Techniques d'enquête, décembre 1995
Vol. 21, n° 2, pp. 181–192
Statistique Canada

Concordance des données censitaires et des registres de l'état civil concernant les aînés aux États-Unis: 1970-1990

LAURA B. SHRESTHA et SAMUEL H. PRESTON[1]

RÉSUMÉ

Les lacunes et les erreurs relevées dans les données censitaires et les registres de l'état civil aux États-Unis laissent planer un doute sérieux sur la qualité des données concernant les aînés et le recensement des décès dans ce pays. Dans le présent article, nous évaluons la concordance des données provenant de ces deux sources pour les populations blanche et afro-américaine. Nous étudions en particulier les aînés (60 ans et plus), où les tendances de la mortalité ont la plus grande incidence sur les programmes sociaux et où les problèmes de données sont les plus graves. Au moyen de l'analyse intercensitaire des cohortes, nous déterminons les discordances en fonction de l'âge entre les deux sources pour deux périodes: 1970-1980 et 1980-1990. Les anomalies relevées quant à la nature de la couverture et aux erreurs de contenu sont analysées à la lumière des données de la documentation spécialisée. Les données sur la population afro-américaine montrent de très nombreuses discordances pour la période 1970-1990, qui sont vraisemblablement attribuables à une exagération de l'âge déclaré lors des recensements, par rapport aux données des registres de l'état civil. On relève également des anomalies pour la population blanche dans la période intercensitaire 1970-1980. Selon nous, ces anomalies sont principalement dues à un sous-dénombrement dans le recensement de 1970, par rapport à celui de 1980 et au recensement des décès. Par contre, les données de 1980-1990 concernant les blancs, et en particulier les femmes, sont très cohérentes, beaucoup plus même que dans la plupart des pays européens.

MOTS CLÉS: Erreurs de déclaration au sujet de l'âge; couverture; mortalité; évaluation des recensements; recensement des décès; qualité des données; croisement des courbes du taux de mortalité, États-Unis.

1. INTRODUCTION

Les méthodes classiques d'estimation de l'ampleur de la mortalité dans les pays développés utilisent des données provenant de deux sources distinctes. Les données sur les décès tirées des registres de l'état civil servent habituellement de numérateurs des taux de mortalité. Les dénominateurs correspondent pour leur part habituellement au dénombrement des personnes vivantes réalisé lors des recensements. L'exactitude des taux ainsi calculés dépend donc de la qualité des données provenant de ces deux sources.

Nous présentons ci-après les résultats d'un test de la qualité des données américaines correspondant à deux périodes intercensitaires: 1970-1980 et 1980-1990. Nous étudions en particulier la concordance des changements relevés dans la taille d'une cohorte d'un recensement à l'autre et du nombre de décès recensés pendant cette période, en procédant aux ajustements voulus pour tenir compte des migrations. Toutes les données sont exprimées en années d'âge, et des tests séparés sont réalisés pour les populations blanche et noire.

Nous nous intéressons essentiellement aux aînés (60 ans et plus), où les tendances de la mortalité ont la plus grande incidence sur les programmes sociaux (Preston 1993) et où les problèmes de données sont les plus graves. La population blanche des États-Unis semble afficher des taux de mortalité plus bas, chez les plus de 80 ans, que dans tous les autres pays industrialisés (Vaupel 1993). Si elle s'avérait valide, cette comparaison pourrait avoir une grande incidence sur l'évaluation de la qualité relative des services de santé. Cependant, la population afro-américaine présente des taux de mortalité encore plus bas que ceux de la population blanche chez les plus de 80 ans, reflétant ainsi le phénomène bien connu du croisement des courbes du taux de mortalité des deux races qui s'observe entre 75 et 85 ans. La fiabilité de ces deux ensembles de données sur la mortalité dépend bien sûr de la qualité des données. Or, on a déjà émis des doutes très sérieux sur la qualité des données concernant la population de race noire (voir à ce propos Zelnik 1969; Coale et Kisker 1990), même si la plupart des observateurs semblent admettre d'emblée la validité du croisement des tendances (Manton et coll. 1986; McCord et Freeman 1990).

Dans le cadre de leurs travaux de détermination de nouveaux profils modèles du taux de mortalité dans les pays à faible mortalité, Condran, Himes et Preston (1991) ont fait état de tests semblables de la qualité des données pour 68 périodes intercensitaires dans 18 pays industrialisés. En général, la concordance des données s'est avérée très bonne pour les cohortes de 65 ans au deuxième recensement (66 des 68 ensembles de données ont passé le test de cohérence). La concordance a cependant diminué avec l'âge: la moitié seulement des ensembles de données montraient une concordance à 85 ans et la proportion des données concordantes passait à 15%, à 95 ans (Condran et coll. 1991: tableau 7). Les États-Unis ne faisaient pas partie des pays qui ont fait l'objet de ces tests parce que leurs données publiées sur la mortalité ne sont pas classées

[1] Laura B. Shrestha, The World Bank, Human Development Department, 1818 H Street, N.W., Washington, DC 20433, U.S.A.; Samuel H. Preston, Population Studies Center, University of Pennsylvania, 3718 Locust Walk, Philadelphia, PA 19104, U.S.A.

par années d'âge. Nous sommes maintenant en mesure de combler cette grave lacune puisque nous avons traité les bandes de données pour chacun des décès enregistrés aux États-Unis de 1970 à 1988. (Les chiffres pour 1989 (année complète) et pour 1990 (janvier à mars seulement) ont été estimés à partir des données par groupes d'âge publiées par le National Center for Health Statistics et des données de l'année 1988. Voir l'annexe A pour plus de détails.) Ces bandes sont produites par le National Center for Health Statistics (NCHS) et sont diffusées par le Inter-University Consortium for Political and Social Research (ICPSR). Pour les années que nous avons examinées, elles portent sur un nombre approximatif de 50 millions de décès.

2. POPULATIONS ET DONNÉES

2.1 Contexte général

Trois sources principales de données ont été utilisées: 1) recensements nationaux du U.S. Bureau of the Census pour les années 1970, 1980 et 1990; 2) données annuelles sur les décès compilées par le NCHS; 3) estimations inédites sur l'immigration nette obtenues auprès du U.S. Bureau of the Census. La description détaillée de nos sources de données est jointe au présent article dont elle constitue l'annexe A. Il nous paraît cependant utile de décrire ci-après brièvement ces données et les ajustements importants que nous leur avons apportés.

2.2 Recensements

Nous utilisons des données censitaires classées en fonction de la race (noire/blanche) et du sexe, et par années d'âge (avec une classe ouverte pour les 100 ans et plus). Les compilations englobent l'ensemble des résidents des 50 états et du district de Columbia; elles comprennent les personnes placées en établissements, les Américains en déplacement temporaire à l'étranger et les étrangers dont la résidence habituelle (légale ou non) est située aux États-Unis (sauf les militaires et le personnel diplomatique). Sont par contre exclus les Américains qui résident à l'étranger pour une période prolongée et les ressortissants étrangers en visite temporaire aux États-Unis. Les statistiques officielles ne sont pas ajustées pour tenir compte du sous-dénombrement (p. ex., résidents légaux et immigrants sans document non recensés).

Le choix de l'expression "population résidente" signifie que les tableaux de recensement comprennent à la fois les résidents légaux et les immigrants sans document. Sur l'ensemble des personnes sans document qui résidaient aux États-Unis à l'époque du recensement de 1970, il semble qu'une proportion négligeable ait été dénombrée. Ainsi, la population des résidents légaux était approximativement comparable à la population résidente totale lors du recensement de 1970. Par contre, lors du recensement de 1980, le U.S. Bureau of the Census a estimé que, pour la toute première fois, un nombre significatif de personnes sans document avaient été dénombrées. Cette population a été évaluée à 2.06 millions de personnes. De ce nombre, dans le groupe d'âge des 60 ans et plus, on a compté 10,000 hommes et 19,000 femmes de race blanche, et 3,000 hommes et 6,000 femmes de race noire (U.S. Bureau of the Census 1988).

On reconnaît que les tableaux officiels du recensement de 1970 contiennent des erreurs parmi lesquelles la plus évidente est le surdénombrement marqué du nombre de personnes âgées de 100 ans ou plus. Même si le recensement a dénombré 106,000 personnes appartenant à ce groupe d'âge, les estimations démographiques indirectes laisseraient plutôt conclure à un nombre oscillant entre 3,000 et 8,000 personnes, le chiffre estimatif le plus plausible étant fixé à 4,800 (Siegel 1974; Siegel et Passel 1976; U.S. Bureau of the Census 1974). Nous utilisons les données inédites du U.S. Bureau of the Census pour le recensement de 1970, où cette surestimation des centenaires a été corrigée. Cette utilisation des estimations corrigées est motivée par deux raisons: premièrement, à défaut d'une correction, l'excédent serait suffisamment important pour biaiser les résultats pour les groupes plus âgés. Deuxièmement, il semble que cette surestimation n'ait pas été due à des déclarations erronées de l'âge par les intéressés, mais plutôt à un malentendu qui a porté certaines personnes à confondre les colonnes destinées à l'inscription du mois et de l'année de la naissance (Siegel et Passel 1976).

Au cours des recensements de 1980 et de 1990, un grand nombre de personnes ont choisi de fournir une réponse de leur cru à la question concernant la race, au lieu de choisir une des catégories préétablies. Sur la population totale, 6.8 millions de personnes, en majeure partie d'origine hispanique, ont choisi cette option en 1980, et 9.3 millions on fait de même en 1990. Les tableaux officiels de ces recensements ne sont donc pas directement comparables à d'autres sources de données puisqu'ils sont seuls à prévoir une catégorie raciale résiduelle. Pour autoriser une comparaison avec d'autres sources de données, le Census Bureau a modifié les données censitaires de 1980 et de 1990 pour qu'elles soient conformes aux catégories historiques de groupes raciaux. Le Bureau a en outre procédé, en 1990, à une "correction" du problème lié à l'âge (pour plus de détails, voir Word et Spencer 1991). Nous avons décidé d'utiliser les statistiques du Bureau de 1980 et de 1990 modifiées pour la race aux fins de notre étude. Ce choix a été motivé par le nombre considérable de personnes qui auraient été exclues autrement, en particulier dans la population blanche.

2.3 Registres des décès

Les registres américains des décès portent sur l'ensemble des décès recensés dans les 50 états et dans le district de Columbia, classés selon la race et le sexe et par années d'âge (jusqu'à 125 +). Pour permettre une comparaison avec les données des recensements, les décès des non résidents (ressortissants étrangers et citoyens américains résidant à l'étranger) ont été exclus.

Aucune correction n'est faite pour tenir compte du sous-dénombrement des décès ou des décès dont la description sur le certificat comporte des erreurs. Nous avons relevé deux problèmes qui influent sur l'utilisation de notre

méthode d'analyse intercensitaire. La période intercensitaire s'étale du 1er avril au 31 mars, alors que les registres des décès utilisent l'année civile. En outre, les registres des décès et les recensements américains utilisent l'âge au dernier anniversaire au lieu de l'année de la naissance. Nous avons corrigé les deux problèmes en répartissant les décès sur des graphes triangulés du temps et de l'âge correspondant aux "années de recensement" débutant le 1er avril. Par exemple, les décès déclarés dans l'intervalle d'un an compris entre la date de recensement du 1er avril 1970 et celle du 1er avril 1971 et concernant des personnes âgées de 60 ans (à leur dernier anniversaire) au moment du recensement peuvent être répartis en quatre catégories: 1) personnes de 60 ans décédées au cours de l'année civile 1970; 2) personnes de 60 ans décédées au cours de l'année civile 1971; 3) personnes de 61 ans décédées au cours de l'année civile 1970; 4) personnes de 61 ans décédées au cours de l'année civile 1971. À l'aide des données sur les dates des décès tirées des bandes du NCHS, nous avons répartis les décès sur des triangles du temps et de l'âge correspondant à l'année de recensement débutant le 1er avril 1970. Ce faisant, nous avons présumé que les décès compris dans chaque triangle étaient répartis uniformément. Cette supposition est rendue nécessaire par l'absence de données fiables sur la naissance pour la plupart des cohortes visées par notre étude, données grâce auxquelles il nous aurait été possible de répartir les décès plus précisément entre les cohortes de naissance adjacentes. Pour une description plus détaillée de cette méthode, voir Shrestha (1993).

2.4 Statistiques sur l'immigration nette

Nous avons utilisé des statistiques inédites sur l'immigration nette obtenues auprès du U.S. Bureau of the Census. Si la qualité des statistiques américaines sur l'immigration est généralement mise en doute (Hill 1985), l'estimation de la taille de la population des aînés est passablement indépendante des variations dans les estimations de la migration intercensitaire. Cette robustesse découle à la fois du nombre moindre de migrants nets chez les personnes plus âgées, comparativement aux plus jeunes, et du rôle plus important de la mortalité dans la diminution du nombre de personnes âgées, comparativement à la migration nette. Par exemple, les données sur l'immigration nette font état d'un apport de 64 hommes de race noire pour la cohorte des personnes âgées de 75 ans et plus (en 1970) pendant la décennie de 1970 à 1980. À titre de comparaison, plus de 141,000 décès ont été enregistrés dans cette cohorte au cours de la même période.

Les estimations concernant les migrations des résidents sans document sont exclues des ensembles de données sur l'immigration nette, mais elles seront prises en compte dans l'interprétation des résultats. Cette exclusion a été motivée par un certain nombre de facteurs. Les estimations de la taille et de la répartition selon l'âge et le sexe des immigrants illégaux (illegal aliens) varient largement par suite de l'insuffisance d'instruments de collecte de données aux États-Unis. Cependant, même les estimations les plus exagérées du nombre de migrants sans document sont négligeables par rapport aux décès en ce qui concerne les aînés.

Nous avons décrit un certain nombre d'ajustements apportés aux données de base: utilisation de compilations de données censitaires inédites de 1970 à cause d'un surdénombrement marqué de la population des centenaires dans les statistiques officielles; utilisation de compilations modifiées pour la race des recensements de 1980 et de 1990 et exclusion des estimations concernant les populations d'étrangers sans document. Pour juger des effets de ces ajustements sur nos résultats, nous avons procédé à plusieurs analyses de sensibilité utilisant des données non corrigées. Même si nous n'avons observé que de légères différences entre les résultats obtenus avec les données officielles et les données corrigées (sauf pour les 100 ans et plus), les analyses intercensitaires de cohortes utilisant des données non corrigées ont généralement laissé voir des écarts plus grands dans les résultats définitifs, confirmant dans l'ensemble la pertinence des corrections apportées.

3. SOURCES D'ERREURS DANS LES RECENSEMENTS ET LES REGISTRES DE DÉCÈS

Les erreurs dans les données démographiques ont été classées en deux catégories: erreurs de couverture et erreurs de contenu. La couverture s'entend de la mesure dans laquelle les personnes ou les événements faisant partie d'un ensemble défini dans un système de données particulières sont pris en compte. Le contenu s'entend de la qualité de l'information effectivement compilée sur les personnes ou les événements. L'un ou l'autre de ces deux types d'erreurs peut créer des discordances entre l'évolution intercensitaire de la taille d'une cohorte et les décès intercensitaires. Toutefois, si les deux sources de données sont marquées par le même taux net d'omissions, la concordance entre les deux sera maintenue. En outre, dans ces conditions, les taux de mortalité resteront également exacts.

Même si les cas d'erreurs dans la déclaration de l'âge suivent la même tendance dans les recensements et dans les certificats de décès, cela ne permettra pas, en règle générale, de maintenir la concordance des changements de taille des cohortes entre les deux sources de données. En effet, comme les taux de mortalité augmentent avec l'âge, la distribution par âge des décès chez les aînés est "plus vieille" que la distribution par âge de la population. Par exemple, si on place par erreur 10% des personnes et des décès du groupe d'âge 75-79 dans le groupe d'âge 80-84, l'incidence proportionnelle de cette erreur sur les données démographiques sera plus grande que l'incidence proportionnelle sur les taux de mortalité. Cette tendance à déclarer un âge erroné aurait pour effet de biaiser les taux de mortalité et influerait également sur la cohérence des tests que nous utilisons.

Le Census Bureau a utilisé des méthodes démographiques et statistiques pour estimer la qualité de la couverture des recensements. Les méthodes démographiques comparent les estimations du nombre réel de naissances,

diminuées des nombres estimés de décès dans la cohorte et des migrations, aux données censitaires (voir à ce propos le résumé dans Robinson et coll. 1993, et Himes et Clogg 1992). Les méthodes statistiques comparent un groupe de personnes tiré d'une source de données de rechange (p. ex., le Current Population Survey) aux données individuelles du recensement. Une troisième méthode consiste à comparer les données censitaires concernant les aînés aux dénombrements des personnes inscrites dans les dossiers du programme fédéral d'assurance-maladie (Medicare).

Les évaluations du recensement de 1970 réalisées par le U.S. Bureau of the Census (1973, 1974, 1975) ont conduit à un certain nombre de conclusions générales concernant la population des aînés. Premièrement, l'erreur nette (combinaison des erreurs de couverture et de contenu) est plus importante dans le cas des aînés que des personnes plus jeunes. Deuxièmement, les taux d'erreur nets ont été plus élevés dans le cas des femmes que dans celui des hommes, par suite du plus grand nombre d'erreurs commises dans la déclaration de l'âge. Cependant, les taux bruts d'omission (qui ne constituent qu'un élément de l'erreur nette) étaient plus élevés dans le cas des hommes. Troisièmement, les taux d'erreur nets, d'omission bruts et d'erreur dans la déclaration des caractéristiques démographiques sont beaucoup plus élevés dans le cas des américains de race noire que de ceux de race blanche. Quatrièmement, il semble que les statistiques officielles contiennent une très grande quantité d'erreurs dans la déclaration de l'âge. Par exemple, il est intéressant de noter que pour les quatre groupes de race-sexe des personnes âgées de 65 à 69 ans en 1970, les estimations tirées de l'analyse démographique laissent conclure à une surestimation nette dans les données censitaires tandis que l'étude fondée sur les données du Medicare laisse transparaître d'importantes omissions dans les données censitaires (de 2.1% dans le cas des femmes blanches à 12.6% dans le cas des hommes noirs et de ceux des autres catégories raciales). Cette comparaison donne à conclure qu'en plus des taux d'omission bruts dans le dénombrement des personnes âgées de 65 à 69 ans, d'autres erreurs encore plus grandes (surtout, semble-t-il, une exagération de l'âge chez les personnes de moins de 65 ans) agissent dans le sens contraire pour augmenter les estimations du dénombrement net des personnes appartenant à ces groupes d'âge. Il découle notamment de cette situation que les caractéristiques d'une portion substantielle de la population classée dans le groupe des 65 ans et plus lors du recensement se rapportent en fait à des personnes âgées de moins de 65 ans (U.S. Bureau of the Census 1976).

Comparativement aux données censitaires de 1970, les taux d'erreurs nets observés en 1980 dans la plupart des groupes d'âge-race-sexe ont été sensiblement moins élevés. Toutefois, comme le faisait observer le Bureau of the Census (1988), les résultats du programme postcensitaire (Post Enumeration Program, ou PEP) et de la Housing Unit Enumeration Duplication study de 1980 donnent à conclure qu'une part considérable du total des dénombrements censitaires, soit plus de 1.1% selon toute vraisemblance, contient des personnes déjà recensées. Les données laissent supposer que le taux de duplication était beaucoup moindre au cours des recensements antérieurs. Ainsi, les améliorations relevées dans la couverture nette du recensement de 1980 semble malheureusement en partie attribuables à un problème de duplication des données (U.S. Bureau of the Census 1988:10).

Le Census Bureau compte procéder à des évaluations complètes de la qualité du recensement de 1990, mais la diffusion des résultats obtenus est demeurée jusqu'à maintenant fragmentaire. Il semble que le problème de sous-dénombrement des données brutes ait été moins grave en 1980 qu'en 1990 (Robinson et coll. 1993), mais cette différence pourrait être due à un taux de duplication plus élevé dans le recensement de 1980. On peut formuler un certain nombre de généralisations concernant la tendance au sous-dénombrement net dans le recensement de 1990, dans le cas des aînés. Premièrement, confirmant la tendance historique, les estimations de l'erreur nette pour les Afro-américains dépassent largement celles des blancs. La différence la plus importante s'observe dans le cas des hommes âgés de 60 à 64 ans. Le taux de sous-dénombrement net pour les hommes de race noire atteint 10.3%, alors qu'il n'est que de 2.6% chez les hommes de race blanche (différence de 7.7 points). Deuxièmement, alors qu'on observe un sous-dénombrement pour tous les groupes d'âge chez les hommes, il existe des problèmes de surdénombrement dans plusieurs des groupes de femmes. Finalement, comme le soulignent Robinson et coll. (*ibid*), les tendances de la couverture nette sont généralement comparables dans les trois derniers recensements pour chaque groupe de race-sexe.

Les statistiques officielles sur les décès produites par le National Center for Health Statistics constituent la source de base de données sur la mortalité annuelle aux États-Unis. Ces chiffres sont généralement utilisés sans ajustement pour tenir compte du sous-dénombrement et des erreurs de déclaration qui se glissent dans les certificats de décès. Toutefois, on présume en général que le système de recensement des décès est pratiquement complet (Wilkin 1981; U.S. Bureau of the Census 1984a; National Center for Health Statistics 1968) même si aucun test national n'a été fait pour confirmer ce diagnostic depuis la mise en place de la Death Registration Area en 1933. Cette supposition s'appuie sur les exigences légales strictes en vigueur concernant le recensement des décès ainsi que sur la nécessité, pour les survivants, de prouver le décès aux fins des arrangements funéraires, du règlement des droits successoraux et de la collecte des prestations d'assurance (U.S. Bureau of the Census 1984a; Wilkin 1981). Les calculs réalisés par Coale et Kisker (1990) donnent toutefois à penser qu'il existe un problème de sous-dénombrement des décès, en particulier chez les aînés. Dans le cas des personnes de couleur, par exemple, le total des décès portés aux registres de l'état civil est inférieur de 7% à celui des décès inscrits aux dossiers du Medicare dans le cas des hommes âgés de plus de 80 ans en 1980, et de 10% dans le cas des femmes. Toutefois, il est possible que l'écart relevé soit dû à des différences dans l'âge déclaré entre les deux sources plutôt qu'à un problème de sous-dénombrement.

Une étude au cours de laquelle on a comparé un échantillon de certificats de décès datant de mai à août 1960 aux données du recensement de 1960 constitue la meilleure évaluation de la concordance des données sur l'âge déclaré contenues dans les tableaux de recensements et dans les registres de l'état civil - sans doute une des sources les plus importantes d'erreurs de contenu influant sur notre test de cohérence (NCHS 1968; Hambright 1969). Même si les données ont été recueillies avant la période visée par notre projet, les résultats de cette étude mettent en lumière un problème qui risque d'exister encore aujourd'hui. Les auteurs de l'étude ont constaté ce qui suit: 1) dans le cas des blancs, les données des deux sources concordent assez bien même lorsque l'âge augmente, mais pour les personnes de couleur, la concordance est moins nette; 2) lorsqu'il y a discordance entre les deux sources de données, la différence d'âge dans le cas des blancs est généralement inférieure à un an, mais cette différence est typiquement supérieure à un an dans le cas des personnes de couleur, en particulier pour celles âgées de 45 ans et plus; 3) pour les blancs de tous âges et pour les personnes de couleur âgées de moins de 45 ans, l'âge indiqué sur le certificat de décès est typiquement plus avancé que celui déclaré au moment du recensement. Par contre, pour les personnes de couleur âgées de 45 ans et plus, l'âge indiqué sur le certificat de décès est en moyenne moins avancé que celui déclaré lors du recensement.

Les auteurs de cette étude n'ont pas été en mesure de déterminer laquelle des deux sources donnait l'âge "réel". Rosenwaike et Logue (1983) ont cherché à vérifier les données sur l'âge indiqué dans les certificats de décès des personnes âgées de 85 ans et plus pour la période de 1968 à 1972. Ils ont à cette fin tiré un échantillon de certificats de personnes décédées à un âge très avancé en Pennsylvanie et aux New Jersey. Ils ont ensuite comparé ces données à celles du recensement manuscrit de l'année 1900. Au total, 1,429 comparaisons ont été établies: 960 concernant des personnes de race blanche et 496 concernant des personnes de couleur.

Ils ont constaté que la concordance des renseignements sur l'âge entre les deux sources diminuait à mesure que l'âge augmentait pour les deux groupes raciaux. Des différences frappantes ont été observées entre les groupes raciaux. La concordance était élevée dans le cas des blancs, sauf pour les personnes âgées de 100 ans et plus. Toutefois, pour les personnes de couleur, la concordance était sensiblement moins bonne. Les auteurs ont par ailleurs relevé qu'à l'intérieur de chaque groupe racial, le sexe avait peu d'incidence sur la concordance des données.

4. MÉTHODE INTERCENSITAIRE D'ÉVALUATION DE LA QUALITÉ DES STATISTIQUES SUR LES AÎNÉS

La présente analyse permet de déterminer l'ampleur de la discordance des sources de données sur les aînés à l'aide d'une méthode intercensitaire fondée sur les cohortes. La taille prévisible d'une cohorte d'âge ouverte dans le second recensement peut être estimée à partir de sa taille dans le premier recensement et des décès survenus au sein de cette cohorte au cours de la période intercensitaire, après l'ajustement tenant compte des migrations (Condran, Himes et Preston 1991). Le recours à une catégorie ouverte permet l'observation de la tendance du rapport tout en atténuant l'effet des valeurs extrêmes dues à des erreurs dans des groupes d'âge particuliers. Cette méthode est insensible aux erreurs commises dans la déclaration de l'âge des personnes décédées ou des personnes recensées dont l'âge est supérieur à celui qui marque le début de la catégorie ouverte.

En utilisant les données censitaires et celles portant sur les décès et les migrations pour une période intercensitaire donnée, l'analyse intercensitaire des cohortes nous permet d'estimer la taille prévisible de chaque cohorte d'âge ouverte du recensement suivant. Les statistiques susmentionnées, classées par années d'âge, par sexe et par race (blanche et noire), ont servi à calculer l'équation suivante pour la population prévue à l'époque du second recensement:

$$\hat{N}_x(2) = N_{x-10}(1) - D_{x-10}(1) + M_{x-10}(1) \qquad (1)$$

où

$\hat{N}_x(2)$ = la population prévue d'âge x et plus, lors du second recensement, réalisé 10 ans après le premier;

$N_{x-10}(1)$ = la population recensée d'âge $x - 10$ et plus au temps 1, soit au premier recensement;

$D_{x-10}(1)$ = les décès survenus pendant la période intercensitaire dans la cohorte des personnes d'âge $x - 10$ et plus (au moment du premier recensement);

$M_{x-10}(1)$ = l'immigration légale nette intercensitaire dans la cohorte des personnes d'âge $x - 10$ et plus (au moment du premier recensement).

On peut également calculer à l'aide d'une méthode analogue la population prévue d'un âge donné (par opposition à celle d'âge x et plus). Dans l'un ou l'autre de ces cas, il est ensuite possible de calculer le rapport de la population observée, dénombrée lors du recensement suivant, sur la population prévue (après simplification et en présumant que la migration nette est nulle) à l'aide de la formule suivante:

$$R_x = \frac{N_x(2)}{N_{x-10}(1) - D}. \qquad (2)$$

Le changement de taille de la cohorte mesuré entre deux recensements successifs ne peut être dû qu'aux décès et aux migrations. Un rapport de 1.00 indiquerait donc une concordance complète entre les sources de données. (Noter toutefois qu'un tel rapport n'est pas nécessairement un signe d'exactitude des données. Ainsi, si l'âge d'une personne donnée a systématiquement été augmenté de n années, la méthode ne permettra pas de relever cette déclaration erronée). Dans les faits, toutefois, le dénombrement

indiqué pourra également subir l'influence: 1) des erreurs de couverture dans l'un ou l'autre ou les deux recensements; 2) des erreurs (en plus ou en moins) dans les registres des décès ou les statistiques de l'immigration; 3) des déclarations erronées (concernant l'âge, la race, *etc.*) dans l'une ou l'autre ou l'ensemble des sources de données (Ewbank 1981; Shryock et Siegel 1976; Condran et coll. 1991). Le rapport des populations observées sur les populations prévues est un outil de diagnostic utile lorsqu'il est possible d'attribuer les écarts par rapport à 1.00 à l'effet de ces erreurs sous-jacentes. Toutefois, il ne constitue pas un outil très précis puisque différents types d'erreurs peuvent produire le même genre d'écarts. Il peut néanmoins permettre un choix plus facile entre diverses solutions concurrentes.

5. EFFETS DES ERREURS SUR LES RAPPORTS DES POPULATIONS OBSERVÉES SUR LES POPULATIONS PRÉVUES

Certains types d'erreurs ont un effet direct visible sur la formule du rapport lui-même (effets qui ont été confirmés par les simulations que nous avons réalisées). Pour simplifier la démonstration, désignons par R_x dans l'équation (2) le rapport de la population observée sur la population prévue pour l'âge x et plus au deuxième recensement. On peut distinguer les possibilités principales suivantes d'erreurs de couverture et leurs répercussions sur la configuration des rapports en fonction de l'âge:

1) Si $N_{x-10}(1)$ et D sont également complets et que $N_x(2)$ est assorti d'un degré de complétude relative de $C(2)$, alors la configuration des rapports en fonction de l'âge sera constante et son niveau sera $C(2)$.

2) Si $N_x(2)$ et D sont également complets et que $N_{x-10}(1)$ est assorti d'un degré de complétude relative de $C(1)$, alors la configuration des rapports en fonction de l'âge sera:

   a) supérieure à 1.00 et augmentera en fonction de l'âge si $C(1) < 1.00$;

   b) inférieure à 1.00 et diminuera en fonction de l'âge si $C(1) > 1.00$.

La raison pour laquelle ces erreurs induisent un effet relatif à l'âge sur le rapport R_x est qu'une erreur proportionnelle donnée dans $N_{x-10}(1)$ engendre des erreurs proportionnelles de plus en plus grandes dans le dénominateur à mesure que les deux termes compensatoires du dénominateur (l'un positif et l'autre négatif) tendent l'un vers l'autre en valeur absolue. Cette égalisation est due au fait qu'une proportion plus grande des membres de chaque cohorte meurent pendant la période intercensitaire à mesure que l'âge avance.

3) Si $N_{x-10}(1)$ et $N_x(2)$ sont également complets et que D est assorti d'un degré de complétude relative de $C(D)$, alors la configuration des rapports en fonction de l'âge sera:

   a) supérieure à 1.00 et augmentera en fonction de l'âge si $C(D) > 1.00$ (c.-à-d., si les certificats de décès sont plus complets que les dénombrements des deux recensements);

   b) inférieure à 1.00 et diminuera en fonction de l'âge si $C(D) < 1.00$.

Ici encore, on observe une tendance relative à l'âge puisqu'une erreur proportionnelle égale dans D engendrera des erreurs proportionnelles plus grandes dans le dénominateur à mesure que les deux éléments de ce dernier tendent l'un vers l'autre en termes absolus.

On peut mieux comprendre les effets des déclarations erronées de l'âge en examinant les éléments qui composent cette formule. Shrestha (1993) et Condran et coll. (1991) ont introduit diverses erreurs dans des ensembles simulés, dépourvus d'erreurs, typiques des conditions démographiques actuelles existant aux États-Unis et aux Pays-Bas respectivement. Ils ont démontré qu'une tendance nette à exagérer l'âge qui se limiterait à deux recensements produirait des rapports oscillant aux alentours de 1.00 jusqu'à un âge avancé, pour chuter ensuite à des valeurs très faibles. La diminution du rapport à des valeurs inférieures à 1.00 s'explique ici encore par le fait qu'une erreur dans un des éléments du dénominateur (dans ce cas, l'inflation de $N_{x-10}(1)$ par une exagération de l'âge) engendre des effets disproportionnés dans le dénominateur. Même si la chute rapide de la distribution par âge peut entraîner une inflation plus grande de $N_x(2)$ par rapport à $N_{x-10}(1)$, l'inflation du dénominateur finira tôt ou tard par excéder celle du numérateur et les rapports diminueront. (À titre d'exemple, voir la figure 1 de Condran et coll. 1991).

Un problème d'exagération de l'âge limité aux recensements des décès donnera un rapport supérieur à 1.00 et qui augmentera avec l'âge. Dans ce cas, le dénominateur est trop petit (sa composante négative est trop grande) et le déficit proportionnel augmente avec l'âge.

L'existence d'une tendance identique à exagérer l'âge dans les certificats de décès et les recensements donnera également des rapports qui finiront par augmenter avec l'âge. Ce résultat important est robuste dans la mesure de l'erreur introduite (Condran et coll. 1991). Il découle du fait que les distributions par âge chutent de plus en plus rapidement à mesure que l'âge avance, de sorte qu'un pourcentage identique de personnes qui exagèrent leur âge véritable entraînera un pourcentage plus grand d'erreurs dans les distributions par âge aux âges très avancés. Autrement dit, $N_x(2)$ est assorti d'un facteur d'inflation plus grand que $N_{x-10}(1)$. Dans ce cas, une certaine inflation de $N_{x-10}(1)$ verra ses effets sur le dénominateur compensés par l'inflation de D.

6. RÉSULTATS

L'analyse intercensitaire des cohortes a porté sur quatre groupes de sexe-race des États-Unis, au cours des périodes 1970-1980 et 1980-1990. Les figures 1 et 2 illustrent les rapports calculés de la population observée sur la population prévue pour des groupes d'âge précis en fonction de la race, du sexe et de la période intercensitaire.

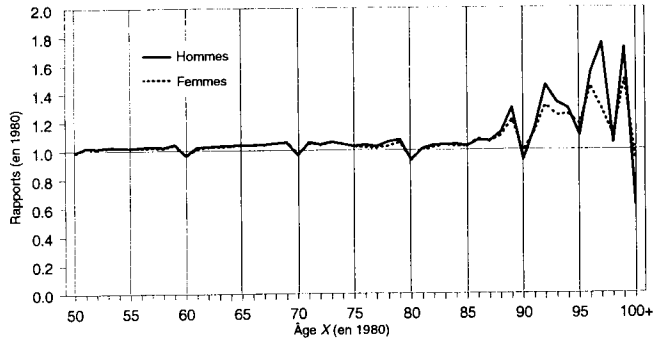

Figure 1A. Rapports intercensitaires de la population observée sur la population prévue: blancs, 1970-1980.

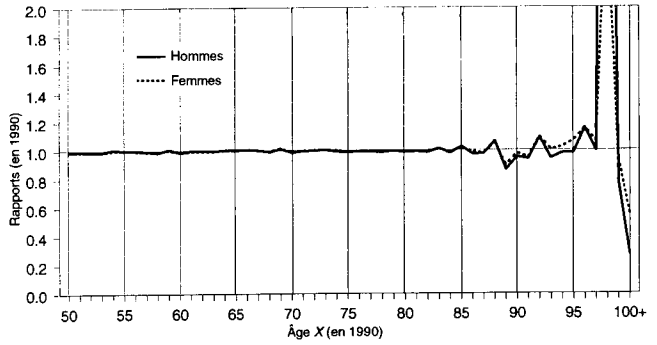

Figure 1B. Rapports intercensitaires de la population observée sur la population prévue: blancs, 1980-1990.

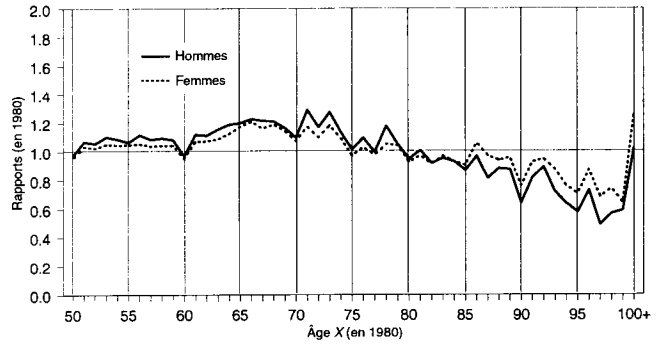

Figure 2A. Rapports intercensitaires de la population observée sur la population prévue: noirs, 1970-1980.

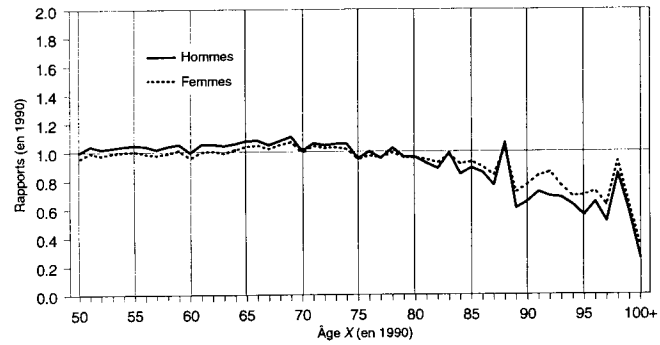

Figure 2B. Rapports intercensitaires de la population observée sur la population prévue: noirs, 1980-1990.

Pour toutes les combinaisons de races et de périodes, l'effet des tendances relatives à l'âge sur les rapports est pratiquement le même pour les hommes et pour les femmes. Dans tous les cas, le degré de discordance augmente avec l'âge, même si on ne commence à observer un écart systématique et significatif par rapport à 1.00 qu'à partir de 95 ans dans le cas des blancs en 1980-1990. On observe une discontinuité nette dans beaucoup de ces séries à l'âge de 100 ans, ce qui reflète le caractère idiosyncratique de la déclaration de l'âge et les méthodes d'ajustement utilisées par le Census Bureau dans le cas des centenaires.

6.1 Résultats pour les blancs

6.1.1 Période intercensitaire 1970-1980

Comme l'indique la figure 1A, les rapports de la population blanche ont tendance à se maintenir au-dessus de 1.00 au cours de cette période et ils augmentent avec l'âge (jusqu'à 100 ans). Cette configuration pourrait être le résultat de divers types d'erreurs de données dont les deux plus plausibles sont:

1) un sous-dénombrement dans le recensement de 1970 par rapport au recensement de 1980 et au recensement des décès;

2) des probabilités à peu près égales d'exagération de l'âge dans les certificats de décès et lors des deux recensements.

Nous croyons que la première explication risque davantage d'être la bonne. Si l'évolution des rapports en fonction du vieillissement découlait de tendances semblables à l'exagération de l'âge déclaré dans les certificats de décès et les recensements, on pourrait s'attendre à une situation comparable au cours de la décennie 1980-1990, notamment du fait que le recensement de 1980 figure dans les deux comparaisons. On ne s'attendrait pas non plus à voir les propensions culturelles à déclarer un âge erroné disparaître soudainement. Pourtant, le graphique de l'évolution des rapports pour les blancs au cours de la décennie 1980-1990 (figure 1B) laisse constater une remarquable concordance, bien meilleure que dans la plupart des pays européens et équivalant à celle observée en Suède et dans les Pays-Bas, deux pays dotés de registres de la population extrêmement efficaces (Condran et coll. 1991). La concordance des données observée pour la décennie 1980-1990 est également beaucoup plus grande que celle obtenue dans d'autres pays de langue anglaise: Angleterre et pays de Galles, Canada, Australie et Nouvelle-Zélande.

Notre préférence pour la première explication est également motivée par le fait que le Census Bureau a conclu que le recensement de 1980 était plus complet que celui de 1970 (U.S. Bureau of the Census 1988; Robinson et coll. 1993). Cette conclusion est fondée en partie sur l'analyse démographique et n'est donc pas entièrement indépendante du genre de preuve que nous examinons. Toutefois, l'analyse démographique du Bureau fait grand usage de groupes d'âge plus jeunes que ceux auxquels nous nous intéressons ici. En outre, la conclusion selon laquelle la couverture des recensements a été améliorée est également confirmée par

le programme postcensitaire au cours duquel des données individuelles du recensement sont comparées à des données provenant d'autres systèmes.

6.1.2 Période intercensitaire 1980-1990

Comme mentionné précédemment, l'évolution des rapports dans le cas des blancs (en particulier les femmes) laisse constater une très grande concordance des données, de loin meilleure que dans la plupart des pays européens. Notre étude semble confirmer l'hypothèse de Vaupel (1993) selon laquelle la population blanche des États-Unis présenterait des taux de mortalité plus bas, au-delà de l'âge de 80 ans, que dans tous les autres pays industrialisés. Toutefois, la prudence est de mise. Si nos méthodes laissent voir une grande concordance des données entre les recensements et les certificats de décès pour la période 1980-1990, rien n'indique que ces données soient nécessairement exactes. Condran et coll. (1991) décrit une situation dans laquelle une tendance à déclarer des âges erronés donne une série de rapports de 1.00 pour tous les âges. Par ailleurs, la méthode intercensitaire ne permet pas de dépister les sujets qui font systématiquement des déclarations erronées de leur âge. Comme le soulignait Horiuchi (1993), une personne qui exagère une première fois son âge - pour être admise à une école ou être autorisée à travailler plus jeune, pour éviter le service militaire lorsqu'elle approche de l'âge limite supérieur ou pour bénéficier plus tôt de la sécurité sociale, du programme Medicare ou d'une pension - pourra avoir tendance à continuer d'exagérer son âge par la suite. Une telle possibilité ne peut être ignorée, même si nous sommes incapables de la mesurer directement.

6.2 Résultats pour les noirs

L'évolution des rapports en fonction de l'âge est beaucoup plus régulière dans le cas des Afro-américains (voir les figures 2A et 2B). Ils commencent à diminuer vers 70 ans pour les deux sexes dans les deux périodes et poursuivent cette chute jusqu'aux âges avancés (jusqu'à 100 ans en 1970-1980). Avant 70 ans, les rapports se maintiennent typiquement au-dessus de 1.00 en 1970-1980, et légèrement au-dessus de 1.00 pour les hommes en 1980-1990.

L'écart observé entre les rapports pour des groupes d'âge donnés entre les résultats de 1970-1980 et de 1980-1990 n'est pas étranger au sous-dénombrement relatif qui a caractérisé le recensement de 1970. Comme nous l'avons déjà mentionné, il est vraisemblable qu'un tel sous-dénombrement se soit produit également dans le cas des blancs. Ce sous-dénombrement ne suffit pas toutefois à expliquer la tendance à la baisse persistante des rapports au-delà de 70 ans pour les deux périodes. Deux raisons principales peuvent permettre d'expliquer ce déclin des séries de rapports pour les Afro-américains:

1) L'inscription des décès aux registres d'état civil n'est pas faite avec la même efficacité que les recensements;

2) L'exagération de l'âge est plus importante dans les recensements que dans les certificats de décès.

Coale et Kisker (1990) préfèrent la première explication. Ils relèvent que les populations reconstituées à partir des données sur les décès à l'aide de méthodes à r variables (Preston et Coale 1982) sont trop petites par rapport aux données censitaires de 1980 pour les plus de 65 ans, ce qui porte à conclure que les décès n'ont pas tous été portés aux registres. Ils relèvent également que chez les Afro-américains d'âge avancé, on observe un déficit dans les décès déclarés par rapport aux dossiers Medicare.

Toutefois, ces deux observations pourraient également s'expliquer par une exagération de l'âge lors des recensements (et dans les dossiers Medicare) par rapport aux données des certificats de décès. Une comparaison directe des certificats de décès de 1960 et des données se rapportant aux mêmes sujets dans le recensement de 1960 (NCHS 1968; Hambright 1969) tend à confirmer l'existence d'une telle tendance. Pour les hommes comme pour les femmes, le total des décès chez les plus de 50 ans, lorsque ces décès sont classés selon l'âge déclaré au recensement, ne s'écarte pas de plus de 1% du total des décès classés selon l'âge inscrit dans les certificats de décès. Toutefois, pour l'ensemble des personnes de 65 ans et plus, le total des décès classés selon l'âge au recensement dépasse de 15.4% celui des décès classés selon l'âge inscrit dans le certificat de décès dans le cas des femmes, et de 7.1% dans le cas des hommes. Pour les personnes de 75 ans et plus, les écarts atteignent 23.3 et 17.8% respectivement et pour le groupe des 85 ans et plus, ils sont de 39.2 et 17.6% respectivement.

Ces écarts importants de l'âge déclaré entre les recensements et les certificats de décès peuvent expliquer la tendance à la baisse des rapports observée dans les graphiques des figures 2A et 2B. Elo et Preston (1994) ont calculé que les valeurs de R_x pour les Afro-américains, entre les périodes 1950-1960 et 1960-1970, périodes qui encadrent la comparaison des données censitaires et celles des certificats de 1960. Ils démontrent que si les âges au décès sont "corrigés" pour correspondre à ceux déclarés lors des recensements, la tendance à la baisse des rapports est éliminée.

Les raisons qui poussent les Afro-américains à déclarer un âge plus avancé lors des recensements, comparativement à l'âge inscrit dans le certificat de décès, ne sont pas claires. Cette tendance n'apparaît pas avant le recensement de 1940, c'est-à-dire le premier réalisé après l'adoption de la Loi sur la sécurité sociale. Ce recensement laisse constater un large excédent d'Afro-américains appartenant aux groupes d'âge 65-69 ans et 70-74 ans, et un déficit des personnes du groupe d'âge 50-64 ans (Elo et Preston 1994). Comme le rappelle Wolfenden (1954:56), "Les distorsions dans les données sur les noirs étaient si marquées qu'il a fallu procéder à une nouvelle répartition préliminaire de ces populations (et de ces décès) entre 55 et 69 ans, aux fins de la préparation des tables de survie." Cet excédent persiste, même s'il s'atténue graduellement, jusque dans les recensements les plus récents (comme le montre la figure 2). Quelle qu'en soit la raison, nous croyons que l'explication principale des écarts importants observés entre les données censitaires et les données des certificats de décès pour la population noire américaine s'expliquent

principalement par une tendance, chez les personnes de ce groupe, à exagérer leur âge au moment des recensements par rapport aux données inscrites dans les certificats de décès. Il s'ensuit de cette tendance que les taux de mortalité des groupes d'âge supérieurs à 65 ans pour les Afro-américains donneront vraisemblablement lieu à de graves sous-estimations. Il est effectivement possible que les courbes des taux de mortalité des blancs et des noirs se croisent à un âge avancé, mais il serait dangereux de fonder une telle conclusion uniquement sur les données des recensements et des registres de l'état civil. La discordance entre ces deux sources de données est tout simplement trop grande pour nous permettre de faire une estimation fiable des taux de mortalité pour les groupes d'âge avancé.

7. CONCLUSION

Les erreurs de couverture et de contenu relevées dans les systèmes de recensement des personnes et des décès aux États-Unis laissent planer un doute sérieux sur la qualité de ces données en ce qui concerne les aînés. Nous avons évalué la concordance des données provenant de ces deux sources pour les populations blanche et afro-américaine, en concentrant notre attention sur les personnes âgées de 60 ans et plus, où les tendances affichées par les taux de mortalité ont la plus grande incidence sur les programmes sociaux et où les données sont les plus problématiques. L'analyse intercensitaire des cohortes laisse constater des discordances liées à l'âge entre les deux sources de données pour les deux périodes: 1970-1980 et 1980-1990.

Afin de déterminer quelles combinaisons de degré de couverture et de tendance à la déclaration d'un âge erroné produirait les résultats empiriques, nous avons réalisé une série de simulations. Les discordances observées dans les données américaines sont examinées à la lumière des résultats de ces simulations et des données de la documentation spécialisée concernant la nature des erreurs de couverture et de contenu dans les sources de données.

Les données concernant les blancs pour la période intercensitaire 1980-1990 laissent constater une concordance remarquable. Leur qualité, jusqu'à 95 ans, s'approche de celle des données de la Suède et des Pays-Bas, deux pays dotés de registres de la population extrêmement efficaces. Les données portant sur les blancs pour la décennie 1970-1980 laissent voir plus de discordances. L'explication la plus plausible de cette différence est le sous-dénombrement relatif net du recensement de 1970 combiné à la tenue de statistiques plus complètes sur les décès. En conséquence, les estimations de la mortalité chez les aînés qui utilisent des numérateurs tirés des certificats de décès et des dénominateurs tirés du recensement de 1970 risquent de surestimer la mortalité.

Les données concernant les Afro-américains laissent voir une tendance différente. Au-delà de 70 ans, la population recensée diminue graduellement comparativement à la population prévisible tant en 1980 qu'en 1990. Il semble que cette discordance soit principalement due à une tendance à exagérer l'âge lors des recensements, par rapport aux données portées aux certificats de décès. Il s'ensuit de cette tendance que les taux de mortalité des Afro-américains plus âgés risquent d'être sérieusement sous-estimés. Il est possible que les courbes des taux de mortalité des noirs et des blancs se croisent à un âge avancé, mais il serait imprudent de fonder une telle conclusion uniquement sur les données des recensements et des registres de l'état civil.

REMERCIEMENTS

Cette recherche a été réalisée au Population Studies Center de la University of Pennsylvania, avec l'aide financière du National Institute of Aging (AG10168) et du Boettner Institute of Financial Gerontology. Nous remercions Irma Elo, Douglas Ewbank, Shiro Horiuchi et J. Gregory Robinson pour leurs commentaires utiles sur notre projet et sur le texte du présent article. Merci également à J. Gregory Robinson et au U.S. Bureau of the Census qui nous ont fournis les données censitaires inédites et les tableaux de données sur les migrations internationales.

ANNEXE A

Source: Shrestha (1993)

Trois sources principales de données ont été utilisées dans notre étude: 1) recensements de 1970, 1980 et 1990; 2) registres officiels des décès; 3) statistiques sur l'immigration nette. Nous décrivons ci-après ces sources et les ajustements que nous avons apportés aux données.

1.A Recensement de 1970

Les tableaux officiels des caractéristiques démographiques de base de la population en 1970 sont présentés dans le document Series B – U.S. Summary of the 1970 Census (U.S. Bureau of the Census 1972). Nous savons que ces données officielles contiennent des erreurs importantes qui risquent de biaiser notre enquête sur le recensement des aînés aux États-Unis. La première est un surdénombrement évident des personnes centenaires. Alors que 106,000 personnes ont été classées dans cette catégorie ouverte, des analyses démographiques indirectes donnent plutôt à conclure que le nombre correct oscillerait plutôt entre 3,000 et 8,000 (Siegel 1974; Siegel et Passel 1976). Ce surdénombrement semble avoir été causé par une interprétation erronée d'une des questions du formulaire plutôt que par une tendance systématique à se déclarer faussement dans la catégorie des centenaires. Le deuxième problème découle d'une erreur de classification de la population en groupes raciaux dans les tableaux du recensement total qui touche 21,000 personnes âgées de 65 ans et plus. Finalement, le recensement officiel a omis plus de 23,000 personnes (de tous âges) dont les données ont été retrouvées après la publication des tableaux officiels.

À cause des erreurs contenues dans les tableaux officiels, nous avons utilisé des tableaux corrigés inédits obtenus auprès du U.S. Bureau of the Census. Ces données statistiques modifiées étaient exemptes des trois types d'erreurs susmentionnés. Les données sont classées selon la race (blanche ou noire), le sexe, les années d'âge (de 0 à 94 ans) et les groupes d'âge (95-99 ans et 100 et plus). Pour répartir les données du groupe des 95-99 ans en années d'âge, nous avons utilisé des distributions de l'âge moyen par sexe et par race des recensements de 1960 et de 1980 pour les blancs, et des recensements de 1950 et 1980 pour les noirs (les données en années d'âge ne sont pas disponibles pour ce groupe d'âge chez les noirs dans le recensement de 1960).

1.B Recensement de 1980

Les tableaux officiels du recensement de 1980 sont présentés dans le document Series B – U.S. Summary of the 1980 Census (U.S. Bureau of the Census 1983). Au cours de ce recensement, un grand nombre de gens (environ 6.8 millions) ont choisi de fournir une réponse de leur cru à la question concernant leur race au lieu de choisir une des catégories préétablies. Comme le recensement de 1980 est le seul à contenir une catégorie raciale résiduelle, il n'est pas directement comparable aux autres sources de données (registres de l'état civil, recensements antérieurs, *etc.*). le Census Bureau a produit un dossier modifié conforme aux catégories raciales historiques (U.S. Bureau of the Census 1984b). La méthode d'adaptation des données comportait une redistribution des macro-données sur la race fondée sur un tableau détaillé croisant la race et l'origine hispanique et découlant des données échantillons et des données intégrales du recensement. Nous décrivons en détails ci-après les modifications apportées par le Census Bureau.

Pour les 219.8 millions de personnes qui ont choisi l'une des 14 catégories raciales préétablies, aucun ajustement n'a été fait. Deux catégories de personnes (6.7 millions au total) ont choisi de fournir une description de leur cru: les personnes d'origine hispanique (5.8 millions) et celles d'une autre origine (0.9 million). Des méthodes d'ajustement différentes ont été élaborées pour ces deux groupes.

Les personnes d'origine hispanique ont été réparties entre les catégories blanche et noire (et non dans celles des autochtones ni des asiatiques ou des insulaires du Pacifique). Toutes les personnes d'origine mexicaine ont été replacées dans la catégorie blanche. Les personnes d'origine porto-ricaine et cubaine et toutes les autres personnes d'origine hispanique ont été replacées dans des groupes modifiés, blancs ou noirs, dans des proportions identiques au choix des Hispaniques de même origine qui avaient précisé la race blanche ou noire dans leur formulaire. Les calculs ont été effectués par cellules d'âge-sexe-comté.

Les personnes d'une autre origine ont été réparties entre trois groupes raciaux modifiés (blancs, noirs et autres) dans des proportions correspondant à celles observées dans les cellules d'âge-sexe-comté de leur État de résidence. Ces proportions étaient fondées sur un échantillon de données tirées du recensement de 1980. Pour de plus amples détails sur ces modifications, voir U.S. Bureau of the Census (1984b).

Les données des tableaux modifiés sont classées selon la race, le sexe, les années d'âge (0 à 99) et le groupe d'âge (100 ans et plus). Aux fins de notre recherche, nous avons utilisé les groupes raciaux modifiés, compte tenu du nombre considérable de personnes qui ont été transférées du groupe racial résiduel aux catégories blanche ou noire.

1.C Recensement de 1990

La publication des tableaux du recensement de 1990 par le U.S. Bureau of the Census n'est pas encore terminée. Les statistiques déjà publiées présentent toutefois un certain nombre de problèmes qui compliquent leur comparaison avec les recensements antérieurs et les autres sources de données. Trois problèmes sont évidents: classification des 9.3 millions de personnes placées dans une catégorie raciale résiduelle non spécifiée; incohérences dans la déclaration de l'âge et modification des méthodes d'attribution de l'âge pour les personnes qui ont omis de fournir ce renseignement.

Un dossier modifié intitulé MARS (Modified Age and Race Statistics) a été préparé par le Census Bureau pour régler les deux premiers problèmes (Word et Spencer 1991). Les modifications du recensement de 1990 ont porté sur les micro-données. Des méthodes d'imputation spéciales (hot-deck) ont servi à attribuer une race particulière aux personnes qui avaient choisi la catégorie "autre non précisé". Cette redistribution est appliquée aux dossiers individuels des tableaux détaillés entièrement révisés du recensement de 1990 (Robinson, Word et Spencer 1991).

Nous avons encore une fois choisi d'utiliser les statistiques modifiées, classées selon la race, le sexe et les années d'âge. Notre décision d'utiliser les statistiques modifiées des recensements de 1980 et de 1990 n'allait pas de soi; pour un examen plus détaillé de la question, voir Shrestha (1993).

2. Le système d'enregistrement des décès

Les statistiques annuelles nationales sur les décès que nous avons utilisées pour notre étude sont venues du National Center for Health Statistics (NCHS). Les données de 1970 à 1988 ont été tirées des bandes de données du NCHS obtenues auprès du ICPSR (NCHS 1970-1988). Elles sont classées selon la race (noire ou blanche), le sexe et les années d'âge (0 à 124 ans; catégorie ouverte pour les 125 ans et plus). Comme les bandes de données de l'année civile 1989 et des trois premiers mois de 1990 n'ont pas encore été diffusées, nous avons élaboré une méthode pour en estimer la distribution. Les statistiques finales sur la mortalité pour 1989 classées selon la race et le sexe ont été publiées par le NCHS en 1992. Les données présentées par groupes d'âge ont été réparties en années d'âge en fonction de la répartition des données de 1988. La répartition selon le mois du décès s'est fondée sur les données des registres de l'état civil (NCHS 1989). Les estimations de la distribution des décès en 1990 sont fondées sur les rapports provisoires mensuels sur la mortalité du NCHS (1990). Ces données provisoires ont été classées en années d'âge en

fonction de la répartition des données du groupe d'âge correspondant en 1988.

Tel que mentionné dans le texte, nous avons ajusté les données disponibles pour corriger deux problèmes. Premièrement, les périodes intercensitaires s'étendent du 1er avril au 31 mars tandis que les registres des décès utilisent l'année civile. Deuxièmement, les deux sources de données indiquent l'âge au dernier anniversaire au lieu de donner l'année de la naissance. Comme la date des recensements est fixée au 1er avril, nous préférons la deuxième méthode qui identifie la cohorte de naissance utilisée dans notre analyse des cohortes. Pour corriger ces deux problèmes, nous présumons que la surface tridimensionnelle du nombre de décès en fonction de l'âge et du temps est égale sur tout l'intervalle. Nous ne tenons pas compte du sous-dénombrement ni des déclarations erronées dans les certificats de décès.

3. Statistiques sur l'immigration nette

Nous utilisons des statistiques inédites sur l'immigration nette obtenues auprès du U.S. Bureau of the Census. Les données sont classées par catégories de "facteurs de changement" pour chacune des deux décennies.

L'immigration nette en fonction de l'âge, de la race et du sexe a été calculée par cohortes à l'aide de l'équation suivante:

Immigration nette = immigration internationale légale + réfugiés et admissions conditionnelles + immigration nette de civils américains + immigration nette de Porto-ricains + immigration nette d'étudiants étrangers + déplacements nets des membres des Forces armées américaines outre-mer − émigration légale.

Compte tenu de l'imprécision des données brutes fournies par le U.S. Census Bureau, un certain nombre d'ajustements ont dû être apportés. Premièrement, la classe d'âge terminale des données fournies était trop précoce (75 ans et plus au début de la décennie). Pour répartir ces données en groupes quinquennaux (75-79, . . . , 95-99; 100 et plus), nous avons présumé que les taux d'immigration nette en fonction de l'âge, de la race et du sexe demeuraient constants pendant toute la période ouverte débutant à 75 ans. Cette estimation plutôt imprécise est adéquate à cause du petit nombre d'immigrants appartenant à ce groupe d'âge. Deuxièmement, pour convertir les groupes quinquennaux en années d'âge, nous avons eu recours aux multiplicateurs ou à l'interpolation osculatrice de Sprague (Sprague, 1880-81).

BIBLIOGRAPHIE

COALE, A.J., et KISKER, E.E. (1990). Defects in data on old-age mortality in the United States: new procedures for calculating mortality schedules and life tables at the highest ages. *Asian and Pacific Population Forum*, 4(1), 1-31.

CONDRAN, G.A., HIMES, C.L., et PRESTON, S.H. (1991). Old-age mortality patterns in low-mortality countries: an evaluation of population and death data at advanced ages, 1950 to present. *Population Bulletin of the United Nations*, 30, 23-60.

ELO, I.T., et PRESTON, S.H. (1994). New estimates of old-age mortality among African Americans, 1930-1990. *Demography*, 31(3), 427-58.

EWBANK, D.C. (1981). *Age Misreporting and Age-Selective Underenumeration: Sources, Patterns, and Consequences for Demographic Analysis*. National Academy of Sciences, Committee on Population and Demography. Report No. 4. Washington, DC: National Academy Press.

HAMBRIGHT, T.Z. (1969). Comparison of information on death certificates and matching 1960 census records: age, marital status, nativity, and country of origin. *Demography*, 6(4), 413-24.

HILL, K. (1985). Illegal Aliens: an assessment. In: Panel on Immigration Statistics. *Immigration Statistics, A Story of Neglect*. Washington, DC: National Academy Press.

HIMES, C.L., et CLOGG, C.C. (1992). An overview of demographic analysis as a method for evaluating census coverage in the United States. *Population Index*, 58(4), 587-607.

HORIUCHI, S. (1993). Personal communication dated April 20, 1993.

MANTON, K.G., STALLARD, E., et VAUPEL, J.W. (1986). Alternative models for the heterogeneity of mortality risks among the aged. *Journal of the American Statistical Association*. 81, 635-644.

MC CORD, C., et FREEMAN, H.P. (1990). Excess mortality in Harlem. *New England Journal of Medicine*. 322, 172-177.

NATIONAL CENTER FOR HEALTH STATISTICS (1968). *Comparability of age on the death certificate and matching census record: United States - May - August 1960*; Vital and Health Statistics: Data Evaluation and Methods Research. By Thea Zelman Hambright. Series 2, No. 29.

NATIONAL CENTER FOR HEALTH STATISTICS (1970-1988). *Mortality Detail Files* (data and codebooks). Data were made available by the Inter-University Consortium for Political and Social Research, University of Michigan.

NATIONAL CENTER FOR HEALTH STATISTICS (1989). Births, marriages, divorces, and deaths for January-December 1989. *Monthly* vital statistics report. Vol. 38, Nos. 1-12. Hyattsville, Maryland: Public Health Service.

NATIONAL CENTER FOR HEALTH STATISTICS (1990). Births, marriages, divorces, and deaths for January-March 1990. *Monthly* vital statistics report. Vol. 39, Nos. 1-3. Hyattsville, Maryland: Public Health Service.

NATIONAL CENTER FOR HEALTH STATISTICS (1992). Advance report of final mortality statistics, 1989. *Monthly* vital statistics report. Vol. 40, No. 8, supp. 2. Hyattsville, Maryland: Public Health Service.

PRESTON, S.H. (1993). Demographic change in the United States, 1970-2050. *Demography and Retirement: The 21st Century*, (Sylvester Scheiber, éd.). New York: Praeger Press.

PRESTON, S.H., et COALE, A.J. (1982). Age structure, growth, attrition, and accession: a new synthesis. *Population Index*. 48(2), 217-59.

ROBINSON, J.G., AHMED, B., DAS GUPTA, P., et WOODROW, K.A. (1993). Estimation of population coverage in the 1990 United States census based on demographic analysis. *Journal of the American Statistical Association*, 88, 1061-1079.

ROBINSON, J.G., WORD, D.L., et SPENCER, G. (1991). Uncertainty for models to translate 1990 census concepts into historical racial classifications. 1990 Decennial Census, Preliminary Research and Evaluation Memorandum (PREM) No. 81, Demographic Analysis Evaluation Project D8.

ROSENWAIKE, I., et LOGUE, B. 1983. Accuracy of death certificate ages for the extreme aged. *Demography*, 20(4), 569-585.

SHRESTHA, L.B. (1993). Age Misreporting and its Effects on Old-Age Population and Death Registration Estimates: United States, 1970-1990. Unpublished doctoral dissertation, University of Pennsylvania, Population Studies Center.

SHRYOCK, H.S., et SIEGEL, J.S. (1976). *The Methods and Material of Demography*. Orlando, Florida: Academic Press, Inc. (Harcourt Brace Jovanovich, Publishers): Studies in Population Series.

SIEGEL, J.S. (1974). Estimates of coverage of the population by sex, race, and age in the 1970 census. *Demography*, 11(1), 1-23.

SIEGEL, J.S., et PASSEL, J.S. (1976). New estimates of the number of centenarians in the United States. *Journal of the American Statistical Association*. 71, 559-566.

SPRAGUE, T.B. (1880-81). Explanation of a new formula for interpolation. *Journal of the Institute of Actuaries*. 22, 270.

UNITED STATES BUREAU OF THE CENSUS (1972). *General Population Characteristics*. 1970 Census of Population and Housing. Final Report PC(1)-B1. U.S. Summary. Washington, DC: U.S. Government Printing Office.

UNITED STATES BUREAU OF THE CENSUS (1973). *The Medicare Record Check: an Evaluation of the Coverage of Persons 65 Years of Age and Over in the 1970 Census*. 1970 Census of Population: Evaluation and Research Program, PHC(E)-7. Washington, DC: U.S. Government Printing Office.

UNITED STATES BUREAU OF THE CENSUS (1974). *Estimates of Coverage of Population by Sex, Race, and Age: Demographic Analysis*. 1970 Census of Population: Evaluation and Research Program, PHC(E)-4. By Jacob S. Siegel. Washington, DC: U.S. Government Printing Office.

UNITED STATES BUREAU OF THE CENSUS (1975). *Accuracy of Data for Selected Population Characteristics as Measured by the 1970 CPS-Census Match*. 1970 Census of Population: Evaluation and Research Program, PHC(E)-11. Washington, DC: U.S. Government Printing Office.

UNITED STATES BUREAU OF THE CENSUS (1976). *Demographic Aspects of Aging and the Older Population in the United States*. Current Population Reports. Series P-23, No. 59. By J.S. Siegel. Washington, DC: U.S. Government Printing Office.

UNITED STATES BUREAU OF THE CENSUS (1983). *General Population Characteristics*. 1980 Census of Population and Housing. Final Report PC80-1-B1. United States Summary. Washington, DC: U.S. Government Printing Office.

UNITED STATES BUREAU OF THE CENSUS (1984a). *Demographic and Socioeconomic Aspects of Aging in the United States*. Current Population Reports. Series P-23, No. 138. By J.S. Siegel and M. Davidson. Washington, DC: US Government Printing Office.

UNITED STATES BUREAU OF THE CENSUS (1984b). Census of Population: 1980. Race detail file. 100% count. Table IV: modified counts (OMB-consistent) by age, race, and sex. Unpublished tabulations.

UNITED STATES BUREAU OF THE CENSUS (1988). *The Coverage of Population in the 1980 Census*. 1980 Census of Population and Housing: Evaluation and Research Reports, PHC80-E4. By R.E. Fay, J.S. Passel and J.G. Robinson. Washington, DC: U.S. Government Printing Office.

VAUPEL, J.W. (1993). Verbal presentation at Research Workshop on Oldest Old Mortality, Duke University, Durham, North Carolina. March, 1993.

WILKIN, J.C. (1981). Recent trends in the mortality of the aged. *Transactions of the Society of Actuaries*. Vol. XXXIII, 11-62.

WOLFENDEN, H.H. (1954). *Population Statistics and their Compilation*. Revised Edition. The University of Chicago Press: published for the Society of Actuaries.

WORD, D.L., et SPENCER, G. (1991). Age, sex, race, and Hispanic origin information from the 1990 census: a comparison of census results with results where age and race have been modified. 1990 CHS-L-74. Draft dated August 1991.

ZELNIK, M. (1969). Age patterns of mortality of American Negroes: 1900-02 to 1959-61. *Journal of the American Statistical Association*. 64, 433-451.