l'Europe (1994). *Statistical Data Editing*, vol. n° 1, *Methods and Techniques*. Conference of European Statisticians, Statistical Standards and Studies Series, n° 44. Numéro de vente : 94.II.E.36.
- Commission de statistique des Nations Unies et Commission économique pour l'Europe (1997). *Statistical Data Editing*, vol. 2 : *Methods and Techniques*. Conference of European Statisticians, Statistical Standards and Studies Series, n°48. Numéro de vente : 96.II.E.30.
- Fellegi, I.P, et D.Holt, (1976). A systematic approach to automatic edit and imputation. *Journal of the American Statistical Association*, vol. 71, n° 353 (March), pp. 17-35.
- Ford, Barry L. (1983). An overview of hot deck procedures. Dans *Incomplete Data in Sample Surveys*, vol. 2, *Theory and Bibliographies*. William G. Madow, Ingram Olkin et Donald B. Rubin, eds.
- Granquist, L. (1984). Data editing and its impact on the further processing of statistical data. Papier présenté dans Workshop on Statistical Computing, Budapest, 12-17 novembre 1984.
- Granquist, L. (1987). The short-term developing program for computer-supported editing at Statistics Sweden. Rapport présenté dans Data Editing Joint Group Meeting, Madrid, 22-24 avril 1987. Stockholm, Statistics Sweden.
- Granquist, L. (1997). The new view on editing. *International Statistical Review*, vol. 65, n° 3, New York, Academic Press, pp. 381-387.
- Granquist, L. et Kovar, J. G. (1997). Editing of survey data: how much is enough? Dans *Survey Measurement and Process Quality*, Lyberg et al (eds), New York, Wiley and Sons pp. 415-435.
- Greenberg, Brian, et Rita Surdi (1984). A flexible and interactive edit and imputation system for ratio edits. Dans *Proceedings of the American Statistical Association, Section on Survey Research Methods*. Alexandria, Virginia, American Statistical Association pp. 421-426.
- Kalton, G. et D. Kasprzyk (1982). Imputing for missing survey responses. Dans *Proceedings of the Section on Survey Research Method*. American Statistical Association, pp. 23-31.
- Kalton, G. et D. Kasprzyk (1986). The Treatment of Missing Survey Data. *Survey Methodology*, vol. 12, pp. 1-16.
- Naus, J.I. (1975). *Data Quality Control and Editing*. New York, Marcel Dekker.
- Nordbotten, S. (1963). Automatic editing of individual statistical observations. Conference of European Statisticians, Statistical Standards and Studies, n° 2. New York, Nations Unies.
- Organisation des Nations Unies (1992). *Handbook of Population and Housing Censuses*, part I : *Planning, Organization and Administration of Population and Housing Censuses*. Studies in Methods, Series F, n° 54. Numéro de vente : E.92.XVII.8.
- _____ (1992). *Handbook of Population and Housing Censuses*, part II : *Demographic and Social Characteristics*. Studies in Methods. Series F, n° 54. Numéro de vente : E.91.XVII.9.
- _____ (1998). *Principes et Recommandations concernant les recensements de la population et de l'habitat*, première révision. Etudes statistiques, Série M, n° 67/Rev.1. Numéro de vente : F.98.XVII.8.
- _____ (1999). *Standard Country or Area Codes for Statistical Use*, Statistical Papers, Series M, n° 49/Rev.4. Numéro de vente : M.98.XVII.9.
- Pierzchala, M. (1995). Editing systems and software. Dans *Business Survey Methods*. B.G. Cox, and others, eds. New York, John Wiley & Sons, pp. 425-411.

- Pullum, T. W., T. Harpham et N. Ozsever (1986). The machine editing of large-sample surveys: the experience of the World Fertility Survey. *International Statistical Review*, vol. 54, pp. 311-326.
- Rubin, D. B. (1987). *Multiple Imputation for Non-response in Surveys*. New York, Wiley.
- Sande, I. G. (1982). Imputation in surveys: coping with reality. *American Statistician*, vol. 36, pp.145-152.
- Särndal, C. E., B. Swensson et J. Wretman (1992). *Model Assisted Survey Sampling*. New York, Springer-Verlag.
- Statistique Canada (1998). *Lignes directrices concernant la qualité*, 3^e édition. Ottawa, Statistique Canada.





كيفية الحصول على منشورات الأمم المتحدة

يمكن الحصول على منشورات الأمم المتحدة من المكتبات ودور التوزيع في جميع أنحاء العالم . استعلم عنها من المكتبة التي تتعامل معها أو اكتب إلى : الأمم المتحدة ، قسم البيع في نيويورك أو في جنيف .

如何购取联合国出版物

联合国出版物在全世界各地的书店和经售处均有发售。请向书店询问或写信到纽约或日内瓦的联合国销售组。

HOW TO OBTAIN UNITED NATIONS PUBLICATIONS

United Nations publications may be obtained from bookstores and distributors throughout the world. Consult your bookstore or write to: United Nations, Sales Section, New York or Geneva.

COMMENT SE PROCURER LES PUBLICATIONS DES NATIONS UNIES

Les publications des Nations Unies sont en vente dans les librairies et les agences dépositaires du monde entier. Informez-vous auprès de votre libraire ou adressez-vous à : Nations Unies, Section des ventes, New York ou Genève.

КАК ПОЛУЧИТЬ ИЗДАНИЯ ОРГАНИЗАЦИИ ОБЪЕДИНЕННЫХ НАЦИЙ

Издания Организации Объединенных Наций можно купить в книжных магазинах и агентствах во всех районах мира. Наводите справки об изданиях в вашем книжном магазине или пишите по адресу: Организация Объединенных Наций, Секция по продаже изданий, Нью-Йорк или Женева.

COMO CONSEGUIR PUBLICACIONES DE LAS NACIONES UNIDAS

Las publicaciones de las Naciones Unidas están en venta en librerías y casas distribuidoras en todas partes del mundo. Consulte a su librero o diríjase a: Naciones Unidas, Sección de Ventas, Nueva York o Ginebra.
