NATIONS UNIES
CONSEIL
ECONOMIQUE
ET SOCIAL

Distr.
LIMITEE

E/CN.14/CAS.4/VS/3
11 novembre 1964

FRANCAIS
Original : ANGLAIS

COMMISSION ECONOMIQUE POUR L'AFRIQUE
Cycle d'études sur les statistiques
d'état-civil
Addis-Abéba, 14 - 19 décembre 1964

LES ERREURS ET DISTORSIONS NON DUES AU SONDAGE DANS LES ENQUETES DEMOGRAPHIQUES RETROSPECTIVES

64-3398

TABLE DES MATIERES

I. INTRODUCTION

1. Bien que le dénombrement complet, soit de la population dans un recensement, soit des événements d'état civil (naissances, mariages, décès, etc.) à l'aide d'un système d'enregistrement demeure le véritable objectif, "on peut mettre les techniques de sondage en oeuvre avec succès au stade du dénombrement ... pour réduire le coût et l'envergure des opérations que comporte un recensement, tout en recueillant, dans le même temps, un maximum de données du genre de celles que l'on souhaite se procurer et de la meilleure qualité possible"[1/].

2. C'est pourquoi dans les pays en voie de développement qui n'ont pas de service de l'état civil ou dans lesquels ce service est insuffisant, les enquêtes rétrospectives par sondage sont actuellement un des moyens possibles d'obtenir une évaluation sûre des taux du mouvement de la population[2/]. Dans les pays plus développés on doit aussi, pour obtenir des renseignements sur la consommation des ménages, la morbidité et l'emploi (sauf dans les pays à économie planifiée), effectuer des enquêtes rétrospectives.

1/ Principales applications des sondages aux programmes de recensement de la population, Commission économique des Nations Unies pour l'Afrique, E/CN.14/ASPP/L.9

2/ Parmi les autres méthodes, citons les registres de population, les secteurs d'inscriptions par sondage, l'observation permanente, qui peuvent être employées séparément ou que l'on peut combiner. Voir Méthodes et problèmes d'inscription et de rassemblement des statistiques d'état civil en Afrique. E/CN.14/CAS.3/8

3. Grâce aux progrès accomplis dans la théorie et dans la pratique des enquêtes par sondage on peut, en connaissant les fonctions du coût et de la variance, mettre au point l'enquête par sondage et réduire au minimum les erreurs de sondage pour un coût donné (nombre d'enquêteurs, nombre d'unités de sondage, etc.). Lorsque ces fonctions ne sont pas connues, on peut les obtenir au moyen d'une enquête pilote.

4. Cependant, certaines enquêtes par sondage et plus souvent certains dénombrements ont donné des erreurs et des distorsions non dues au sondage qui ont parfois rendu sans valeur les résultats d'une enquête par sondage ou d'un dénombrement complet bien préparés.

5. On connaît des exemples de cas où l'inscription à l'état civil dans des pays africains n'a pas donné des résultats exacts parce que les erreurs et les distorsions non dues au sondage n'ont pas été diminuées[1/]. Pour la récente étude sur les tendances et les niveaux démographiques en Afrique, rédigée par le secrétariat de la Commission économique pour l'Afrique, exception faite des petites îles, seules la Tunisie (pour les naissances) et la République Sud-africaine (pour la population non Bantoue) ont fourni des taux d'enregistrement acceptables[2/]

6. On trouve des exemples d'erreurs et de distorsions non dues au sondage dans des enquêtes démographiques rétrospectives dans le recensement de la population africaine en Ouganda (1959). Le taux brut de mortalité évalué d'après le nombre des décès inscrit sur le bulletin de recensement était loin d'être exact et on a dû faire d'autres évaluations par des méthodes analytiques[3/]. On en trouve d'autres dans

---

1/ Voir Méthodes et problèmes d'inscription et de rassemblement des statistiques d'état civil en Afrique, op. cit.

2/ Tendances et niveaux démographiques récents en Afrique dans Bulletin économique des Nations Unies pour l'Afrique (sous presse). Pour la comparaison des taux de natalité, de mortalité et de mortalité infantile tirés des inscriptions à l'état civil et d'autres sources, voir Méthodes et problèmes d'inscription et de rassemblement des statistiques d'état civil en Afrique, op. cit.

3/ Uganda Census 1959, African Population, Gouvernement de l'Ouganda, 1961, Appendice XII.

l'enquête sur la fécondité et la mortalité menée en 1951-1952 par le Gokhale Institute of Politics and Economics de Poona, dans le district de Poona en Inde, où toutes les données sur la mortalité étaient à ce point inexactes qu'elles ont dû être rejetées[1/]. Il ne s'agit pas de cas isolés d'erreurs et distorsions non dues au sondage dans les statistiques démographiques des pays en voie de développement qui ne disposent que de statistiques de l'état civil défectueuses.

7. Récemment, on a accordé une plus grande attention au problème des erreurs et des distorsions non dues au sondage[2/] particulièrement dans les enquêtes démographiques[3/]. L'Union internationale pour l'étude scientifique de la population a consacré aux "Nouvelles méthodes d'observation du mouvement naturel de la population dans les pays sous-développés", une réunion qui s'est tenue à New York en 1961[4/] et à l'"Evaluation des taux du mouvement de la population d'après des données limitées" une autre réunion qui s'est tenue à Ottawa en 1963. Etant donné l'importance d'une étude des différents types d'erreurs dans les enquêtes démographiques rétrospectives, qui ont fourni ou fourniront dans l'avenir la base pour l'estimation des taux du mouvement de la population dans de nombreux pays africains, la troisième Conférence des statisticiens

---

1/ V.M. Dandekar et K. Dandekar, Survey of Fertility and Mortality in Poona District, Publication No.27 (Gokhale Institute of Politics and Economics, Poona, 1953).

2/ Pour une liste complète des études sur les erreurs des données voir S.S. Zarkovich, Sampling methods and Consuses, Vol II, Quality of Statistical Data (F.A.O. 1963)

3/ Pour des études sur l'Afrique, voir Problèmes de démographie en Afrique, Colloque de Paris, 20 au 27 août 1959 (Union internationale pour l'étude scientifique de la population, Paris, 1960); F. Lorimer, et M. Karp (éd.), Population in Africa, Report of a Seminar held at Boston University (Boston University Press, 1960); F. Lorimer, Demographic Information on Tropical Africa (Boston University Press, 1961). Voir aussi les communications suivantes présentées à la réunion de l'Union internationale pour l'étude scientifique de la population (Ottawa, 21-23 août 1963) : R. Blanc, L'analyse des données actuelles sur la fécondité africaine, utilisées pour une évaluation du mouvement de la population en Afrique au Sud du Sahara et F. Lorimer, Possibilities and Problems in the Estimation of Vital Rates in Africa.

4/ Congrès international de la population, New York, 1961, Tome II (Union internationale pour l'étude scientifique de la population, Londres, 1963)

africains, tenue en 1963, a recommandé l'organisation d'un cycle d'études qui s'attacherait particulièrement à l'étude des problèmes du rassemblement des statistiques de l'état civil en Afrique[1/].

1/ Commission économique des Nations Unies pour l'Afrique, Rapport de la troisième conférence des statisticiens africains (E/CN.14/CAS.3/22).

II. TYPES D'ERREURS ET DE DISTORSIONS DANS LES DONNEES RECUEILLIES

8. Erreurs et distorsions dans les données recueillies. On peut généralement les classer comme suit : i) erreurs de sondage dans les données obtenues par sondage parce que tout "l'univers" des unités constitutives n'est pas observé et ii) erreurs et distorsions non dues au sondage qui se produisent dans tout rassemblement de données, que ce soit par sondage ou par dénombrement complet (recensement) et dont l'ampleur est liée à l'habilité de l'enquêteur et au temps à consacrer à l'entrevue (dans une enquête par entrevues), à la qualité du bulletin et aux connaissances et à la compréhension des enquêtés. Aux erreurs et aux distorsions dans les données recueillies, peuvent s'en ajouter d'autres au stade de l'exploitation.

9. Précision et exactitude. Il est commode de distinguer la précision et l'exactitude des données. La différence entre les résultats d'une enquête par sondage et ceux d'un dénombrement complet, menés dans les mêmes conditions générales (avec ou sans contrôle) donne la précision des données obtenues par sondage et se mesure par l'erreur de sondage. La différence entre le résultat d'une enquête, qu'elle soit par sondage ou par dénombrement complet, et la valeur vraie donne l'exactitude des données. "C'est l'exactitude d'une enquête qui nous intéresse; le plus souvent nous ne pouvons mesurer que la précision".[1/] Or, il est d'habitude difficile de mener une enquête par sondage et un dénombrement complet dans les mêmes conditions générales. L'échelle des opérations étant relativement petite, l'enquête par sondage permet un meilleur contrôle du rassemblement et de l'exploitation des données, par le choix d'enquêteurs plus compétents, par une formation intensive et par des entrevues plus poussées.

1/ M.H. Hansen, W.N. Hurwitz et W.G. Madow, Survey Methods and Theory, Vol. I, Methods and Applications (John Wiley & Sons, Inc., New York, 1953), page 10.

10. Des erreurs et des distorsions non dues au sondage peuvent venir de plans de sondage, notamment de méthodes d'évaluation défectueuses (par exemple, la distorsion inhérente à l'estimation par la méthode des quotients) dans une enquête par sondage.

11. Dans le présent document, les erreurs et les distorsions non dues au sondage sont étudiées ensemble, bien qu'il soit possible de les distinguer. Les erreurs dont on peut penser que la somme algébrique sera nulle lorsque le nombre d'unités sera suffisamment grand peuvent être dites erreurs non dues au sondage: sans changer les évaluations si ce n'est au sens statistique, elles augmentent la variabilité ou l'erreur de sondage des évaluations. Les distorsions sont un autre type d'erreurs, qui ne s'annullent pas et modifient les évaluations elles-mêmes sans changer nécessairement la variabilité des évaluations[1/].

12. Réduction des erreurs de sondage. Grâce aux progrès de la théorie des enquêtes par sondage, on peut réduire et mesurer les erreurs de sondage dans une enquête par sondage convenablement préparée en choisissant une bonne stratification, et des probabilités de tirage des unités de sondage aux différents degrés et en utilisant des formules appropriées pour l'évaluation[2/] reliant tous ces éléments à la variable importante

---

1/ Ainsi, dans un recensement ou dans un sondage, si les différences qui tiennent aux enquêteurs s'annullent, on peut n'y voir que des erreurs non dues au sondage, mais elles ne s'annullent pas, on a affaire à une distorsion très nette commune à tous les enquêteurs et on peut dire qu'il s'agit de distorsions non dues au sondage (voir Annexe, paragraphes A8 - A9).

2/ Voir Hansen, Hurwitz et Madow (1953) op.cit.; W.G. Cochran, Sampling Techniques, 2e éd. John Wiley & Sons, Inc., New York, 1963; P.V. Sukhatme, Sampling Theory of Surveys with applications (Indian Society of Agricultural Statistics, New Delhi, et Iowa State College Press, Iowa, 1953); F. Yates, Méthodes de sondage pour recensements et enquêtes (Paris, Masson, 1949).

à évaluer, taux de natalité et taux de mortalité dans une enquête démographique rétrospective.

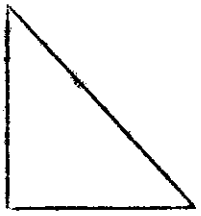
13. Erreurs de sondage et erreurs et distorsions non dues au sondage

L'erreur globale d'une évaluation résultant d'une enquête (qu'elle soit par sondage ou par dénombrement complet) est donnée par la formule suivante :

Erreur quadratique moyenne de l'estimation
= variance d'échantillonnage de l'estimation + carré de la distorsion dans l'estimation.

Dans un dénombrement complet, la variance d'échantillonnage n'existe pas, mais l'élément de distorsion peut exister[1/]. La relation ci-dessus peut être représentée par un triangle rectangle dont l'hypoténuse est la racine carrée de la moyenne de l'erreur quadratique moyenne :

(Variance d'échantillonnage)



(Erreur quadratique moyenne)

Distorsion

14. Pour avoir une idée de la grandeur des erreurs de sondage et des erreurs et distorsions non dues au sondage, on peut se reporter aux enquêtes démographiques dans les régions rurales de l'Inde, effectuées par le Gouvernement de l'Inde en 1953-54 par l'intermédiaire du National Sample Survey[2/], et à l'enquête démographique dans l'ancien

---

1/ Les lacunes d'enregistrement et les comptages multiples dans les recensements sont des exemples de distorsions non dues au sondage. R.K. Som, A.K. De, N.C. Das, B.T. Pillai, H. Mukherjee, S.M.U. Sarma.

2/ Preliminary Estimates of Birth and Death Rates and of the Rate of Growth of Population, National Sample Survey Report, No. 48 (Government of India, 1961), Appendice 3.

Etat de Mysore menée conjointement par le Gouvernement indien et l'Organisation des Nations Unies en 1951-52[1]. Les taux de natalité et de mortalité observés ont été corrigés pour fournir des estimations des taux réels et les erreurs de sondage dans les taux enregistrés ont été calculées. Les résultats ainsi que les estimations de l'erreur globale sont indiqués au Tableau I. On peut y voir quelle est l'erreur globale par rapport à l'erreur de sondage[2].

15. Ces chiffres montrent que la connaissance des erreurs de sondage n'a pas grande utilité lorsque les données contiennent des erreurs et des distorsions considérables non dues au sondage. Des taux du mouvement de la population certainement trop faibles ou trop élevés peuvent résulter d'erreurs de sondage relativement peu importantes et il faut être sur ses gardes. Si seules les erreurs de sondage entraient en ligne de compte, on n'hésiterait pas à recommander la méthode du dénombrement complet.

1/ Nations Unies, The Mysore Population Study. Etude démographique No 34, No de vente : 61.(XIII.3).

2/ Voir les paragraphes A3 - A5 de l'annexe, où ces erreurs sont étudiées selon la théorie des probabilités.

TABLEAU I

Taux observés et taux ajustés de natalité et de mortalité, erreur de sondage et erreur globale: régions rurales de l'Inde (1953 - 54) et Etat de Mysore, Inde (1951 - 52)

Enquête	Naissances par 1000 personnes					Décès par 1000 personnes				
	Taux observé	±	Erreur de sondage	Taux ajusté	Erreur globale b/	Taux observé	±	Erreur de sondage	Taux ajusté	Erreur globale b/
(1)	(2)		(3)	(4)	(5)	(6)		(7)	(8)	(9)
Inde: régions rurales (1953-54) a/	34,6	±	1,0	40,9	6,4	16,6	±	1,1	24,0	7,5
Mysore, Inde (1951-52):										
Zone rurale I	40,4	±	2,72	44,4	4,84	13,3	±	1,87	15,1	2,60
" " II	41,6	±	2,55	44,7	4,01	15,7	±	1,84	18,9	3,69
" " III	37,5	±	2,40	39,9	3,39	17,6	±	1,29	18,6	1,63
Centres urbains	36,2	±	2,84	38,9-39,8	3,92-4,58	12,4	±	1,52	13,8-23,9	2,07-11,60
Villes	30,0	±	1,27	33,0	3,26	8,1	±	0,75	13,7	5,65

a/ Les taux d'inscription pour l'ensemble de l'Inde en 1952-1954, soit 24,2/1000 pour les naissances et 13,6/1000 pour les décès, étaient inférieure aux taux obtenus par l'enquête.

b/ Racine carrée de l'erreur quadratique moyenne.

III. CORRECTIONS D'ERREURS ET DE DISTORSIONS NON DUES AU SONDAGE

16. Méthodes utilisées. Les méthodes généralement employées pour déceler des erreurs et des distorsions non dues au sondage et pour les éliminer sont les suivantes: répétition de l'enquête, sous-échantillons superposés, méthodes analytiques, y compris les analyses rétrospectives et contrôle externe[1].

17. Répétition de l'enquête. Une méthode classique pour la mesure et la correction des erreurs et des distorsions non dues au sondage dans un recensement ou dans une enquête par sondage est la répétition de l'enquête ("double sondage") sur un sous-échantillon des unités prises pour la première enquête, à l'aide du même bulletin (ou, si possible, un bulletin plus détaillé) et, dans les entrevues directes, à l'aide d'un personnel plus qualifié. Cette répétition de l'enquête peut prendre la forme d'une vérification des échantillons. Cette méthode est utilisée presque toujours pour les recensements. Elle peut aussi faire partie intégrante de l'enquête originale. C'est le cas de l'enquête démographique par sondage en Guinée (1954-55) dans laquelle, en outre, chaque fois que le taux de mortalité infantile obtenu était jugé beaucoup trop bas, une équipe médicale interrogerait à nouveau les femmes[2]. Dans le recensement de la population par sondage effectué au Soudan en 1955-56, environ un septième des ménages ont été dénombrés une deuxième fois par les institutrices ou par d'autres enquêteuses de contrôle dans six des neuf provinces et à la suite de cette enquête de contrôle on a pu mesurer le degré d'inexactitude des premières déclarations et corriger les premiers résultats[3].

1/ R.K. Som On adjustments for Non-Sampling Errors and Biases in the Estimation of Vital Rates, Session de l'Union internationale pour l'étude scientifique de la population, Ottawa, 21 - 23 août 1963.

2/ Etude démographique par sondage, Guinée, 1954-1955, 1ère partie, Techniques d'enquête (Ministère de la France d'Outre-Mer, 1956).

3/ First Population Census of Sudan, 1955/56, Methods Report, Vol.I, (République du Soudan, Bureau de statistique, 1960).

18. Sous-échantillons superposés. La méthode des sous-échantillons superposés consiste à subdiviser le nombre total des unités de l'enquête (enquête par sondage ou recensement) en un certain nombre de groupes au hasard. Elle permet par le procédé de Fisher d'analyse de la variance, d'étudier les différences qui résultent des enquêteurs notamment (variations dans les réponses inscrites dans les bulletins de recensement, dans les méthodes de rassemblement)[1/]. On s'est servi de cette méthode dans les enquêtes démographiques par sondage menées en Rhodésie du Sud en 1948 et en 1953-1955. Elle est utilisée dans l'enquête nationale par sondage en Inde depuis 1951 et dans l'enquête statistique sur les ménages aux Philippines (1957).

19. Dans les enquêtes démographiques par sondage sur la population africaine effectuées en Rhodésie du Sud en 1948[2/], on a pris des échantillons superposés pour vérifier l'exactitude des opérations sur le terrain. On a confié à chaque enquêteur un sous-échantillon complet dans trois districts différents pour pouvoir comparer son travail à celui d'au moins trois autres enquêteurs en vérifiant si les deux ou trois sous-échantillons venaient de la même population. "Dans la pratique, les vérifications n'ont fait apparaître aucune différence notable pour la population totale; sur plus de 70 sous-échantillons trois seulement ont été rejetés pour les taux de natalité et de mortalité infantile mais dix-huit l'ont été pour les taux de mortalité".[3/]

1/ P.C. Mahalanobis, Recent Experiments in Statistical Sampling in the Indian Statistical Institute réimpression par la Asia Publishing House, Bombay, 1960 (Nations Unies, Sous-Commission des statistiques de sondage, deuxième session)

2/ Report on the Demographic Sample Survey of the African Population of Southern Rhodesia, The Central African Statistical Office, Salisbury, 1951.

3/ Communication personnelle de M. C.A.L.Myburgh, Directeur du Bureau des recensements et des statistiques de la Rhodésie du Sud.

20. Dans l'enquête démographique par sondage de la population indigène africaine menée en Rhodésie du Sud en 1953-1955, on a utilisé des échantillons superposés pour déterminer les erreurs de sondage[1]. D'après les données de base, nous avons calculé la variance des données sur la natalité et sur la mortalité de six districts pour l'année 1953 (voir Tableau 2). Les différences des données de dix sous-échantillons n'étaient pas importantes sur le plan statistique et montraient que l'enquête était statistiquement bien conduite. Cependant, la variation des données sur la natalité et la mortalité des six districts était notable.

TABLEAU 2

Analyse de la variance du nombre des naissances et des décès
Enquête démographique par sondage de la population indigène africaine en Rhodésie du Sud, 1953[a]

(6 districts x 10 sous-échantillons)

Source	Degrés de liberté	naissances		décès	
		Carré moyen	Rapport F	Carré moyen	Rapport F
1. districts	5	123105,95	15,48[b]	48184,43	19,45[b]
2. sous-échantillons	9	7531,81	1,06[c]	2050,44	1,21[c]
3. erreur	45	7950,03	-	2476,77	-
4. total	59	17645,04	-	6285,27	-

a/ Les données de base ont été communiquées par M. C.A.L. Myburgh, Directeur du Bureau des recensements et des statistiques de la Rhodésie du sud.

b/ Significatif au niveau de 0,1 pour 100.

c/ Erreur/sous-échantillon dans le rapport F.

1/ The 1953-1955 Demographic Sample Survey of the Indigenous African Population of Southern Rhodesia (The Central African Statistical Office, Salisbury, 1959).

21. Il est également possible, par l'emploi du plus petit coefficient de Student, de vérifier l'importance de la plus petite différence entre deux enquêteurs[1/] : si le plus petit rapport des variances est significatif, on peut supposer que les deux enquêteurs se sont entendus.

22. Périodes rétrospectives. Dans les recensements et les enquêtes démographiques utilisant la méthode des recherches rétrospectives, des questions sont posées au sujet d'évènements du passé proche ou lointain. On peut penser que les réponses données par les personnes enquêtées, seront influencées par des pertes de mémoire et par d'autres causes de distorsion. Ces distorsions dépendent souvent de la période rétrospective, c'est-à-dire du temps qui s'est écoulé entre l'évènement et la date de l'entrevue. Ces distorsions, qui peuvent être exprimées en fonction de la période rétrospective, sont des erreurs de rétrospection. La mise en tableau par périodes rétrospectives ou d'après un autre élément (des études de cohortes de mariages dans les études historiques de la fécondité par exemple) révèle souvent l'existence d'erreurs de rétrospection qui pourraient autrement passer inaperçues dans les estimations globales. Des erreurs de rétrospection ont été relevées ou présumées dans de nombreux cas, aussi bien dans les pays développés que dans les pays sous-développés, se rapportant à des enquêtes démographiques, notamment sur la morbidité[2/] et sur la consommation des ménages. Etant donné leur importance, les erreurs de rétrospection seront étudiées plus en détail dans les Sections IV et V de la présente étude.

1/ K.V. Ramachandran. On the Studentized Smallest Chi-Square dans Journal of American Statistical Association, Vol. 53, No.284 (1959), p.868-872.

2/ Pour une étude des erreurs de rétrospection dans les enquêtes sur la morbidité, W.P.D. voir Logan et E.M. Booke, The Survey of Sickness, 1943-52 (General Register Office, Londres, 1957); F.E.Linder, National Health Surveys as a Source of Morbidity Data for Genetic and Radiation Studies dans The Use of Vital and Health Statistics for Genetic and Radiation Studies, Proceedings of the Seminar sponsored by the United Nations and the World Health Organization, Genève, 1960, Organisation des Nations Unies, 1962, A/AC.82/Seminar (No. de vente : 61.XVII.8).

IV. LES ERREURS DE RETROSPECTION DANS LES DONNEES COURANTES D'ETAT CIVIL

23. Généralités. Nous avons vu que les données obtenues au moyen d'enquêtes rétrospectives (qu'elles soient par sondage ou par dénombrement complet) contiennent souvent des erreurs et des distorsions d'un type particulier, non dues au sondage, dites erreurs de rétrospection, lorsque les erreurs et les distorsions non dues au sondage peuvent être exprimées en fonction de la période rétrospective (intervalle entre la date de l'événement d'état civil et la date de l'entrevue). Des erreurs et des distorsions de ce type non dues au sondage donnent lieu généralement à des omissions dans l'enregistrement du nombre total des naissances pendant toute la période de reproduction, bien que le taux relatif à une période déterminée, une année par exemple, puisse être aussi bien trop élevé que trop faible, ce qu'il est le plus souvent.

24. Dans la présente section, nous examinerons d'une manière générale la méthode élaborée récemment pour expliquer, tout d'abord, le phénomène de l'erreur de rétrospection dans les données "courantes" d'état civil relatives à une période rétrospective (de référence) d'une ou de deux années, et pour corriger, ensuite, l'erreur de rétrospection afin d'obtenir une approximation plus exacte des taux "réels". On trouvera à l'Annexe[1/] l'exposé théorique de cette méthode. Elle s'applique aussi bien aux cas d'omissions d'enregistrement qu'aux comptages multiples. Même lorsqu'il semble ne pas y avoir d'omissions ou de comptages multiples pour la période considérée, l'analyse par périodes rétrospectives peut faire apparaître des contradictions dans les données.

25. La méthode d'analyse par périodes rétrospectives. On suppose que dans l'enquête (par sondage ou par recensement) les unités (de sondage) sont déplacées au hasard (ou uniformément) au cours de la période de l'enquête qui peut s'étendre sur une ou plusieurs années. Des données sont réunies sur les événements (naissances ou décès) qui ont eu lieu pendant la période rétrospective (de référence) d'une ou plusieurs années et des renseignements sont notés au sujet du jour ou tout au moins du mois pendant lequel l'événement a eu lieu. Supposons par exemple qu'une enquête est menée sur une période d'un an (janvier à

1/ R.K. Som, Recall Lapse in Demographic Studies, Congrès international de la population, Vienne 1959 (Union internationale pour l'étude scientifique de la population, 1959).

décembre 1964) et que l'on prend pour période rétrospective (de référence) les douze mois précédents ("année précédente").

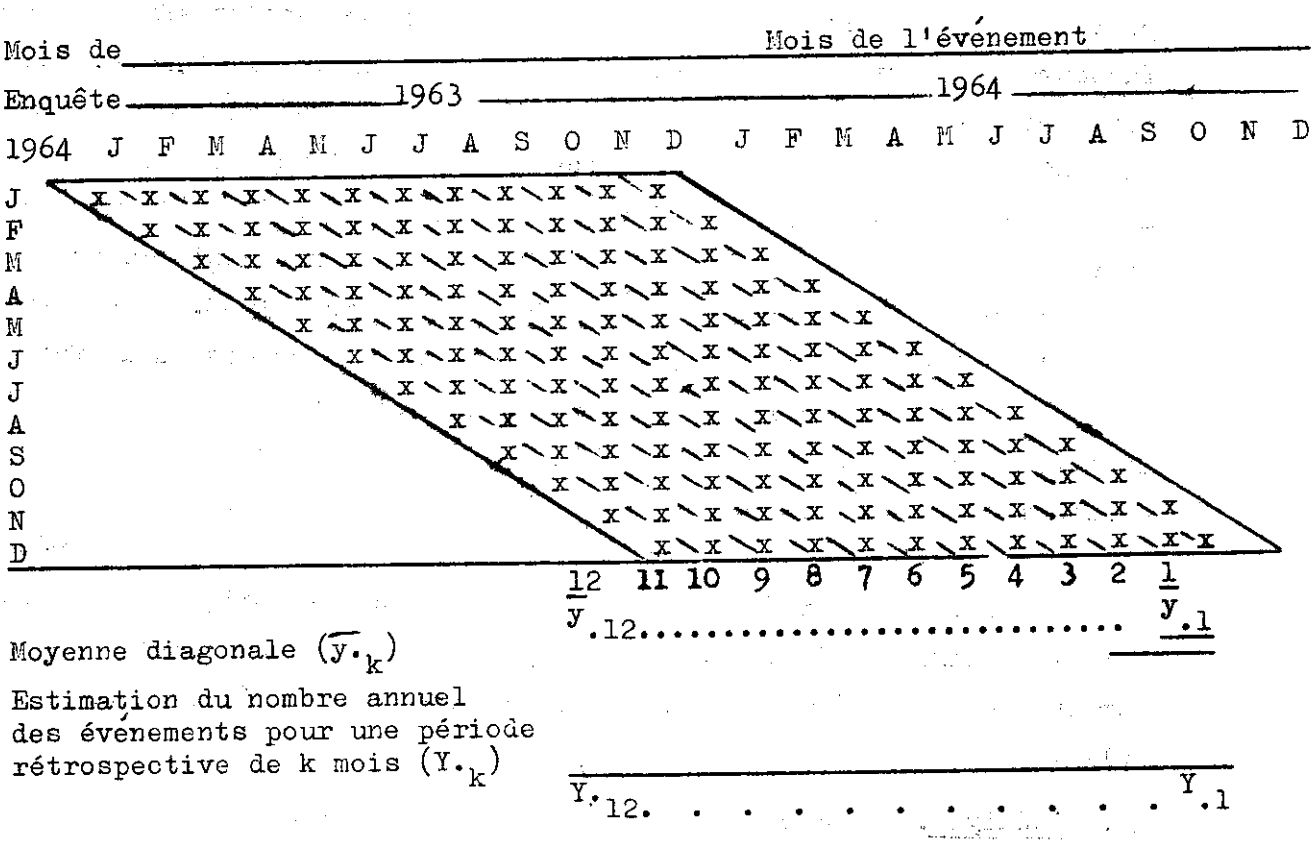


Fig. 1. Représentation schématique

26. Afin de simplifier l'exposé et sans nuire à son caractère général, supposons que l'enquête a lieu le premier jour de chaque mois. Les croix représentent le nombre des événements déclarés au moment de l'enquête entreprise au cours d'un mois quelconque comme ayant eu lieu pendant un autre mois. (Pour un dénombrement complet, on note le nombre réel des événements; dans une enquête par sondage, on devra étendre à l'univers le nombre des événements de l'échantillon au moyen de coefficients appropriés pour obtenir une estimation du nombre total des événements). Ainsi, la première croix à droite représente le nombre total (estimé) des événements déclarés comme ayant eu lieu pendant le mois de décembre 1963, l'enquête ayant lieu en janvier 1964.

27. Elimination des facteurs saisonniers. Il importe que la durée de l'enquête soit d'une ou de plusieurs années et que la période rétrospective (de référence) soit d'une ou de plusieurs années afin que les facteurs saisonniers soient éliminés. Si on sait, par exemple, que les facteurs saisonniers donnent un plus grand nombre d'événements au cours des mois d'octobre, novembre et décembre et si l'enquête ne commence qu'au mois de janvier 1964 (comme c'est le cas dans le schéma d'enquête ci-dessus et s'il apparaît que le nombre des événements déclarés est plus élevé pour les mois d'octobre, novembre et décembre 1963, on ne saurait dire s'il s'agit de l'influence des facteurs saisonniers ou du fait que les événements sont chronologiquement moins éloignés du moment de l'enquête. En d'autres termes, l'élément saisonnier, dont on sait qu'il a des effets marqués sur les événements d'état civil, se confondra inextricablement avec l'erreur de rétrospection. Si, d'autre part, l'enquête est étalée sur une ou plusieurs années, le cadre théorique des opérations permet d'éliminer l'effet saisonnier, car la somme des effets saisonniers sur une année entière (ou sur plusieurs années) peut être considérée comme nulle.

28. L'analyse. Dans la Figure 1, la somme des croix comprises dans le trapèze délimité par des lignes pleines, représente le nombre total (estimé) d'événements pour une période rétrospective d'une année, multiplié par 12. La dernière croix de la première ligne représente le nombre des événements déclarés comme ayant eu lieu au mois de décembre 1963, l'enquête ayant commencé en janvier 1964, et correspond, par conséquent à une période rétrospective d'un mois. Il en est de même de la dernière croix de chaque ligne, ces croix étant séparées des autres par une ligne de traits. Leur somme, divisée par 12 (nombre des lignes) donne le nombre moyen estimé d'événements par mois pour le premier mois de la période rétrospective ($\bar{y}_{.1}$). En multipliant ce résultat par 12 ($=12 \div 1$), on obtient le nombre annuel estimé d'événements pour une période rétrospective d'un mois ($Y_{.1}$). De même, la deuxième croix de chaque rangée à partir de la droite correspond au deuxième mois de la période rétrospective. Leur somme en diagonale, divisée par 12 donne le nombre moyen estimé d'événements par mois pour le deuxième mois de la période rétrospective. ($\bar{y}_{.2}$). La somme des nombres moyens estimés

d'événements par mois pour les périodes rétrospectives du premier et du deuxième mois, divisée par 2, donne le nombre moyen estimé d'événements par mois pour la période rétrospective de deux mois. Si on multiplie ce chiffre par 12, on obtient le nombre annuel estimé d'événements pour une période rétrospective de deux mois (Y.2). En continuant ainsi jusqu'à la kème période rétrospective (k = 1,2,...,12) nous obtenons le nombre annuel estimé d'événements pour une période rétrospective de k mois ($Y._k$) en calculant la moyenne de la moyenne des diagonales jusqu'au kème mois précédent inclusivement et en multipliant le chiffre obtenu par 12. Ayant calculé les événements pour une période rétrospective de 12 mois, en prenant la moyenne de toutes les moyennes diagonales et en la multipliant par 12, nous obtiendrons le nombre total annuel des événements pour toute la période de référence ($Y._{12}$). Ces nombres annuels estimés d'événements obtenus pour les différentes périodes rétrospectives d'un mois, de deux mois, ..., de 12 mois ($Y._1$, $Y._2$, ..., $Y._{12}$) sont des estimations d'une même valeur, à savoir du nombre annuel d'événements. Ces estimations, mises en tableau et/ou portées sur un graphique, permettent une excellente vérification de cohérence interne, car si il n'y avait aucune erreur de rétrospection, toutes ces estimations seraient identiques, dans les limites permises par l'échantillonnage (si les données sont obtenues par une enquête par sondage). Le nombre estimé d'événements pour les différentes périodes retrospectives, divisé par la population totale (estimée), donne les taux estimés du mouvement de la population (par exemple, taux de natalité et de mortalité) pour les différentes périodes rétrospectives.

29. Ajustement des résultats pour tenir compte des erreurs de rétrospection.

Lorsqu'on a constaté l'existence d'une erreur de rétrospection au moyen de l'analyse expliquée ci-dessus, pour obtenir une meilleure approximation du nombre annuel réel des événements d'état civil et des taux relatifs à ces événements, on peut, pour tenir compte de l'erreur de rétrospection, ajuster les résultats comme suit. On lisse les $Y._k$ (estimés) en choisissant une courbe appropriée : $Y._k = f(k)$ à partir de laquelle on peut calculer $\hat{Y}._o = f(k)$, qui correspond à la période rétrospective "0"

et qui peut être considéré comme étant égal à la valeur réelle à supposer qu'aucune erreur systématique n'existe. Quantitativement, $\hat{Y}_o$ représenterait le nombre annuel probable (ou supposé) d'événements si les événements étaient enregistrés au moment même où ils se produisaient (période de rétrospection nulle), l'erreur de rétrospection n'intervenant pas.

30. Effets de décalage. Il existe deux sources importantes d'erreurs qui tiennent aux réponses obtenues : 1) la non déclaration d'un événement, c'est-à-dire l'omission complète, qui donne lieu à des erreurs par défaut, quelle que soit la durée de la période de référence (rétrospective) et 2) l'indication inexacte de la date des événements, qui donne lieu à des erreurs par excès pour une période de référence, à des erreurs par défaut pour une autre, mais ne fausse pas le nombre total d'événements pour une période de référence non délimitée. Le second type d'erreur provient de la délimitation rigoureuse de la période de référence (rétrospective), ce qui produit un "effet de décalage". Cet effet de décalage agit non seulement aux extrémités de la période de référence (rétrospective), mais aussi à n'importe quel moment de la période de référence délimitée[1/].

31. La méthode exposée ci-dessus peut être utilisée pour compenser le premier type d'erreurs dues aux réponses, c'est-à-dire les erreurs par défaut qui peuvent être exprimées comme une fonction de la période rétrospective (Fig. 2a). En outre, si les erreurs dues à l'indication inexacte des dates auxquelles ont eu lieu les événements varient plus ou moins au hasard, ces erreurs pourraient s'annuler.

1/ Voir aussi R.Blanc, Manuel de recherche démographique en pays sous-développé (Institut national de la statistique et des études économiques, Paris, 1962) G. Théodore et R.Blanc, L'amélioration des statistiques des mouvements naturels de population en Afrique dans Congrès international de la population, New York, 1961, Tome II (Union internationale pour l'étude scientifique de la population, Londres, 1963).

32. Nous en venons maintenant à l'étude du déplacement net unidirectionnel appréciable dans l'indication des dates des événements, qui conduit à une surestimation ("effet de contraction") ou à une sous-estimation ("effet d'extension") pour une période de référence donnée, par exemple une année. A la suite de deux essais dans des pays africains, le premier en Guinée en 1955 et le second en Côte-d'Ivoire en 1955-56 et de la comparaison des résultats aux données obtenues au moyen d'observations continues, on a constaté que les taux obtenus au moyen d'enquêtes rétrospectives étaient trop élevés 1/, alors que, dans les autres pays africains et dans ceux des autres régions, ces taux sont le plus souvent trop faibles. Etant donné que l'on peut exclure la possibilité de déclarations d'événements faux, la seule explication dans les deux cas précités est la déclaration de certains événements extérieurs à la période d'une année (effet de contraction). Ainsi, alors que généralement l'estimation obtenue pour toute la période de référence de 12 mois est inférieure à la valeur réelle (Y), c'est-à-dire $Y._{12} < Y$, dans ces deux cas particuliers on a obtenu $Y._{12} > Y$.

33. Avant d'examiner plus en détail l'effet de contraction et l'effet d'extension, il convient de signaler un point important : même lorsque le nombre estimé d'événements pour une période de référence d'une année semble plausible, il peut s'agir d'un hasard, les effets des erreurs de rétrospection dues à des omissions d'enregistrement et à l'effet de décalage s'étant annulés.

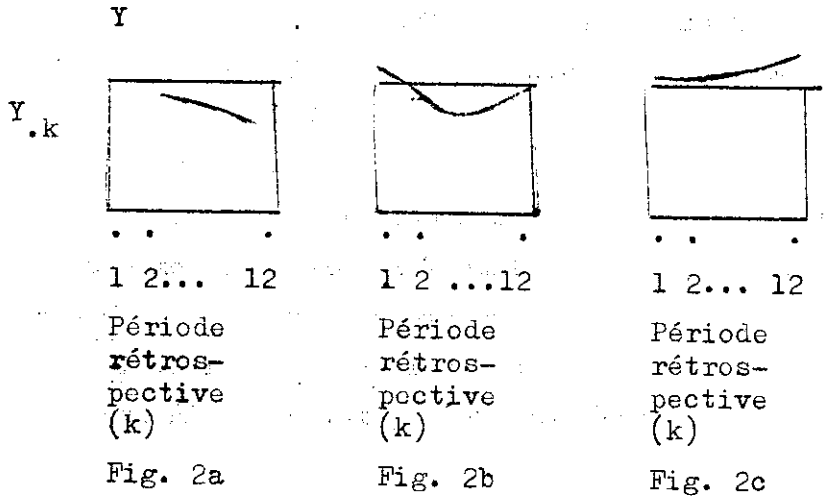
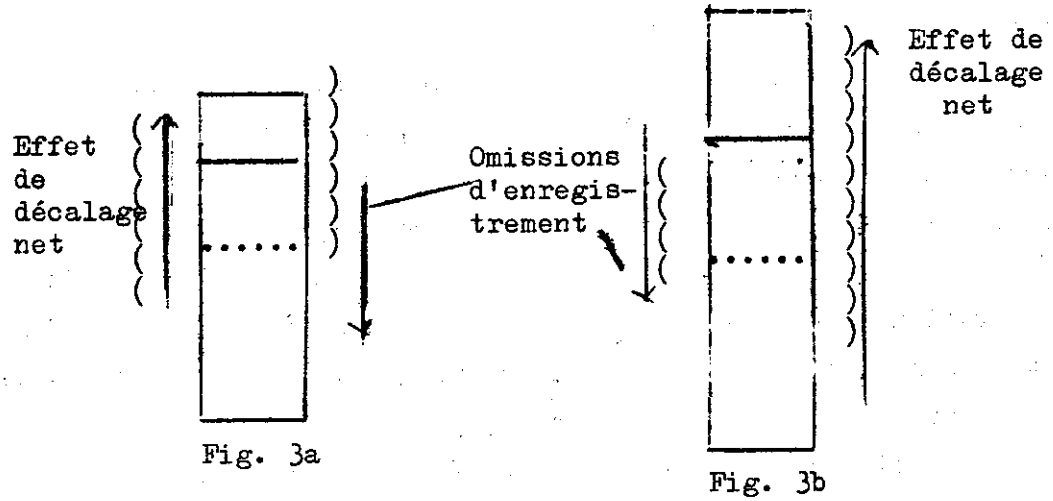


Fig. 2a Fig. 2b Fig. 2c

1/ Voir paragraphes 35 et 36

Considérons la figure 2b. Le nombre des événements pour une période rétrospective d'une année coïncide avec la valeur réelle (Y), mais l'analyse par périodes rétrospectives révèle une incohérence interne des données.



34. Examinons maintenant l'effet combiné des omissions d'enregistrement et des effets de décalage. Considérons d'abord la figure 3a. Pour le kème mois précédent, supposons que le rectangle délimité vers le haut par la ligne pleine représente le nombre réel d'événements. A cause d'omissions le nombre descend jusqu'à la ligne de points. Nous devons ajouter l'effet de décalage. Supposons d'abord que l'effet net de décalage (= report des mois précédents moins report aux mois suivants) élève le nombre des événements jusqu'à la ligne de traits mais ne suffit pas à le remonter au niveau réel (Fig. 3a). S'il en est ainsi pour tous les k et que le nombre des événements signalés peut s'exprimer comme une fonction de la période rétrospective, le mode d'analyse décrit ci-dessus peut s'appliquer, à supposer que la valeur estimée pour la période rétrospective "0" donnerait le nombre réel des événements (Fig. 2a). Si, en revanche, l'effet net de décalage fait plus que compenser les omissions du kème mois précédent, le nombre des événements observés étant cette fois encore représenté par la ligne de traits, le nombre estimé des événements est erroné par excès (Fig. 3b). Là encore, si

le nombre des événements observés peut s'exprimer comme une fonction croissante de k (période rétrospective), on pourrait, au moyen du mode d'analyse décrit plus haut, corriger les données, en supposant aussi, que pour la période rétrospective "0", l'effet net obtenu est nul (Fig. 2c). Cela est vrai également si le report d'événements d'une période à l'autre a lieu dans le sens contraire, c'est-à-dire s'il s'agit d'un effet d'extension au lieu d'un effet de contraction.

35. Exemples d'effet de contraction. Les essais faits en Guinée et en Côte-d'Ivoire fournissent des exemples d'erreurs par excès pour une période de référence d'une année. Les taux ont été calculés d'après les événements signalés au cours d'enquêtes rétrospectives sur le terrain et ont été comparés aux résultats obtenus à l'aide d'un système d'observations continues (ou répétées) des ménages. En Guinée, l'essai a duré de août 1955 à juillet 1956 et a porté sur trois cantons d'une population totale de 33.000 habitants. Trois enquêteurs ont effectué une série de visites dans les villages (10, 12 et 10 tournées respectivement) pour recueillir des renseignements sur les événements d'état civil[1/]. Nous avons comparé au Tableau 3 les résultats de cette enquête à ceux de l'enquête par sondage dans les régions rurales, effectuée en Guinée en 1955.

TABLEAU 3

Taux de natalité, de mortalité et de mortalité infantile: Enquête rurale par sondage, Guinée, 1955 et observations continues dans trois cantons, 1955 - 1956, Guinée

Taux de	Enquête rurale par sondage	Observations continues
Natalité par 1000 personnes	63	50
Mortalité par 1000 personnes	41	27
Mortalité infantile par 1000 enfants nés vivants	220	153

1/ Theodore et Blanc, op. cit., page 3; Lorimer, Frank (1961), op. cit., page 38.

36. En Côte-d'Ivoire, l'expérience a été faite à Bongouanou en 1955-1956. On a recueilli des renseignements sur un échantillon de dix mille habitants dans neuf villages, au sujet des naissances et des décès qui avaient eu lieu en 1953, 1954 et 1955. En outre, des enquêteurs résidents ont inscrit les événements pendant une période de 12 mois en 1955. Le Tableau 4 donne les taux bruts de natalité[1].

TABLEAU 4

Taux de natalité obtenus par l'enquête rétrospective, 1953 et 1954 et par l'observation continue; 1955, Côte-d'Ivoire

	Enquête rétrospective		Observation continue
	1953	1954	1955
Taux de natalité par 1000 personnes	54,3	55	49

En Guinée, les taux les plus bas résultant de l'observation continue semblent plus acceptables, mais la nature des erreurs de comptage multiple dans l'enquête rétrospective ne sera bien connue qu'après une exploitation plus poussée des données des différentes périodes rétrospectives.

37. Sources de distorsions. Quelle que soit la nature des distorsions (par défaut ou par excès), on doit en examiner les sources d'après le rapport de masculinité à la naissance et la structure de la mortalité par âge et par sexe pour chaque mois de la période de référence[2]. Une source importante de distorsion est l'omission des décès des enfants de moins d'un an et on pourrait examiner si les omissions dans l'indication des naissances et des décès viennent en totalité ou en partie des décès non signalés des enfants de moins d'un an, dont on peut aussi ne pas avoir compté la naissance, ce qui entraîne une diminution des taux de

1/ Enquête nutrition - niveau de vie, Subdivision de Bongouanou, 1955 - 1956 (Territoire de la Côte-d'Ivoire, Service de la statistique et de la mécanographie) p. 36 à 44 et 187.

2/ Méthodes et problèmes d'inscription et de rassemblement des statistiques d'état civil en Afrique, op. cit. paragraphe 75.

natalité et de mortalité. On en trouve un exemple dans les données obtenues à la suite de l'enquête nationale par sondage effectuée en Inde en 1953 - 1954. Le taux de natalité pour toute la période de référence (douze mois) était de 34,6 pour 1000 et le taux de natalité corrigée par la méthode décrite au paragraphe 29 ci-dessus était de 40,9 pour 1000. Ainsi, les naissances omises étaient de 40,9 - 34,3 = 6,6 pour 1000. Sur 1000 enfants nés vivants inscrits au cours des douze mois précédents, 111 ont été signalés comme étant décédés avant l'entrevue. Ainsi, la proportion des décès d'enfants de moins d'un an, pour une période moyenne de 6 mois, dans la cohorte des naissances des 12 mois précédents était de 111 pour 1000 enfants nés vivants, soit 0,111 x 34,3 = 3,8 pour 1000. Si les omissions estimées de naissances se rapportent uniquement à la mortalité infantile, c'est-à-dire si seules sont omises les naissances d'enfants qui mourront avant d'avoir atteint l'âge d'un an, l'estimation de la proportion des décès d'enfants de moins d'un an s'obtient en additionnant les naissances omises (= omissions des décès d'enfants de moins d'un an) et la proportion des décès d'enfants de moins d'un an inscrite, soit 3,8 + 6,6 = 10,4 pour 1000 habitants, soit (10,4/40,9) x 1000 = 254 pour 1000 naissances d'enfants vivants après ajustement. Il s'agit clairement d'une surestimation de la proportion d'enfants qui quittent une cohorte de naissances, mais sur une période moyenne de six mois. On peut alors supposer qu'un certain nombre d'enfants vivants au moment de l'enquête ont aussi été omis.

38. Exemple d'analyse rétrospective : L'enquête démographique par sondage effectuée en Haute-Volta en 1960-1961, a donné un taux brut de mortalité de 30,5 pour 1000[1/]. Un tableau des décès déclarés a également été construit pour les mois de l'enquête (d'octobre 1960 à avril 1961) et les mois des décès (de novembre 1959 à avril 1961) : la répartition

1/ La situation démographique en Haute-Volta : Résultats partiels de l'enquête démographique 1960-1961, République de Haute-Volta, Service des statistiques; République française, Ministère de la coopération, I.N.S.E.E., Service de coopération.

est reproduite au tableau 5. Bien que, pour éliminer l'élément saisonnier, il eût été théoriquement préférable que la période de sondage s'étende sur une année complète, les données recueillies sont néanmoins utilisées ci-après pour décrire la méthode à suivre en vue d'obtenir des taux corrigés des événements d'état-civil.

39. Les diagonales du tableau montrent le nombre des décès pour des périodes rétrospectives de un à 12 mois. Les totaux apparaissent également au tableau 6. On voit que, sur un total de 2769 décès, 271 se sont produits dans le mois précédant l'entrevue, 276 dans les deux mois précédant l'entrevue et ainsi de suite. En faisant la somme cumulative des chiffres (col.3 du Tableau 6), puis en la divisant par le nombre de mois de la période correspondante, on obtient le nombre mensuel moyen des décès déclarés pour les différentes périodes rétrospectives (col.4 du tableau 6).

40. S'il n'y avait pas d'erreur de rétrospection, toutes les moyennes mensuelles seraient statistiquement identiques, sous réserve des erreurs de sondage.

Les valeurs obtenues font penser qu'une fonction exponentielle du type

$$D_k = ae^{-bk^2}$$

correspondrait bien aux données. Dans cette fonction, D_k est le nombre mensuel moyen des décès, k est la période rétrospective (en mois) et a et b sont des paramètres à évaluer d'après les données. Par la méthode des moindres carrés, on trouve que

$$a = 252,27 \text{ et } b = -0,00071343$$

de sorte que $Dk = 252,27e^{-0,00071343k^2}$

En supposant k = 0 dans l'équation ci-dessus, on obtient $D_o = 252,27$, ce qui est le nombre mensuel moyen de décès déclarés pour une période de rétrospection nulle, c'est-à-dire dans l'hypothèse où l'enquêteur pourrait interroger le ménage au moment exact du décès, avant que l'erreur de rétrospection ait pu intervenir. A partir de cette valeur, le taux annuel corrigé de mortalité peut être estimé à

33,34 pour 1000[1/] alors que le taux brut était de 30,5 pour 1000. Ceci correspond à une omission de 9 pour 100 dans les données qui ont servi au calcul du taux établi d'après les déclarations.

1/ Pour le calcul du taux corrigé de mortalité, il y a lieu d'ajuster également la base de population pour tenir compte du nombre plus grand de décès et de naissances. On ne l'a pas fait faute d'une analyse équivalente des naissances.

TABLEAU 5

Répartition des décès déclarés,
d'après les mois de l'enquête et les mois des décès :
enquête démographique en Haute-Volta, 1960 - 1961[a]

Mois de l'enquête	Mois du décès																	
	1959		1960												1961			
	Nov.	Déc.	Jan.	Fév.	Mars	Avr.	Mai	Juin	Juil.	Août	Sept.	Oct.	Nov.	Déc.	Jan.	Fév.	Mars	Avr.
Octobre 1960	33	21	17	25	39	34	52	53	23	25	31	32						
Novembre 1960		27	10	24	49	58	78	59	35	29	33	42	30					
Décembre 1960			10	12	13	39	28	33	23	20	25	26	27	34				
Janvier 1961				15	43	28	61	28	29	33	33	33	25	37	35			
Février 1961					38	50	37	27	27	27	18	24	30	15	41	59		
Mars 1961						40	36	39	21	42	31	40	25	36	43	61	66	
Avril 1961							41		18	33	16	12	23	16	26	30	37	15
Totaux des chiffres des diagonales							204	207	176	255	238	289	248	206	202	197	276	271

a/ Données de base fournies par M. Pierre Cantrello, expert des Nations Unies en statistiques d'état civil, Sénégal.

TABLEAU 6

Analyse rétrospective des données sur la mortalité :
Enquête démographique en Haute-Volta, 1960-1961

Période rétrospective k (mois)	Nombre de décès a/	Nombre cumulatif de décès	Nombre mensuel moyen de décès D_k	Taux corrigé de mortalité pour 1000	Taux corrigé de mortalité Indice b/
(1)	(2)	(3)	(4)=(3)/(1)	(5)	(6)
1	271	271	271	33,3	99,9
2	276	547	273,5	33,2	99,8
3	197	744	248	33,1	99,4
4	202	946	236,5	33,0	98,9
5	206	1.152	230,4	32,8	98,3
6	248	1.400	233,3	32,5	97,5
7	289	1.689	241,3	32,2	96,6
8	238	1.927	240,9	31,9	95,6
9	255	2.182	242,4	31,5	94,4
10	176	2.358	235.8	31,0	93,1
11	207	2.565	233,2	30,6	91,8
12	204	2.769	230,8	30,1	90,2
Total	2.769	-	-	-	-

a/ Totaux des chiffres des diagonales du tableau 5.

b/ Le taux de mortalité pour une période de rétrospection nulle, soit 33,34 pour 1000, étant pris pour base = 100.

41. Une analyse similaire peut être effectuée pour les chiffres relatifs aux taux de natalité. Il est probable que le taux d'accroissement naturel de la population, qui est la balance des taux de natalité et de mortalité, prendrait également une valeur différente après ajustement de ces deux derniers taux.

42. D'autres exemples d'analyse devant servir à la correction des erreurs de rétrospection dans les données courantes d'état civil, sont fournis par l'enquête nationale effectuée en Inde.[1/] Un aperçu des résultats est donné au Tableau 7. Dans tous les exemples cités, il y a eu ajustement en hausse des taux établis d'après les déclarations; il est possible, toutefois, comme on l'a vu aux paragraphes 30 à 34, qu'un ajustement en basse de ces taux doive être opéré dans d'autres cas, après l'analyse par périodes rétrospectives.

TABLEAU 7

Taux bruts et taux corrigés (pour tenir compte des erreurs de rétrospection) dans certaines enquêtes démographiques

Enquête	Nature du taux	Taux pour 1000 personnes Brut	Taux pour 1000 personnes Corrigé	Indice d'omission
(1)	(2)	(3)	(4)	(5)=[1-(3)/(4)]x 100
Inde (zones rurales) (1953-1954)	Taux de natalité	34,3	40,9	16
	Taux de mortalité	16,6	24,0	31
Inde (zones rurales) (1955-1956)	Taux de natalité	32,0	38,9	18
	Taux de mortalité	11,4	20,8	45
Inde (zones urbaines) (1957-58)	Taux de natalité	30,5	36,8	17
	Taux de mortalité	8,5	13,7	38
Haute-Volta	Taux de mortalité	30,5	33,3	9

Source : Pour l'Inde : Som et autres auteurs, (1961), op.cit.; pour la Haute-Volta, Tableau 6 ci-dessus.

1/ Les premiers résultats de l'essai effectué au Pakistan indiquent une proportion d'omission plus grande encore que celle qui a été constaté en Inde. "On pouvait s'attendre que les énumérations basées sur les souvenirs d'un passé éloigné rendent compte d'un moins grand nombre d'événements d'état civil que les enregistrements d'événements du passé récent, mais la diminution (non indiquée dans la présente étude) du nombre des événements non oubliés semble être plus forte que ne le montre, par exemple, la courbe de Som". K.J. Krotki, First Report of the population Growth Estimation Experiment, (Réunion de l'Union internationale pour l'étude scientifique de la population, Ottawa, 21 - 23 août 1963).

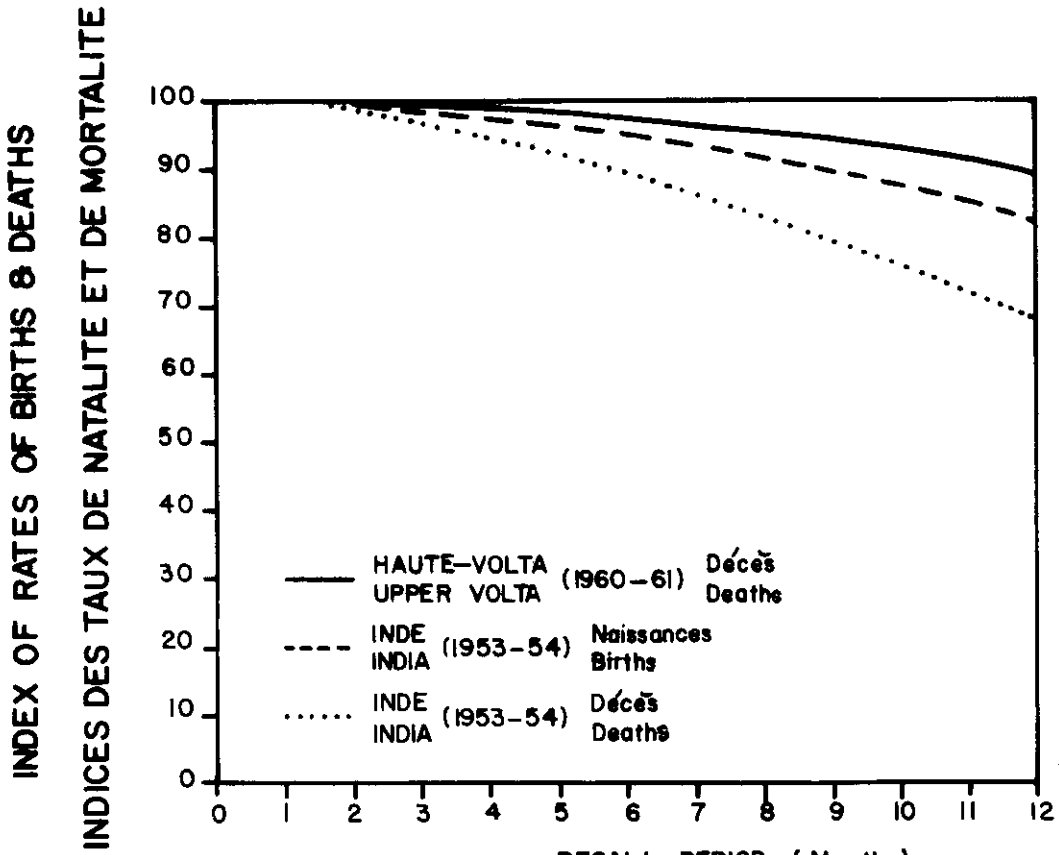


Fig. IV INDEX OF FITTED RATES BY RECALL PERIODS
INDICE DES TAUX AJUSTES PAR PERIODES DE RAPPEL

43. Les taux de natalité et de mortalité dans les zones rurales en Inde (1953-1954) et les taux de mortalité en Haute-Volta (1960-1961) pour les différentes périodes rétrospectives sont indiqués sur le graphique 4. Le "renflement" des courbes correspond à la période pour laquelle les taux demeurent à peu près constants, ce qui peut être considéré comme étant la période optimale pour la collecte des données d'état civil. Par exemple, pour le taux de mortalité, la période optimale dure un mois pour les zones rurales de l'Inde (1953-1954) et deux mois pour la Haute-Volta (1960-1961). Cela signifie non pas qu'on devra, dans les enquêtes futures, récueillir les données de la période optimale ainsi déterminée, mais qu'il faudra choisir une plus longue période de référence et analyser les taux par périodes rétrospectives.

44. Comparaison des taux relatifs aux événements déclarés survenus "l'année précédente" et les taux relatifs aux événements déclarés survenus "l'avant-dernière année". Dans l'enquête nationale par sondage effectuée en Inde pendant deux années consécutives, à savoir 1958-1959, (14ème "phase" de l'enquête nationale par sondage), et 1959-1960, des renseignements ont été recueillis sur les naissances et les décès survenus au cours des deux années précédentes. Durant la 15ème phase, la période de référence précédemment appelée "l'année précédente" est devenue "l'avant-dernière année", de sorte que les deux phases ont procuré deux estimations indépendantes correspondant une même période d'une année (la 14ème et la 15ème phases) : il a ainsi été possible de faire une comparaison permettant d'apprécier le degré de confiance des données courantes d'état civil que l'on pouvait espérer atteindre. Les taux de natalité et de mortalité tirés des données réunies lors de la première partie de la 15ème phase (en zone rurale) sont présentés séparément au Tableau 8[1/] pour "l'année précédente" et pour "l'avant-dernière année", en même temps que les données obtenues à la 14ème phase. Pour la même période d'une année allant d'août 1957 à juillet 1958,

1/ Indian Statistical Institute, The Use of the National Sample Survey in the Estimation of Current Birth and Death Rates in India, Congrès international de la population, New York, 1961 (Union internationale pour l'étude scientifique de la population, Londres, 1963).

on a obtenu un taux de natalité de 38,7 pour 1000 dans le cas de la période rétrospective dite "l'année précédente" au cours de la 14ème phase, mais ce taux est tombé à 28,2 pour 1000 dans le cas de la période retrospective dite "l'avant-dernière année", lors de la 15ème phase. Ceci correspond à une sous-estimation de l'ordre de 27 pour 100 du taux de natalité de "l'avant-dernière année" par rapport à celui de "l'année précédente", pour la même période d'une année. Les taux de mortalité correspondants, soit 19,3 pour 1000 et 10 pour 1000 respectivement, font apparaître une sous-estimation de 38 pour 100.

TABLEAU 8

Taux de natalité et de mortalité pour 1000 durant "l'année précédente et "l'avant-dernière année": enquête nationale par sondage en milieu rural en Inde, première partie de la 14ème phase (juillet-août 1958 : 430 villages échantillons et 38.362 ménages) et première partie de la 15ème phase (juillet - août 1959 : 434 villages échantillons et 39.726 ménages).

Période de référence	Taux de natalité		Taux de mortalité	
	Période retrospective		Période retrospective	
	Avant-dernière année	Année précédente	Avant-dernière année	Année précédente
Août 1956 - Juillet 1957	31,6+		10,1+	
Août 1957 - Juillet 1958	28,2	38,7+	10,0	19,3+
Août 1958 - Juillet 1959		38,7		17,1
.....Indice : 100 x taux de "l'avant-dernière année" (15ème phase)/ taux de "l'année précédente" (14ème phase).....				
Août 1957 - Juillet 1958	72,9	100	51,8	100

+ D'après les renseignements recueillis lors de la 14ème phase; les taux non accompagnés d'astériques reposent sur des renseignements recueillis lors de la 15ème phase.

45. Enquêtes périodiques : le taux d'acroissement de la population, compte tenu des migrations, peut être déterminé directement au moyen d'enquêtes périodiques, dans lesquelles on cherche à déterminer ce que sont devenus tous les individus lors de l'enquête précédente : chaque sortie est due soit au décès, à l'émigration par suite de mariage, à la dissociation d'un ménage ou à d'autres causes et chaque entrée est due à la naissance, à l'immigration par suite de mariage, au regroupement d'un ménage ou à d'autres causes. Dans les enquêtes périodiques, il ya lieu de choisir des unités d'énumération de nature géographique et non des ménages pour qu'il soit possible de tenir compte des ménages disparus et de ménages nouvellement constitués. Les informations relatives aux événements d'état civil qui n'apparaissent pas dans les dossiers, par exemple celles qui concernent les enfants nés et morts durant la période considérée, peuvent être obtenues par le moyen de l'entrevue. Lors de l'enquête nationale par sondage effectuée en Inde en 1958-1959, on a utilisé cette méthode pour calculer directement le taux d'accroissement de la population et pour vérifier l'exactitude des données d'état civil obtenues par la méthodes rétrospective d'enregistrement des naissances et des décès, par mois, pour les deux années précédant chacune des deux visites. En outre, pour vérifier l'exactitude de la seconde énumération effectuée par les enquêteurs, un certain nombre de notations fictives ont été incorporées. La méthode de dénombrement appliquée à cette nouvelle énumération de personnes précédemment dénombrées a donné une légère augmentation du taux de mortalité (du fait qu'il a été tenu compte des enfants nés et décédés durant la période intermédiaire et des décès de migrants) par rapport au taux non corrigé de mortalité pour "l'année précédente".[1/] Lorsque l'analyse des données par périodes rétrospectives mois par mois sera effectuée, on doit pouvoir comparer les trois taux de mortalité : taux non corrigé pour "l'année précédente", taux corrigé pour une période de rétrospection nulle, et taux directement tiré de la nouvelle énumération.

46. Confrontation avec les données enregistrées. C. Chandra Sekhar et W.E. Deming ont mis au point une formule permettant de corriger le nombre des événements d'état civil observés,[2/] par référence aux événements enregistrés.

1/ Indian Statistical Institute (1961), op.cit.; Som et autres auteurs, (1961), op. cit.

2/ C. Chandra Sekhar et W.E. Deming : On a Method of Estimating Birth and Death Rates and the Extent of Registration, dans Journal of the American Statistical Association, Vol. 44, No.245, 1949, pages 101 à 115.

Cette formule a trouvé d'importantes applications dans les enquêtes démographiques de l'Inde[1/] et du Pakistan[2/]. La même population doit théoriquement être à la fois enregistrée et dénombrée (dénombrement rétrospectif), mais les deux opérations doivent être effectuées indépendamment l'une de l'autre et les événements doivent tous être confrontés. Après avoir pris en considération les différences de définition (on prend généralement la population de fait pour l'enregistrement et la population de droit pour l'énumération des naissances et des décès) et tenu compte des migrations on obtient une matrice à quatre éléments :

TABLEAU 9

	Enregistré	Non enregistré	Total
Déclaré au cours de l'enquête	a	b	a + b
Non déclaré au cours de l'enquête	c	d	c + d
Total	a + c	b + d	a + b + c + d

Evidemment, d, nombre des événements ayant échappé dans les deux cas (c'est-à-dire n'ayant été ni déclarés au cours de l'enquête, ni enregistrés), n'est pas connu. Le problème consiste à estimer ce nombre.

1/ The Mysore Population Study, op. cit., (Chapitre 16).

2/ Krotky (1963) op. cit. Voir aussi A.J. Coale, The Design of an Experimental Procedure for Obtaining Accurate Vital Statistics, Congrès international de la population, New York, 1961, Tome II, (Union internationale pour l'étude scientifique de la population, Londres, 1963).

47. En supposant que les omissions (ou inclusions) d'événements, au cours de l'enquête et dans l'enregistrement sont indépendantes, le calcul des probabilités permet d'écrire :

$$d = b./ca$$

ce qui donne une valeur estimée de d. L'estimation de l'inconnue se trouve ainsi ramenée à la résolution d'une simple règle de trois.

48. D'après l'équation ci-dessus, le nombre total estimé d'événements est :

$$\hat{n} = a + b + c + bc/a$$

La complétude des déclarations recueillies lors de l'enquête rétrospective pouvant être un peu moindre dans le cas d'événements ayant échappé à l'enregistrement que dans celui d'événements enregistrés, il se peut que l'hypothèse de l'indépendance des omissions conduise à exagérer légèrement la complétude des résultats de l'enquête. Toutefois, il a été démontré que cette hypothèse n'entraîne qu'une erreur négligeable lorsque 80 pour 100 au moins des événements enregistrés sont signalés lors de l'enquête.[1/]

49. Malgré ces imperfections, la formule proposée demeure un excellent moyen d'augmenter le degré de confiance à accorder aux événements d'état civil observés. Toutefois, parce qu'elle exige le contrôle individuel (confrontation événement par événement), l'application de cette méthode ne peut se faire qu'à une échelle réduite. Il semble donc que la méthode la plus intéressante serait de prendre un sous-échantillon, dont la taille pourrait être déterminée en fonction du coût de l'opération et de l'exactitude requise et à partir duquel les coefficients de correction pourraient

---

1/ Chandra Sekhar et Deming, op. cit.; S. Shapiro, Development of Birth Registration and Birth Statistics in the United States, dans Population Studies, Vol. V, No.1 (1950), pages 86 à 111.

être calculés d'après la formule, puis appliqués aux données initiales de l'enquête par la méthode de la régression ou par celle du quotient.

50. Conditions d'application à l'analyse rétrospective. Il y a lieu de considérer maintenant la manière dont la formule de Chandra Sekhar et Deming peut être adaptée à la méthode d'analyse par période rétrospectives. Il faut d'abord supposer que le temps qui sépare le moment où se produit un événement et le moment où cet événement est enregistré est minime, pour ceux des événements qui sont effectivement enregistrés. Les deux groupes d'événements a et b peuvent alors être analysés par périodes rétrospectives, ce qui permet d'obtenir les estimations pour une période rétrospective nulle. Si on appelle ces estimations a_o et b_o on peut obtenir une estimation de d à l'aide de la formule de Chandra Sekhar-Deming :

$$\hat{d}_o = b_o c/a_o$$

et le nombre totale estimatif d'évenements donné par l'équation :

$$\hat{n} = a_o + b_o + c + b_o c/a_o$$

V. L'ERREUR DE RETROSPECTION DANS LES DONNEES HISTORIQUES SUR LA FECONDITE

51. Généralités. La présente section a pour objet l'étude des erreurs dues à la rétrospection dans les données historiques sur la fécondité, c'est-à-dire dans les renseignements recueillis à une date relativement récente et servant à déterminer d'évolution passée de la fécondité. Etant donné l'erreur de rétrospection constitué dans les données courantes sur les événements d'état civil, une forte erreur de rétrospection est à prévoir pour les données histriques sur la fécondité. Toutefois, l'effet de décalage des périodes, dû au découpage de l'ensemble de l'âge de procréation en périodes délimitées ne se manifesterait sous la forme d'un effet de contrection ni sous celle d'un effet d'extension, dans le nombre total d'enfants nés vivants (y compris ceux qui sont morts par la suite), à moins qu'on ne cherche à analyser, par années depuis le mariage, la fécondité effective et l'incidence des taux de mortalité parmi les enfants.

52. Méthode d'analyse. Comme dans le cas de naissances et décès courants, une analyse par périodes rétrospectives peut ici encore mettre en lumière des contradictions internes dans les données. Pour plus de commodité, l'analyse peut s'appliquer à la cohorte de mariages ou être effectuée en fonction de l'âge de la mère; ces deux variables sont liées à la période rétrospective, mais d'autres périodes rétrospectives interviennent dans l'analyse, à moins qu'on n'opère aussi une mise en tableaux par nombre d'années écoulées depuis le mariage. L'analyse par cohortes a pour avantage de permettre l'étude du même groupe de femmes dans le temps et d'indiquer ainsi les tendances des différentes générations, bien qu'ici intervienne les niveaux courants de fécondité.[1]

1/ A. Das Gupta, Estimation of Vital Rates for Developing Countries, dans Réunion de l'Union internationale pour l'étude scientifique de la population, Ottawa, 21 - 23 août 1963.

53. Analyse du rapport de féminité des naissances et des décès d'enfants de moins d'un an par cohorte de mariages. Considérons d'abord le cas où des renseignements ont été recueillis sur l'âge actuel ou sur l'âge au mariage de la mère (ou de chacun des membres du couple) et où une analyse par cohorte de mariages a été effectuée pour les naissances par mère (ou par couple), pour le rapport de féminité des naissances et pour la mortalité infantile.

54. A titre d'illustration, appliquons la méthode d'ajustement aux données de fécondité recueillies à l'occasion de l'enquête nationale par sondage effectuée en Inde en 1951-1952.[1] Le nombre déclaré des enfants nés vivants par couple marié, pour les différentes périodes, est donné dans la colonne (13) du Tableau 10 : bien que le nombre des enfants issus de la cohorte de mariages la plus ancienne soit inférieur à l'estimation qui pouvait en être faite d'après d'autres renseignements, les chiffres n'indiquent pas clairement l'origine de l'omission dans les déclarations. Considérons maintenant le rapport de féminité des naissances, défini comme étant le nombre des naissances d'enfants du sexe féminin pour chaque naissance d'enfant du sexe masculin, qui est donné dans la colonne (2), et la mortalité infantile définie comme étant le nombre des décès avant l'âge d'un an pour chaque naissance indiquée dana la colonne (6) du même tableau. Alors qu'on peut, en se basant sur des considérations biologiques, admettre que le taux de féminité des naissances n'est pas susceptible de se modifier sensiblement dans le temps, on constate que le nombre des naissances déclarées d'enfants du sexe féminin déclaré pour chaque naissance déclarée d'un enfant du sexe masculin est toujours moins élevé dans le cas des cohortes de mariages les plus anciennes. En outre, alors qu'on sait que la mortalité infantile a récemment fortement diminué, les taux de mortalité infantile établis d'après les déclarations révelent une évolution absolument opposée. Il apparaît clairement que le nombre des omissions a été proportionnellement plus élevé pour les naissances d'enfants du sexe féminin que pour celles d'enfants du sexe masculin, et pour les décès d'enfants de moins d'un an que pour l'ensemble des naissances.

1/ A. Das Gupta, R.K. Som, M. Majumdar et S. Mitra, Couple Fertility, National Sample Survey Report No. 7 (Gouvernement de l'Inde, 1955).

55. La méthode d'ajustement adoptée est expliquée ci-après, dans l'hypothèse que l'erreur de rétrospection en ce qui concerne le sexe a faussé les déclarations de naissances d'enfants du sexe féminin. Toutefois, cette hypothèse est de nature à conduire à une sous estimation de l'erreur totale, dans la mesure où des naissances d'enfants mâles peuvent également avoir été omises, et il peut y avoir un certain effet de compensation dans le rapport de féminité. La valeur de ce rapport pour la plus récente cohorte de mariages, soit 0,9434 naissance de fille par naissance de garçon pour la cohorte de 1946-1951, a été prise pour valeur estimative du rapport de féminité des naissances pour toutes les cohortes de mariages.

56. En ce qui concerne la mortalité infantile, les taux estimés ont été tirés des événements enregistrés et on a supposé que la proportion des lacunes d'enregistrement était la même que pour la plus récente cohorte de mariages.[1/]

57. L'explication de la correction de l'erreur de rétrospection dans les déclarations relatives aux naissances d'enfants du sexe féminin et aux décès d'enfants de moins d'un an est donnée ci-après pour la plus ancienne cohorte de mariages (couples mariés avant 1910). Le rapport de féminité des naissances, d'après les déclarations, est de 0,7937 naissance de fille par naissance de garçon (col.2), alors que le rapport de féminité retenu est de 0,9434. Le total estimé des naissances est 1 + 0,9434 = 1,9434 par naissance de garçon, au lieu de 1 + 0,7937 = 1,7937 d'après les déclarations. A supposer que seules des naissances d'enfants du sexe féminin ont été omises dans les déclarations, l'omission est de 0,9434 - 0,7937 = 0,1497 naissance de fille par naissance de garçon (col.3) : rapportée au nombre estimé de toutes les naissances (par naissance de garçon), le taux d'omission est de 0,1497/1,9434 = 0,0770 (col.4). Ce chiffre représente la probabilité d'omettre une naissance, par suite de l'erreur de rétrospection, dans la déclaration des naissances d'enfants du sexe féminin. En retranchant ce

---

1/ Les taux tirés des événements enregistrés, après ajustement en fonction de la durée des mariages, étaient manifestement inférieurs aux taux réels, mais ils étaient supérieurs aux taux non corrigés tirés des données de l'enquête, sauf en ce qui concerne la cohorte de mariages de 1946-1951.

chiffre de l'unité, on obtient 0,9230 (col.8), qui est la probabilité de ne pas omettre de compter une naissance à cause de l'erreur de rétrospection dans la déclaration des décès d'enfants de moins d'un an.

58. Pour la cohorte de mariages considérée, le nombre déclaré des décès d'enfants de moins d'un an par naissance déclarée est 0,107 (col.6) et le nombre estimé de décès d'enfants de moins d'un an par naissance déclarée est 0,273 (col.7). Le taux d'omission de déclaration de décès d'enfants de moins d'un an par naissance déclarée est donc de 0,273 - 0,107 = 0,166 (col.8). Le nombre corrigé des naissances est de 1 + 0,166 = 1,166 par naissance déclarée (col.9). Le taux d'omission de déclaration des décès d'enfants de moins d'un an par naissance est donc de 0,166/1,166 = 0,1424 (col.10). Ce chiffre donne la probabilité de ne pas omettre de compter une naissance par suite de l'erreur de rétrospection dans la déclaration des décès d'enfants de moins d'un an. En retranchant cette valeur de l'unité, on obtient 0,8576 (col.11), qui est la probabilité de ne pas omettre de compter une naissance par suite de l'erreur de rétrospection précitée.

59. La probabilité composée de ne pas omettre de compter une naissance par suite de l'effet combiné des erreurs de rétrospection dans la déclaration des naissances d'enfants du sexe féminin et des décès d'enfants de moins d'un an, à supposer que ces erreurs sont indépendantes l'une de l'autre, est de 0,9230 x 0,8576 = 0,7916 (col.12). Autrement dit, si 7.916 enfants sont déclarés, le nombre réel est égal à 10.000. Le nombre déclaré d'enfants par couple, soit 4,26 (col.13), divisé par 0,7916, donne le nombre corrigé d'enfants par couple, soit :

$$4,26/0,7916 = 5,38 \text{ (col.14)}$$

60. <u>Analyse de la fécondité totale par âge de la mère</u>. Il peut arriver qu'une simple mise en tableau du nombre totale des naissances par âge de la mère révèle que l'erreur de rétrospection se rapporte aux données relatives aux femmes ayant dépassé l'âge de procréation, puisque le nombre des naissances par mère devrait rester constant après la période de procréation, dans l'hypothèse d'une fécondité stable et de l'absence de relation nette entre la fécondité et la mortalité dans ce groupe d'âges.[1/]

1/ <u>Communications présentées à la réunion de l'Union internationale pour l'étude scientifique de la population, Ottawa, 1963</u> par Das Gupta, (1963), <u>op. cit</u>, R. Blanc (1963), <u>op. cit</u>. M.A.R. El Badry, <u>Errors in Parity Data</u>.

61. Das Gupta, Chotechanapibal, Chalothorn et Siripak ont appliqué aux données sur la fécondité en Thaïlande, en 1960, une technique consistant essentiellement à remplacer la courbe de fécondité "totale" décroissante par une ligne horizontale passant par le maximum de cette courbe, maximum qui est atteint à la fin de la période de procréation : cette technique appelle toutefois une restriction, à savoir que les taux de fécondité à cet âge, ainsi d'ailleurs que dans tous les groupes d'âges, peuvent eux-mêmes être faussés par l'erreur de rétrospection; autrement dit, l'erreur de rétrospection qui agit au delà de l'âge de procréation se manifeste également durant la période de procréation en réduisant le taux de fécondité tiré des déclarations.[1/] Devant cette situation, Das Gupta et d'autres chercheurs ont eu l'idée ingénieuse de prolonger vers la période antérieure la courbe de fécondité totale décroissante obtenue et de déterminer ainsi la valeur du taux existant au début de l'âge de procréation (soit 15 ans environ), et ils ont supposé que, si l'erreur de rétrospection n'avait pas joué, cette valeur constituerait une estimation du taux de fécondité après la fin de la période de procréation.

62. Cette technique n'est pas applicable dans les cas où le taux de fécondité ne diminue pas après la fin de la période de procréation. D'après l'enquête effectuée sur la population de Mysore en 1961, ce taux à continué d'augmenter dans deux sur trois des zones rurales, mais a diminué après l'âge de 55 ans dans la ville de Bangalore.[2/] Une diminution du taux de fécondité totale lorsque la mère a dépassé l'âge de procréation n'a été constatée que dans quatre des quinze groupes de population des pays francophones de l'Afrique de l'ouest et du centre dans lesquels des enquêtes démographiques ont été effectuées entre 1955 et 1962.[3/]

1/ A. Das Gupta, S. Chotechanapibal, T. Chlothorn et W. Siripak, Population Perspective of Thailand (Bangkok, 1963, ronéotypé).

2/ The Mysore Population Study, op. cit., Tableau 10.1, page 112.

3/ R. Blanc, op. cit., (1963), graphique.

Correction de le moins d'un an

Cohorte de mariages	Rapport de féminité des naissances (nombre de naissances de filles par naissance de garçons)	Nombre d'omissions de déclaration de naissance de filles par naissance de garçons	Pro d'o de une san sui mis déc de ces les	Probabilité composée de la déclaration d'une naissance	Nombre de naissances par couple	
					D'après les déclarations	Après correction
(1)	(2)	(3) = 0,9434-(2)	(4) 1,	(12)= (5) x (11)	(13)	(14) (13)/(12)
1946-51	0,9434	0	0	1	0,34	0,34
1940-45	0,9346	0,0088	0,0	0,9545	1,45	1,52
1930-39	0,9259	0,0175	0,0	0,9236	3,07	3,32
1920-29	0,8333	0,1101	0,0	0,8623	4,23	4,91
1910-19	0,8197	0,1237	0,0	0,8345	4,26	5,10
Avant 1910	0,7937	0,1497	0,0	0,7916	4,26	5,38

Source : D'après Das Gupta et autres au

63. Le taux de fécondité moyen par groupe d'âges tiré du recensement effectué en Thaïlande en 1960 est donné au Tableau 11. Le nombre déclaré de naissances par femme accuse une baisse sensible au délà du groupe d'age de 45 à 49 ans. Admettant que la fécondité serait restée stable au cours des dernières décennies, ce que certains indices permettent de supposer, et admettant, en outre, l'absence de toute relation nette entre la mortalité des femmes ayant dépassé l'âge de procréation, d'une part, et la fécondité effective, d'autre part, Das Gupta et autres (1963) ont considéré la différence entre le taux de fécondité totale du groupe d'âge de 45 à 49 ans (b_w), et le taux observé pour les autres groupes d'âge au delà de 50 ans (b_{w+}), comme une mesure de l'erreur de rétrospection dans le taux de fécondité totale des mères de plus de 50 ans.

TABLEAU 11

Taux de fécondité moyen par groupes d'ages :
Recensement en Thaïlande , 1960

Groupe d'âges des (en années)	15-19	20-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64	65-69	70+
Taux de fécondité moyenne	0,07	0,86	2,29	3,78	4,98	5,73	5,91	5,74	5,64	5,43	5,33	5,16

64. La droite représentant la fonction suivante a été ajustée de manière à passer par le point correspondant à un taux de 5,91 pour le groupe d'âge de 45 à 49 ans :

$$f\ (b_{w+}) = 5,91 - 0,02\ w+ \qquad .. \quad (1)$$

l'âge (w+) étant mesuré à partir de 47 ans 1/2, milieu du groupe de 45 à 49 ans d'âge. L'erreur de rétrospection au delà de 50 ans d'âge est alors donné par :

$$b_w - f\ (b_{w+}) = 0,02\ w+ \qquad .. \quad (2)$$

La première approximation du taux au delà de 50 ans d'âge est donnée par :

$$b'_{w+} = b_w = 5,91$$

65. On a opéré un ajustement supplémentaire visant à compenser une omission possible de déclaration dans le groupe de 45 à 49 ans en faisant passer la ligne représentant la relation (2) par le point correspondant au début de l'âge de procréation, considéré comme étant 15 ans, ce qui donne :

$$f\ (b_{15}) = 5{,}91 - 0{,}02\ (15 - 47{,}5) = 6{,}56 \qquad .. (3)$$

Cette valeur a été prise pour estimation définitive de la fécondité réelle dans le groupe d'âge de 45 à 49 ans et plus :

$$b''_{w} = b''_{w+} = f\ (b_{15}) = 6{,}56$$

la correction totale à ajouter étant de l'ordre de 10 pour 100.

66. Comparaison du taux de fécondité totale et des taux courants de fécondité par âge de la mère. Dans des conditions de fécondité stable et en l'absence de relation nette entre la mortalité et la fécondité des mères, la comparaison entre le taux de fécondité totale et les taux courants de fécondité par âge cumulée permet une vérification immédiate de la cohérence des données : si la courbe représentant les taux courants de fécondité cumulée se trouve au-dessus de celle représentant le taux de fécondité totale, on peut en déduire qu'une omission de déclaration fausse les taux courants de fécondité. La figure 5 représente ces courbes pour quinze groupes de population vivant dans les anciens territoires français d'Afrique, et les données de base concernant quatre pays sont indiquées dans le Tableau 12. Dans huit de ces cas, la courbe des taux courants de fécondité cumulée se trouve au dessus de la courbe de fécondité totale pour tous les groupes d'âges; dans cinq des cas, elle s'y trouve à partir de l'âge de 20 ans, et dans deux des cas, à partir de l'âge de 25 ans.

67. En supposant, d'après ces données, qu'une certaine confiance peut être accordée aux renseignements recueillis sur la fécondité totale en ce qui concerne les jeunes femmes, mais que ceux qui concernent les femmes âgées sont sujets à des erreurs par omission, et en admettant en outre que les erreurs dans les déclarations courantes de naissances sont proportionnellement les mêmes pour toutes les femmes, quel que soit leur âge, Lorimer a décrit une méthode graphique permettant d'ajuster les données sur la fécondité courante.[1/]

1/ F. Lorimer, L'Analyse des données démographiques sur l'Afrique au service de la planification économique et sociale, Commission économique des Nations Unies pour l'Afrique, Cycle d'étude des problèmes démographiques en Afrique, (E/CN.14/ASPP/L.6).

TOTAL FERTILITY (—)

CUMULATED CURRENT FERTILITY (---)

FECONDITE TOTALE (—)

ET FECONDITE ACTUELLE CUMULEE (---)

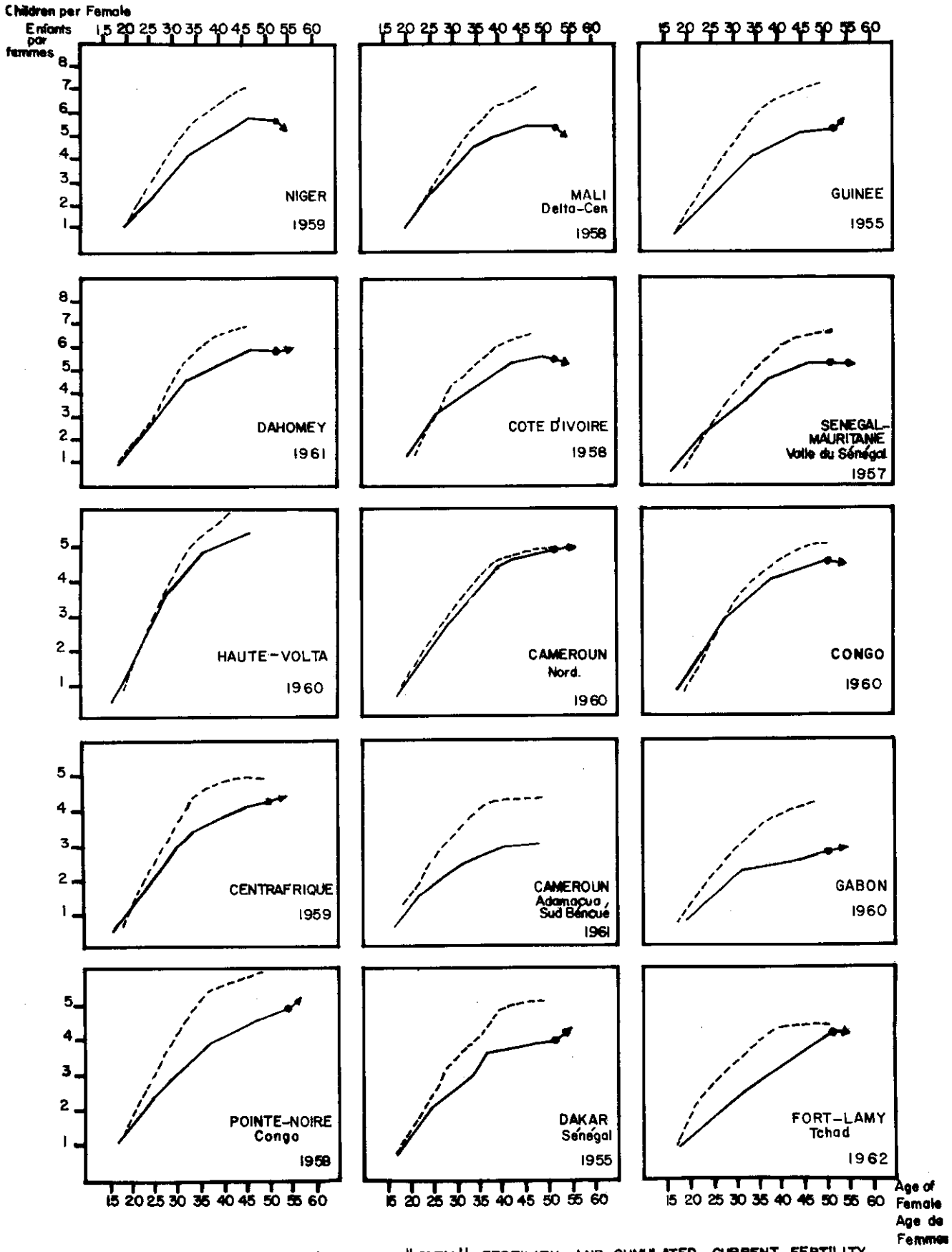


Fig. V COMPARISON OF "TOTAL" FERTILITY AND CUMULATED CURRENT FERTILITY

COMPARAISON DE LA "FECONDITE" TOTALE" ET DE LA FECONDITE CUMULEE

SOURCE: BLANC (1963) op. cit.

TABLEAU 12

Taux de fécondité totale (a) (Nombre total d'enfants nés vivants par femme) et taux courant de fécondité actuelle cumulée (b) par âge

Age des femmes (années)	République centrafricaine		Gabon		Côte-d'Ivoire		Niger	
	(a)	(b)	(a)	(b)	(a)	(b)	(a)	(b)
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
15-19	0,3	0,5	0,85	0,91	0,6	1,3	0,41	1,05
20-24	1,2	1,8	1,50	1,92	2,1	3,1	1,79	2,76
25-29	2,1	2,7	1,96	2,77	3,3	4,8	2,93	4,29
30-34	2,5	3,4	2,27	3,45	4,4	5,9	4,17	5,58
35-39	3,2	3,8	2,41	3,93	5,0	6,8	4,71	6,54
40-44	3,7	4,0	2,58	4,11	6,1	7,2	5,16	7,03
45-49	4,0	4,1	2,72	4,24	5,9	7,5	5,82	7,27
50-54	-	-	-	-	-	-	5,73	-

Source : République centrafricaine : Enquête démographique, Centre-Oubangui, 1959, Méthodologie - Résultats provisoires (République centrafricaine, Mission socio-économique, Centre-Oubangui, Bureau d'étude et recherches du Plan et Service de la statistique générale, et République française, Secrétariat d'Etat aux relations avec les Etats de la Communauté, 1960).

Gabon : Recensement et Enquête démographiques, 1960-1961, Résultats provisoires, ensemble du Gabon, (République gabonaise, Service de statistique; République française, Ministère de la coopération, INSEE, Service de coopération, 1963).

Côte-d'Ivoire : Etude démographique et agricole, 1er Secteur - Côte-d'Ivoire, 1957-1958, (Administration générale des services de la France d'outre-mer), Service des statistiques, et République de la Côte-d'Ivoire, Ministère du Plan, Service de la statistique et de la mécanographie, 1959).

Niger : Etude démographique du Niger, 2ème fascicule, Données individuelles, Résultats définitifs, (République du Niger, Mission démographique du Niger, 1960; République française, Ministère de la coopération, INSEE, Service de coopération, 1963).

68. Une comparaison entre les relevés "rétrospectifs" de fécondité (série B) et les données "courantes" sur le nombre d'enfants (série A) mis au monde par des femmes de classes d'âge comparables permet d'ajuster les taux de fécondité par âge de la série A, et par là d'obtenir une estimation raisonnable de la fécondité totale, étant entendu que 1) les données "rétrospectives" fournies par des femmes jeunes sont la source d'information la plus sûre, et que 2) les erreurs dans les données "courantes" fournies sur les naissances sont proportionnellement les mêmes quel que soit l'âge des femmes.

69. Nous exposerons d'abord une méthode graphique permettant de faire cette comparaison d'une façon approximative : 1) porter sur le graphique le nombre total des enfants mis au monde par les femmes de toutes les classes d'âge successives (série B), en attribuant le nombre moyen des naissances pour chaque classe à l'âge exact auquel on estime que, dans cette classe, la descendance actuelle est égale à la moyenne pour le groupe. A cette fin, on peut utiliser pour toutes les classes, sauf la première (15 à 19 ans), le milieu des classes de cinq années (22 ans 1/2, 27 ans 1/2, etc.). Etant donné que la fécondité augmente rapidement aux environs de la vingtième année, il est préférable de porter sur le graphique la moyenne de la première classe aux environs de 18 ans. 2) porter les sommes successives des taux spécifiques courants (série A) aux âges terminaux, moins la moitié de l'intervalle visé par les relevés, c'est-à-dire à 19 ans 1/2, 24 ans 1/2, etc., 3) Examiner ces courbes qui devraient être à peu près semblables si toutes les données étaient exactes, et voir quelles sont leurs relations générales. 4) Relever sur la courbe B les valeurs correspondant aux âges utilisés dans le tracé de la courbe A (19 1/2, 24 1/2, etc.) et prendre les différences entre ces valeurs pour obtenir une série de taux de fécondité du moment, par âge, pour les classes d'âge successives de 15 à 19 ans, 20 à 24 ans, etc. 5) Calculer les rapports B/A pour les classes d'âge successives (rapports de la série obtenue au point 4 ci-dessus, aux taux courants de la série A). 6) si on obtient par ce moyen une série ajustée, le rapport B/A pour les âges de 20 à 24 ans, ou une moyenne de rapports pour les âges de 20 à 24 ans et de 25 à 29 ans, peut servir de coefficient de correction de tous les taux spécifiques de la

série A.7). La somme des valeurs de la série A ainsi corrigées fournit une estimation de la "fécondité totale", c'est-à-dire du nombre moyen de naissances vivantes par femme ayant vécu pendant toute sa période de procréation.

70. C'est là une méthode assez grossière, et les rapports B/A obtenus de cette façon constituent bien souvent une série aberrante. Une bonne méthode d'ajustement, mise au point par Brass[1/], permet d'obtenir une série plus satisfaisante de taux de fécondité estimés par âge à partir des relevés "rétrospectifs", et par là une série plus satisfaisante de rapports B/A ainsi qu'une estimation plus sûre de la fécondité totale. Cette méthode est d'une application simple et directe, et elle donne une série de valeurs mieux ajustées pour la série B. Les valeurs de la série A n'ont pas été ajustées et, par conséquent, les rapports B/A pourront encore être plus ou moins aberrantes en raison des variations dans l'échantillonnage ou d'autres fluctuations. En particulier, toutes les valeurs relatives à la classe d'âge de moins de 20 ans, correspondant à une période où la fécondité augmente rapidement, doivent être utilisées avec une extrême prudence et le rapport B/A pour les âges de 15 à 19 ans ne doit pas être utilisé comme coefficient de correction. En principe, c'est le rapport B/A pour les âges de 20 à 24 ans qui permet le mieux d'ajuster la série A. Mais comme ce rapport peut être influencé par une erreur d'échantillonage ou par quelque autre irrégularité, on ne l'utilisera pas sans précaution pour corriger les valeurs voisines, surtout celle des âges de 25 à 29 ans. La fécondité totale estimée obtenue de cette manière paraît être l'indice de fécondité le plus sûr qu'il soit possible d'établir actuellement, compte tenu des conditions existant dans la plupart des pays d'Afrique.

71. Cette méthode permet d'obtenir directement une estimation du taux brut de natalité, par l'application du même coefficient d'ajustement au nombre total des naissances de l'année précédente.

1/ W. Brass, The Graduation of Fertility Distributions by Polynomial Functions dans Populations Studies (Londres), Vol. 14 No.2, pages 148 à 162.

VI. RESUME ET RECOMMANDATIONS

72. Dans les pays en voie de développement, dont les systèmes d'inscription à l'état-civil sont défectueux ou inexistants, comme c'est le cas dans la plus grande partie de l'Afrique, les enquêtes rétrospectives par sondage sont actuellement l'un des moyens de déterminer des taux fiables pour les événements d'état civil. Or, s'il est possible, dans une enquête par sondage bien organisée, de mesurer et de réduire les erreurs dues au sondage, il importe également de tenir compte des distorsions et des erreurs non dues au sondage, qui se produisent tant dans les dénombrements complets que dans les enquêtes par sondage et qui ont souvent faussé les résultats d'investigations par ailleurs convenablement préparées.

73. Deux méthodes permettent d'ajuster les données recueillies pour tenir compte des distorsions et des erreurs non dues au sondage : la répétition de l'enquête sur un sous-échantillon composé d'unités retenues pour l'enquête initiale, et l'emploi de réseaux superposés d'échantillons permettant de déceler les différences dans l'observation dues aux enquêteurs ainsi que d'autres éléments tels que les différences entre les bulletins utilisés sur le terrain et entre les méthodes de collecte.

74. Parmi les distorsions non imputables au sondage, un type particulier aux enquêtes rétrospectives est l'erreur de rétrospection du fait de laquelle on voit diminuer (ou augmenter) le nombre des évènements d'état civil déclarés, à mesure qu'augmente l'intervalle entre la date de l'événement et la date de l'enquête. On a mis au point une méthode permettant de déceler ces distorsions et d'ajuster en conséquence les données courantes d'état civil. Sans cela les distorsions pourraient rester dissimulées dans le taux général applicable à l'ensemble des douze mois de la période de référence. Il est recommandé d'adopter cette méthode d'analyse dans les enquêtes rétrospectives pour déceler dans les données toute contradiction interne due à ce type d'erreur que les données soient erronées par défaut ou par excès.

75. Les enquêtes périodiques peuvent également servir à procurer des estimations directes du taux de croissance de la population, et à vérifier l'exactitude des données recueillies à l'aide des enquêtes rétrospectives.

76. On a également suggéré une méthode combinant l'analyse rétrospective des événements d'état civil et la confrontation avec les données enregistrées et des données obtenues par des enquêtes pour obtenir des taux plus proches des taux réels.

77. Dans le cas des données historiques sur la fécondité obtenues par enquêtes rétrospectives, la cohérence interne peut se vérifier au moyen, d'une part, d'une analyse du rapport de féminité des naissances et du nombre des décès des enfants de moins d'un an par périodes de mariage, ainsi que du taux de fécondité totale par âge de la mère, et, d'autre part, d'une comparaison de l'indice synthétique de fécondité (nombre total de naissances vivantes par femme) et des taux de fécondité actuelle selon l'âge des femmes.

78. Qu'il s'agisse d'un dénombrement complet ou d'un sondage, il est recommandé d'incorporer au plan d'enquête les méthodes de détection et d'ajustement des distorsions et des erreurs non dues au sondage dans les données recueillies, pour que l'analyse suggérée puisse se faire.

ANNEXE

LE CADRE THEORIQUE

A1. Modèle statistique. Le modèle statistique suivant, qui se retrouve dans tous les travaux sur le sujet, est une adaptation du modèle de Hansen-Hurwitz-Mark-Mauldin-Bershad[1/] Cf. suppose que l'univers se compose de N éléments, que chacun de ces éléments peut être considéré comme ayant une valeur Y_j et que la proportion à mesurer

$$\overline{Y} = \sum_j Y_j / N \quad \text{où } Y_j = 1,$$

si le jème élément appartient à une classe particulière et où $Y_j = 0$ si tel n'est pas le cas.

A2. Si l'on suppose, pour faciliter l'exposé que l'enquête est un recensement complet ou un sondage équiprobabiliste et si on désigne une observation relative à la jème unité de l'enquête par Y_{jtG}, dont la valeur est égale à l'unité si le jème élément (j = 1, .., n) appartient à la classe particulière considérée pour le tème essai ou à zéro si tel n'est pas le cas (l'indice G souscrit indiquant les conditions générales dans lesquelles l'enquête a été conduite), la moyenne de Y_{jtG} pour l'enquête (essai) est

$$\overline{y}_{tG} = \sum_j y_{jtG} / n$$

dont l'espérance mathématique, qui peut être considérée comme une estimation de $\overline{Y}$ dans les conditions générales de l'enquête, est donnée par

$$E(\overline{y}_{tG}) = \overline{Y} + B_G, \text{ BG étant la distorsion de l'estimation.}$$

1/ M.H. Hansen, W.N. Hurwitz, E.S. Marks et W.P. Mauldin, Response Errors in Surveys dans Journal of American Statistical Association, Vol. 46 (1951), p. 147-190; M.H. Hansen, W.N. Hurwitz et M.A. Bershad, Measurement Errors in Censures and Surveys dans Bulletin de l'Institut international de statistique, Tome 38, p. 359-374.

On constatera que ce modèle s'applique aux recensements ($n = N$) aussi bien qu'aux enquêtes par sondage ($n < N$). Toutefois, les conditions générales indiquées par G, différeront selon qu'il s'agit d'un recensement ou d'une enquête par sondage. La valeur réelle $\bar{Y}$ ne sera bien entendu pas connue. Admettre que $\bar{Y}_{tG}$ est une estimation de $\bar{Y}$, revient donc à supposer a priori que $B_G = 0$, c'est-à-dire que la distorsion est nulle ou négligeable[1/].

A3. Erreurs de sondage et erreurs et distorsions non dues au sondage. Exprimons la variance de sondage de l'estimation $\bar{Y}_{tG}$ de l'enquête, comme on le fait habituellement, par

$$\sigma^2_{\bar{Y}_{tG}} = E\,(\bar{Y}_{tG} - E(\bar{Y}_{tG}))^2$$

D'autre part, l'erreur quadratique moyenne ou l'erreur totale de l'estimation est donnée par :

$$MSE_G = E\,(\bar{Y}_{tG} - \bar{Y})^2 = \sigma^2_{\bar{Y}_{tG}} + G_B^2$$

Il apparaît donc que, sauf le cas où l'élément de distorsion est nul, la variance de sondage de l'estimation donnera l'erreur totale de l'enquête. En outre, une valeur trop faible, même si la variance de sondage n'existe pas dans un recensement complet, l'erreur totale reflètera la distorsion de l'estimation.

1/ Si une distorsion constante, donnant une estimation systématiquement faussée de la valeur de la population, n'influe pas sur la variance de sondage, d'autres erreurs non dues au sondage, même dans le cas d'une espérance mathématique égale à zéro, augmenteront la variance. On trouvera une étude détaillée de cette question dans Hansen et autres auteurs (1961), op. cit.

A4. Comme le montre le Tableau 1, dans le cas des taux non corrigés de l'enquête démographique dans les zones rurales de l'Inde, les erreurs de sondage ont été de 1/1000 pour le taux de natalité et de 1,1/1000 pour le taux de mortalité. Sur la base de ces estimations et de leurs erreurs dues au sondage, on peut admettre que les taux réels de natalité et de mortalité se situent, avec un degré de confiance de 95 pour 100[1], dans les limites suivantes

Estimation de l'échantillon ± Déviation normale correspondant à une probabilité de 95 pour 100 x erreur de sondage;

c'est-à-dire : 34,6 ± 1,96 x 1 ou 32,64 et 36,56 pour le taux de natalité
et 16,6 ± 1,96 x 1,1 ou 14,44 et 13,76 pour le taux de mortalité

à condition que les estimations de l'échantillon ne contiennent aucune erreur ou distorsion non due au sondage. La différence entre les taux non corrigés et les taux corrigés est toutefois égale à 6,3 fois l'erreur de sondage pour le taux de natalité et à 6,7 fois cette erreur pour le taux de natalité. La probabilité d'un tel événement est inférieure à 1/100.000. La sous-estimation des taux non corrigés ne peut donc être attribuée aux variations dues au sondage.

A5. Lors de l'enquête démographique dans l'ancien Etat de Mysore, en Inde, les taux de natalité et de mortalité enregistrés ont été corrigés à la suite de vérifications par recoupement à l'aide de chiffres fournis par d'autres bulletins d'enquête ou par les registres de l'état

---

1/ On trouvera une étude de la probabilité fiducielles dans D.A.S. Fraser, Statistics - An Introduction (John Wiley and Sons, Inc., New-York, 1958), Chapitre II ou dans M.G. Kendall, The Advanced Theory of Statistics, Vol. II (Charles Griffin and Co. Ltd., Londres, 1946), Chapitre 20.

civil[1]. Les taux enregistrés et les taux corrigés figurent au Tableau 1 pour les différentes zones. Il a fallu invariablement ajuster en hausse les taux enregistrés. Si, dans ce cas également, on peut calculer les limites de confiance avec un degré de confiance pour les taux de 95 pour 100, il est possible néanmoins d'adopter une analyse plus rigoureuse car les taux obtenus pour les cinq zones peuvent être considérés comme étant indépendants les uns des autres. La distribution étant supposée normale, on peut calculer au moyen de l'erreur type la probabilité d'obtenir un taux égal ou inférieur au taux corrigé (en l'occurrence, on ne considère que la probabilité unilatérale). Si on examine le taux de natalité et le taux de mortalité séparément, on peut combiner les probabilités relatives aux cinq zones. Les résultats peuvent faire l'objet d'un test par χ^2, $-2 \text{ Log}_e P$ étant réparti comme un χ^2 avec 2a degrés de liberté, P étant la probabilité et a le nombre de zones[2]. Le χ^2 est significatif au niveau 0,01, pour le taux de natalité et au niveau 0,001 pour le taux de mortalité. Les écarts des taux enregistrés par rapport aux taux corrigés ne pouvaient donc pas être attribués aux seules variations dues au sondage.

A6. Répétition de l'enquête. Le modèle statistique pour une répétition de l'enquête $E(\bar{Y}_{tG'}) = \bar{Y}$, dans l'hypothèse que l'élément de distorsion $B_{G'} = 0$, où G' indique les conditions générales de cette répétition. La différence $(\bar{Y}_{tG} - \bar{Y}_{tG'})$ donne donc une estimation de l'élément de distorsion B_G de l'enquête primitive. On peut aussi opérer les ajustements par régression ou au moyen d'estimations par la méthode du quotient[3].

1/ The Mysore Population Study, op. cit.

2/ R.A. Fisher, Statistical Methods for Research Workers, 10ème éd. (Oliver and Boyd, Edimbourg, 1958), Section 21.1.

3/ Hansen et autres auteurs (1953), op. cit. ; Cochran (1962), op. cit.; Sukhatme (1953), op. cit. ; Yates (1953), op. cit.

A7. Dans une opération de contrôle au hasard, les résultats des vérifications sur le terrain effectuées par les contrôleurs peuvent servir pour la mesure et le contrôle des erreurs et des distorsions non dues au sondage dans les travaux des enquêteurs. Dans le cas où le contrôle n'est pas fait au hasard, même si les enquêteurs les moins qualifiés sont bien contrôlés les résultats de base sont améliorés, mais on ne peut pas évaluer les erreurs globales.

A8. Réseaux de sous-échantillons superposés. Une simple mise en tableaux peut faire apparaître des différences entre les estimations faites par des enquêteurs différents s'occupant de la même population. Dans le modèle ci-dessus, l'élément de distorsion (B_{iG}) dans lequel i se rapporte à l'enquêteur, peut être considéré comme se composant de deux parties, une distorsion constante (B'_G) qui se retrouve chez tous les enquêteurs, et un autre élément (B'_{iG}) qui se rapporte au ième enquêteur. Le quotient F de la variance "entre enquêteurs" par la "variance résiduelle" appropriée permet de vérifier l'hypothèse B'_{iG} = constante et donc de vérifier statistiquement l'existence de toute distorsion différentielle venant des enquêteurs.[1]

A9. Une distorsion nette commune à tous les enquêteurs (B'_G) n'est pas décelée par cette méthode. Il y a lieu de remédier à ces distorsions par l'amélioration des méthodes d'enquête, par une comparaison avec des estimations obtenues directement et par l'adoption de procédés spéciaux d'analyse.

A10. Erreur de rétrospection dans les données courantes sur les événements d'état civil[2]. Dans le cas d'une enquête dont les unités (échantillons) changent au hasard (ou uniformément) pendant la période d'enquête d'un an lorsque des renseignements sur les événements survenus au cours de 365 jours précédents (c'est-à-dire avec une période rétrospective d'un an) sont rassemblés avec l'indication du jour ou au moins du mois, où ils se sont produits, si Y_{ij} représente le nombre (estimé) d'événements mensuels déclarés pour le mois i de l'enquête (i = 1, 2, ..., 12)

1/ Cochran (1962), op. cit.
2/ R.K, Som, (1959), op. cit.

comme s'étant produits pendant le mois j la relation peut aussi se représenter par y_{ik} qui est le nombre (estimé) d'événements mensuels déclarés pendant le mois i de l'enquête comme s'étant produits pendant le kème mois précédent (k = 1,2, ..., 12). Considérons le modèle

$$y_{ik} = y + b_k + s_j + e_{ik} \qquad ..(A1)$$

dans lequel y est la moyenne mensuelle "réelle" du nombre des événements, s_j l'élément saisonnier correspondant au mois j (j pouvant être remplacé par j' si j = j', mod.12), b_k la distorsion de rétrospection, supposée être une fonction de la période de rétrospection k, et e_{ik} l'erreur aléatoire. Par hypothèse,

$$\sum_j s_j = 0; \quad E(e_{ik}) = 0. \qquad ..(A2)$$

Le nombre d'événements par année estimé d'après les observations du ième mois de l'enquête est obtenu par la sommation de y_{ik} par rapport à k, l'espérance mathématique correspondante étant :

$$E\left(\sum_k^{12} y_{ik}\right) = 12\,y + \sum_k^{12} b_k \qquad ..(A3)$$

et on obtient le nombre estimé d'événements par année pour la totalité de la durée de l'enquête, qui est d'un an, en faisant la moyenne de la relation ci-dessus pour i = 1,2,..,12, l'espérance mathématique étant donnée par :

$$E\left[(1/12)\sum_i^{12}\sum_k^{12} y_{ik}\right] = (12 \times 12/12)\,y + (12/12)\sum_k^{12} b_k$$

$$= 12\,y + \sum_k^{12} b_k \qquad ..(A4)$$

Telle est la forme dans laquelle les données sont généralement disponibles; elle ne permet pas l'analyse plus poussée par périodes de rétrospection.

All. En revanche, le nombre moyen estimé d'événements par mois pour le kème mois précédent est obtenu par la sommation de y_{ik} par rapport à i (i = 1, 2, .., 12) divisée par 12; cette opération s'exprime par :

$$\bar{y}_{.k} = \sum_{i}^{12} y_{ik}/12 \qquad ..(A5)$$

l'espérance mathématique étant donnée par :

$$E(\bar{y}_{.k}) = y + b_k \qquad ..(A6)$$

La moyenne cumulée de $\bar{y}_{.k}$ jusqu'au kème mois précédent inclus est

$$Y_{.k} = (12/k) \sum_{k=1}^{k} \bar{y}_{.k} \qquad ..(A7)$$

qui donne le nombre moyen estimé d'événements par année pour une période rétrospective de k mois. L'espérance mathématique de $Y_{.k}$ est

$$E(Y_{.k}) = (12 \times k \times y/k) + (12/k) \sum_{k=1}^{k} b_k$$

$$= 12y + (12/k) \sum_{k=1}^{k} b_k \qquad ..(A8)$$

$$= Y + B_K \qquad ..(A9)$$

où Y = 12y, le nombre "réel" d'événements par an, et $B_k = (12/k) \sum_{k=1}^{k} b_k$ la moyenne de la distorsion de rétrospection pour la période de rétrospection de k mois.

A12. Un moyen de vérifier nos relations est que le nombre estimé d'événements par an pour toute la durée de l'enquête (un an)(k = 12) s'obtient en introduisant k = 12 dans la relation (A7), ce qui donne

$$Y_{.12} = \sum_{k=1}^{12} \bar{y}_{.k} \qquad ..(A10)$$

l'espérance mathématique correspondante s'obtient en faisant k= 12 dans la relation (A8), ce qui donne

$$E(Y_{.12}) = 12\,y + \sum_{k=1}^{12} b_k \qquad ..(A11)$$

qui est la même relation que (A4).

A13. S'il n'y a pas de distorsion de rétrospection, c'est-à-dire si $B_K = 0$, on a

$$E(Y_{.k}) = Y \qquad ..(A12)$$

qui est une constante. C'est la vérification de la cohérence. Si les $Y_{.k}$ (estimés) sont portés sur un graphique en fonction de k, période de rétrospection, la courbe doit être statistiquement parallèle à l'axe des k et tout écart par rapport à ce parallélisme probable, qui se révèle statistiquement important, indiquera la présence d'une distorsion, qui est une fonction de la période de rétrospection.

A14. Correction pour tenir compte des erreurs de rétrospection. La correction à appliquer pour tenir compte des erreurs de rétrospection (quand celles-ci ont été décelées à la suite de l'analyse ci-dessus), correction qui doit servir à obtenir une meilleure approximation du nombre réel d'événements d'état civil par année et des taux des événements d'état civil, est la suivante. On ajuste les $Y_{.k}$ (estimés) au moyen d'une courbe appropriée

$$Y_{.k} = f(k) \qquad ..(A13)$$

a partir de laquelle on peut estimer $\hat{Y}_{.0} = f(0)$ qui correspond à la de rétrospection nulle et est considéré comme égal à Y dans l'hypothèse que $B_K = 0$. Quantitativement, $\hat{Y}_{.0}$ indique le nombre probable (ou hypothètique) d'événements par année, si les événements étaient enregistrés au moment même où ils se produisent (période de rétrospection nulle) lorsque les erreurs de rétrospection n'interviennent pas encore. Les distorsions dues à la rétrospection se tirent de la relation

$$B_K = Y - Y_{.k} = \hat{Y}_{.0} - Y_{.k} \qquad ..(A14)$$

Le taux estimé de l'événement d'état civil est alors donné par ce quotient de $\hat{Y}_{.0}$ par la population exposée au risque.

A15. Etude de l'élément saisonnier. Il est également possible d'isoler et d'étudier l'effet saisonnier dans le procédé d'analyse ci-dessus. Dans la figure I, considérons verticalement les croix inscrites au-dessous du mois de décembre 1963; les y_{ik} représentent $y_{1,1}$; ..... ; $y_{11,11}$ et $y_{12,12}$, mais ces y_{ik} se rapportent au même mois d'un événement $j - 12$.

La somme est

$$\sum_{i=k=1}^{12} y_{ik} \text{ et la moyenne } (1/12) \sum_{i=k=1}^{12} y_{ik}, \text{ l'espérance}$$

mathématique étant donnée par la relation

$$y + (1/12 \sum_{k=1}^{12} b_k + s_{12} = y + \bar{b}_. + s_{12}$$

dans laquelle $\bar{b}_.$ est la moyenne des b_k et s_{12} l'effet saisonnier de décembre ($j = 12$). Considérons ensuite les sept croix verticales inscrites au-dessous du mois de l'événement juillet 1963; les y_{ik} représentent $y_{1,6}$; $y_{2,7}$; $y_{7,12}$ et se rapportent au même mois de l'événement $J = 7$. La somme est $\sum_{k=i+5=6}^{12} y_{ik}$.

De même, la somme des cinq croix inscrites au-dessous du mois de juillet 1964 ($j = 7$) est $\sum_{k=i-7=1}^{5} y_{ik}$. Si on additionne verticalement les douze croix inscrites au dessous des mois de juillet 1963 et de juillet 1964 et si l'on divise par 12, on obtient la moyenne du nombre des événements survenus pendant le même mois de juillet ($j=7$), l'espérance mathématique étant donnée par la relation suivante :

$$y + \bar{b}_. + s_7 \qquad .. \text{(A16)}$$

dans laquelle s_7 est l'effet saisonnier correspondant à juillet (j=7). Il est évident que l'espérance mathématique applicable au nombre moyen des événements survenus pendant un mois quelconque j (janvier,...., décembre, correspondant à j = 1, .., 12) est donnée par la relation

$$E(\bar{y}_{.j}) = y + \bar{b}_{.} + s_j \qquad ..(A17)$$

Cette relation permet le test de l'élément saisonnier. S'il n'y a pas d'effet saisonnier, c'est-à-dire, si s_j = constante = 0 (étant donné qu'on a déjà admis par hypothèse que $\sum_{j=1}^{12} s_j = 0$), $E(\bar{y}_{.j})$ = $y + \bar{b}_{.}$ = constante. Les $\bar{y}_{.j}$ observés peuvent être portés sur un graphique et tout écart par rapport à la valeur constante, qui se révèle statistiquement significatif et qui peut faire l'objet d'une vérification à l'aide de la méthode de l'analyse de la variance faite dans certaines hypothèses ou non, révèlerait la présence d'un effet saisonnier. Une courbe ajustée au $y_{.j}$ quand ceux-ci sont nettement différents, donnerait la mesure des éléments saisonniers pour les différents mois.

A16. <u>Ajustement des courbes de rétrospection</u>. Dans l'enquête nationale par sondage de l'Inde, les chiffres relatifs aux naissances et aux décès courants pour une période totale de référence d'une année (les 365 jours précédents la date de l'enquête), enquête qui a eu lieu en 1953-54, 1955-56 et 1957-58 (dans les seules zones urbaines) et dans l'enquête démographique de la Haute-Volta les chiffres afférents aux décès (1960-61), aux naissances et aux décès par période de rétrospection (mois) ont fait apparaître un fléchissement de plus en plus net du nombre estimé des événements si la durée de la période de rétrospection augmente. On a constaté qu'une courbe du type suivant s'adaptait convenablement aux données :

$$f(k) = c\, e^{-ak^2}$$

$$= \hat{Y}_{.o}\, e^{-ak^2} \qquad ..(A18)$$

c et a sont les paramètres de la courbe et e la base du logarithme népérien = 2.7183... La courbe a été ajustée à l'aide de la méthode des moindres carrés, une valeur estimée de c donnant l'estimation $\overline{Y}_{.o}$, estimation du nombre réel des événements par année. Dans le cas également de l'essai d'estimation de la croissance démographique, effectué au Pakistan en 1962, on a constaté que le même type de courbe s'appliquait.

A17. La forme de la courbe de rétrospection a été choisie de manière à assurer un bon ajustement aux données, qu'elle ait un maximum pour une période de rétrospection nulle et soit de forme simple avec seulement deux paramètres, c qui donne une estimation du nombre réel d'événements par an et a qui détermine le taux de fléchissement de l'estimation en fonction de la période de rétrospection. Dans les données de l'Inde, de la Haute-Volta et du Pakistan a est de signe négatif, si la surestimation croît avec la durée de la période de rétrospection, a sera affecté d'un signe positif, cette situation est examinée ci-après. L'équation différentielle qui aboutit à la courbe de rétrospection ci-dessus (A18) est :

$$\frac{d\,\overline{Y}_{.k}}{dk} = -2ak\,Y_{.k} \qquad ..(A19)$$

qui indique un fléchissement plus important de la valeur estimée quand la période de rétrospection augmente. Aucun modèle analogue n'existait pour la courbe d'oubli en psychologie expérimentale, la courbe classique de mémoire étant la forme.

$$R = a - b \log t \qquad ..(A20)$$

avec l'équation différentielle $\frac{d\,R}{d\,t} = -\frac{b}{t}$ ..(A20)

A18. Effet de décalage de période. Appelons C la résultante de l'effet des omissions dues à l'erreur de rétrospection et de l'effet de décalage (par contraction ou par extension) provenant d'un glissement, dans la déclaration des dates des événements. Si C est aléatoire, on peut la considérer comme fonction de k, à savoir C_k. on a alors

$$E(Y_{.k}) = Y + (B_k + C_k) \qquad ..(A21)$$

d'où l'on tire $\hat{Y}_{.o} = Y + (B_o + C_o)$

en supposant comme précédemment que $B_o + C_o = 0$. ..(A22)

A19. Si C n'est ni aléatoire ni une fonction de k, on a alors $C_k = C =$ constante, ce qui donne

$$E(Y_{.K}) = (Y + C) + B_k \qquad ..(A23)$$

de sorte que $\hat{Y}_{.o} = Y + C$ ..(A24)

en supposant toujours que $B_o = 0$.

La résultante nette des omissions de déclaration et de l'effet de décalage se trouvera inextricablement incorporée à la valeur réelle. Il est très peu probable que l'effet de décalage net soit indépendant de la période de rétrospection. Cependant, à supposer que tel soit le cas et si on constate que $\hat{Y}_{.o}$ est une estimation plausible de la valeur réelle, on peut considérer l'effet de décalage net comme étant nul ou négligeable.

A20. <u>Causes de distorsion : non déclaration des décès d'enfants de moins d'un an</u>. Le nombre total des naissances déclarées pendant le ième mois de l'enquête comme étant survenues pendant le kème mois précédent peut être divisé en deux parties : y_{ik}^{a}, qui est le nombre des enfants vivants à la date de l'enquête, et y_{ik}^{d}, le nombre des enfants dont le décès a été déclaré à l'époque de l'enquête. La proportion des décès d'enfants de moins d'un an déclarés, pour une exposition moyenne de six mois, sur la cohorte des naissances survenues pendant les 12 mois précédents est alors

$$p = \sum_k \sum_i y_{ik}^{d} \Big/ \sum_k \sum_i y_{ik}$$

Si on suppose que les omissions des naissances concernent uniquement les enfants de moins d'un an décédés, il convient de corriger le nombre où

des décès déclarés d'enfants de moins d'un an pour inclure ces naissances et ces décès, mais le nombre des enfants survivants déclarés ne changera pas. Le nombre des enfants survivants par personne avant et après la correction tenant compte des décès non déclarés d'enfants de moins d'un an est alors $b(1 - p) = \beta(1 - p')$, b étant le taux de natalité observé, β le taux de natalité estimé (obtenu dans la période de rétrospection nulle, p la proportion des décès observés d'enfants de moins d'un an et p' cette proportion corrigée. En partant de cette équation, on obtient une estimation de p' au moyen de la relation

$$p' = 1 - (1 - p)\ b/\beta$$

Si l'hypothèse, selon laquelle seuls les décès d'enfants de moins d'un an ont été omis, aboutit à un taux de mortalité infantile corrigé qui paraît peu plausible, on pourra en déduire logiquement qu'un certain nombre d'enfants vivants à l'époque de l'enquête ont été également omis.

A21. Erreur de rétrospection dans les données historiques sur la fécondité : analyse du rapport de féminité des naissances et des décès d'enfants par cohorte de mariages[1/]

En se fondant sur des considérations biologiques, on peut admettre que le rapport de féminité à la naissance ne saurait changer considérablement avec le temps. Appelons M_c le nombre d'enfants du sexe masculin et F_c le nombre d'enfants du sexe féminin mis au monde par les mères appartenant à la cohorte de mariages c. Le rapport de féminité des naissances, défini comme étant le rapport des naissances féminines aux naissances masculines est $r_c = F_c/M_c$. Si l'on trouve que $r_c < r_{c'}$, statistiquement, si c'est une cohorte de mariages plus récente que c, la prédominance d'une erreur de rétrospection touchant un sexe plus que l'autre (erreur sélective) s'en trouvera établie, les naissances féminines faisant l'objet de déclarations moins nombreuses que les naissances masculines. Il en est de même pour le cas où $r_c > r_{c'}$.

1/ Das Gupta et autres auteurs (1955), op. cit.

A22. La présence d'une erreur sélective de rétrospection étant ainsi établie, on peut adopter la méthode d'ajustement suivante pour tenir compte de ces distorsions, en supposant que seules les naissances d'un des deux sexes sont omises. Cette hypothèse, toutefois, aboutira probablement à une sous-estimation de l'erreur totale, puisque des naissances de l'autre sexe peuvent aussi être omises, dans une plus faible mesure néanmoins, et qu'il peut y avoir un certain effet compensatoire sur le rapport de féminité.

A23. Appelons ρ_c, le rapport réel entre les sexes, qui pourra être estimé à partir de sources extérieures ou que l'on pourra déduire des renseignements relatifs à la cohorte de mariages la plus récente, si le rapport est plausible. Si on suppose, par exemple, que l'erreur sélective ne joue qu'au détriment des naissances féminines, le nombre de naissances par sexe enregistré et le nombre estimé par cohorte de mariages c sont :

	Naissances féminines	Naissances totales
Nombre déclaré	F_c	$M_c(1 + r_c)$
Nombre estimé	$F'_c = M_c \rho_c$	$M_c(1 + \rho_c)$
Omissions en valeur absolue	$F'_c - F_c = M_c(\rho_c - r_c)$	$M_c(\rho_c - r_c)$
en valeur relative	$(F'_c - F_c)/F'_c = 1 - r_c/\rho_c$	$M_c(\rho_c - r_c)/M_c(1 + \rho_c) = (\rho_c - r_c)/(1 + \rho_c)$

A24. Dans ce cas, la mesure correspondante de l'erreur sélective de rétrospection est la proportion des naissances féminines déclarées, qui est $F_c/F'_c = r_c/\rho_c$ et les chiffres observés peuvent être ajustés au moyen d'une courbe appropriée. La probabilité de la non déclaration d'une naissance du fait de la seule erreur sélective de rétrospection est

$$P(\overline{S}) = (F'_c - F_c)/(F'_c + M_c) = (\rho_c - r_c)/(1 + \rho_c)$$

A25. Pour la cohorte de mariages c, appelons p_c le taux de mortalité infantile enregistré pour une naissance (déclarée) et appelons Λ_c le taux de mortalité infantile estimé au moyen des renseignements extérieurs, ou établi à la suite d'une opération analytique effectuée sur les renseignements fournis par l'enquête. Les lacunes dans la déclaration des décès d'enfants de moins d'un an se mesurent donc par $\Lambda_c - p_c$ qui est aussi la mesure de l'omission de déclaration dans les naissances totales dues à l'omission de déclaration dans les seuls décès d'enfants de moins d'un an. Le nombre réel de naissances pour une naissance déclarée est donc $1^+ \ (\Lambda_c - p_c)$ et la probabilité de l'omission de déclaration d'une naissance due à l'omission de déclaration des seuls décès d'enfants de moins d'un an est :

$$P\ (\overline{I}) = (\Lambda_c - p_c) / \{1 + (\Lambda_c - p_c)\}$$

A26. La probabilité de l'omission d'une naissance du fait de ces deux facteurs à la fois est $P\ (\overline{I}, \overline{S}) = P(\overline{I})\ P\ (\overline{S})$

quand les facteurs sont indépendants.

La probabilité de l'omission d'une naissance du fait d'un au moins des facteurs est

$$P\ (\overline{I} + \overline{S}) = P(\overline{I}) + P(\overline{S}) - P(\overline{I}, \overline{S})$$
$$= P(\overline{I}) + P(\overline{S}) - P(\overline{I})P(\overline{S})$$

quand les facteurs sont indépendants.

La probabilité de la déclaration d'une naissance est donc :

$$1 - P(\overline{I} + \overline{S}) = 1 - P(\overline{I}) - P(\overline{S}) + P(\overline{I})\ P(\overline{S})$$
$$= \{1 - P(\overline{I})\}\{1 - P(\overline{S})\}$$

quand les facteurs sont indépendants

Le facteur de correction (grossissement) applicable au nombre total de naissances déclarées est donc donnée par la relation :

$$1 \quad 1 - P(I + S) = 1 / \{1 - P(I)\}\{1 - P(S)\}$$

Dans ces conditions, si b_c est le nombre d'enfants mis au monde et déclarés par femme, le nombre corrigé est :

$$b'_c = b_c / \{1 - P(I + S)\} = b_c / \{1 - P(I)\}\{1 - P(S)\}$$

A27. Il y a lieu de noter particulièrement que la correction ne donne qu'une limite inférieure aux chiffres, pour les raisons suivantes : premièrement toute omission de déclaration d'enfants du sexe masculin et du sexe féminin intervenant dans des conditions telles que le rapport de féminité n'est pas changé ne sera pas décelé; deuxièmement, l'hypothèse de l'indépendance peut avoir pour effet d'exagérer l'efficacité des chiffres non corrigés. Enfin, il peut exister d'autres causes de sous-déclaration.

A28. Analyse du nombre total d'enfants mis au monde selon l'âge des mères[1].

Appelons b_x le nombre des enfants mis au monde pour chaque femme âgée de x années, la période de procréation s'étendant de a (15 ans, par exemple) à w (50 ans par exemple). Si la fécondité s'est maintenue constante au cours des décennies passées, alors dans les conditions stables de fécondité caractéristiques des pays sous-développés, abstraction faite de toute relation entre la fécondité et la mortalité survenant après la période de procréation, ce qui semble peu probable, le taux réel de fécondité au delà de l'âge w est donné par la relation

$$b_{w+} = b_{w'} \text{ la fécondité réelle à l'âge } w.$$

Si $b_{w+} < b_w$ statistiquement, la présence de l'erreur de rétrospection jouant dans la période suivant la période de fécondité, doit apparaître immédiatement.

Cependant, même si $b_{w+} = b_{w'}$, la présence de l'erreur de rétrospection dans les données historiques sur la fécondité n'est pas écartée, car l'erreur peut jouer dans des conditions telles que la relation est satisfaite.

A29. Toutefois, si $b_{w+} < b_w$ statistiquement, et si b_{w+} peut être ajusté au moyen d'une courbe $f(b_{w+})$ telle que $f(b_w) = b_{w'}$, une estimation de l'erreur de rétrospection concernant le nombre des enfants mis au monde

1/ Das Gupta et autres auteurs (1963), op. cit.

par chacune des femmes des âges postérieurs à la période de fécondité est donnée par $b_w - f(b_{w+})$. La première approximation du taux afférent aux âges postérieurs à la période de fécondité est $b'_{w+} = b_w$; on peut obtenir un complément d'ajustement tenant compte d'une éventuelle omission de déclaration à l'âge w en faisant passer la courbe $f(b_{wt})$ par l'âge a, la valeur obtenue étant alors $b''_{w+} = b'_w = f(b_a)$.

A30. Comparaison du nombre total d'enfants mis au monde et taux courant de fécondité selon l'âge des mères

Si pour un groupe d'âge quelconque (de cinq années) x, F_x est le taux de fécondité totale, (aux points milieux des groupes), f_x étant le taux courant dont les sommes successives $(F'_{x'})$ peuvent être portées sur un graphique pour les âges x' (les âges terminaux moins la moitié de l'intervalle sur lequel portent les déclarations), les graphiques de F_x et de $F'_{x'}$ doivent être identiques statistiquement si toutes les données sont de bonne qualité.

A31. Si l'on peut admettre que les déclarations concernant la fécondité totale des femmes jeunes sont relativement bonnes et que celles qui concernent les femmes plus âgées sont sujettes à des erreurs dues à des omissions, on peut ajuster les chiffres relatifs à la fécondité courante en recourant à une méthode graphique et en faisant une autre hypothèse, à savoir que les erreurs dans les déclarations courantes de naissances sont proportionnellement les mêmes pour les femmes des divers groupes d'âge[1].

A32. Dans ces conditions, sur le graphique de F_x (x = 22,5, 27,5, etc,) on peut lire immédiatement les valeurs correspondant aux âges x' (24,5, 29,5, etc.). Si l'on représente ces valeurs par F''_x, $F''_{y'}$ (pour le groupe suivant), etc, une estimation de f_x (par exemple x = 22,5) est donnée par $F''_{y'} - F''_{x'}$ (par exemple $F''_{29,5} - F''_{24,5}$) = F''_x. Si on suppose que le taux déduit f''_x est plus sûr pour les femmes jeunes (l'indice souscrit "o" indiquant le groupe des âges les moins élevés), les taux courants observés peuvent être corrigés pour donner

1/ F. Lorimer, (1962), op. cit.

$$f''_{x'} = f_x \cdot f''_o / f_o$$

et la sommation de $f''_{x'}$ donnera une estimation corrigée de la fécondité totale.

A33. On peut obtenir également une estimation du taux de natalité au moyen de ce procédé, en appliquant le même facteur de correction f''_o/f_o au nombre total des naissances survenues pendant l'année précédente.