

Quality of age data from African censuses. Evidence from census data matching in Mali. (*French version below*)

Assa Gakou Doumbia (INSTAT, Bamako, Mali) et Véronique Hertrich (INED, Paris, France)

Short abstract.

Estimating ages and dates remains a significant issue in many sub-Saharan countries, in terms of data collection and data analysis. In populations where these notions are not commonly used, the estimates are imprecise and follow standard pattern of errors that skew the results on population structure but also on demographic trends. Various indicators have been designed to evaluate the quality of data on age; however they rarely allow to assess and to document the causes of distortion. In this paper we use a unique body of longitudinal data concerning a rural population of Mali, which provides linkage of individual records from 4 national censuses (1976, 1987, 1998, 2009). These data give the opportunity to examine the dynamics of statements on age as individuals progress in their life cycle (age, marital status, children ...). The paper has two objectives: 1) evaluate the accuracy of age reports; 2) document the pattern of errors related to the lifecycle status of individuals.

Long abstract.

The estimation of age and dating events remain problematic in many sub-Saharan countries. On one side, this information is tedious to obtain when people are unfamiliar with these concepts and where civil registration is deficient and covers only part of the events (Ewbank, 1981; Lohle-Tart and Clairin 1988; Hertrich, 1992; Lohle-Tart and Francis, 1999; Johnson et al., 2009). On the other side, the results often reflect bias and significant distortions that affect their interpretation (Pison, 1979; Blanc and Rutenberg, 1990; Rutstein et al., 1990; Gage, 1995; Pullum and Stokes, 1997; Pullum, 2006). Indeed, vagueness and inaccuracies are not random but they often follow standard patterns, varying with gender, age and social attributes associated with age. These limits are obvious on cross-sectional data, especially with irregularities at certain ages (heaping) or age groups (age transfer at the upper and lower limits of eligibility) and distortions of sex ratios (Roger et al, 1981; Waltisperger 1988). They have an impact on demographic indicators and also on demographic trends. For instance, recent studies show that, because of these distortions, the decline of fertility in sub-Saharan Africa has probably been overestimated (Moultrie and Dorrington, 2008; Schoumaker, 2009) and trends in child mortality are probably distorted. Despite improvements in survey procedures and methods for adjusting the data, the problem is far from resolved and remains a challenge not only for the production of data but also to enhance the quality of knowledge on demographic changes in sub-Saharan Africa.

In this paper, we propose to analyze the recording errors on age with longitudinal data from a small scale and follow-up survey conducted in rural Mali for 25 years.

Classical indicators used to check the quality of the data on age are cross-sectional, they provide evidence on the reality of errors but they fail to assess and to document the causes of distortions. For the study population, we have a unique body of data with linkage of individual records from 4 independent observations: those provided by the Malian national censuses conducted in 1976, 1987, 1998 and 2009. These data give the opportunity to examine the dynamics of statements on age as individuals progress in their life cycle (age, marital status, children ...).

The paper has two objectives: 1) to evaluate the quality of age reports and its trends over 3 decades; 2) to document the pattern of errors related to the lifecycle status of individuals. To meet the first

objective, we will calculate indicators on the consistency of age reports, for individuals recorded by several censuses. To address the second point, we will examine the pattern and the magnitude of the discrepancies on age reports between two successive censuses, according to the progression of the individual life cycle from one census to the next and the gender of the respondent. We expect that the probability for two respondents from the same age at one census to be recorded with different ages at the next census increases when their status differs (for instance one getting married and the other staying single; one becoming household head and the other not; one having children and the other not...).

Population and data

The study population (7 villages, 4300 inhab. in 2009) is in the Bwa ethnic region of Southeast Mali, about 450 kilometers from Bamako, and near Burkina Faso. Socio-economic and demographic characteristics are those commonly shared by Sahelian farmers. The population lives of family-based agricultural production, mainly food crops. Fertility is still high (8 children per women). Domestic groups are large and complex. Schooling, hitherto marginal, developed with the advent of democracy and decentralization in the 1990s, with around half of all children in the study area attending school at the end of the 2000s. Other social changes since the 1990s include the generalization of temporary labor migration for adolescent girls and boys and the weakening of traditional marriage patterns (rise in age at marriage and in informal unions).

A demographic surveillance system (named SLAM - Suivi longitudinal au Mali / Longitudinal Survey in Mali) has been implemented in 1987-89, with a new round every 5 years, the last update in 2009-10. The SLAM observation system combines various operations (panel survey, life-event survey, genealogy, etc.). This paper is based mainly on data from a follow-up survey (« Enquête renouvelée »), conducted exhaustively (both sexes, all ages) in the 7 villages.

This follow-up survey includes individual data from different censuses. A new local census is realized each five years, which is merged with the previous ones, and completed with information recorded through interviews in each family. The objective is to follow every individual recorded by (at least) one census in order to know his status and place of residence at every other census. Thanks to a partnership between Ined and the Malian institute of statistics (Instat), data from the national censuses were also incorporated in the database. We have therefore a unique body of data based on the linkage between individual information recorded by independent national censuses. Basically, it means that we have, for each individual, a number of independent reports of age which corresponds to the number of times he/she was registered by a national census.

The current version of the database includes 9 censuses over the 1976-2009 period: 4 national censuses (1976, 1988, 1998, 2009) and 5 local censuses (1988, 1994, 1999, 2004, 2009). The follow-up was done for 8,300 individuals who were registered as resident at least at one census.

The analysis will focus on national censuses, and on individuals who were registered by two consecutive national censuses. (N=4853 observations with linkages between 2 consecutive censuses).

Structure of the presentation (provisional)

Introduction.

The issue of quality on age reporting in SubSaharan Africa : Background and littérature review.

Section 1. Quality of data on ages in Malian censuses

Evaluation of the quality of data on ages in Malian censuses; trends between the first (1976) and the last (2009) censuses. Assessment using classical indicators on the reliability of reports and the patterns of discrepancies (population pyramid, sex ratios, Whipple indicators...)

Section 2. Focus on a small-scaled longitudinal survey

- Presentation of the follow-up survey and the procedure of individual matching
- Short présentation of the context

Section 3. Proportion of discrepancies.

Evaluation of the quality and consistency of the reports on age, based on 2 types of indicators :

- Classical indicators (cross-sectional mesures)
- Longitudinal indicators (% of consistent/inconsistent observations by sex age)

Section 4. The mechanisms at work.

- Insights and analysis of the patterns of discrepancies, with a special focus on the critical ages : adolescence/transition to adulthood, end of reproductive age/transition to old age... And special attention to the sex-specific distorsions.
- Probability of distorsion, according the life events experienced between the censuses, and controlling by birth cohort and gender. Special attention on events which are driving changes in social and family status : marriage, widowhood, birth of children, access to the status of household head,...

Conclusion : appraisal and perspectives

Références

- Blanc Ann et Naomi Rutenberg, (1990). Assessment of the quality of data on age at first sexual intercourse, age at first marriage, and age at first birth in the demographic and health surveys, in: Institute for Resource Development, 1990. - *An Assessment of DHS-I Data Quality*. - Maryland, Macro Systems Inc. (DHS Methodological Report, No.1)
- Ewbank Douglas, (1981). *Age misreporting an age-selective underenumeration : sources, patterns, and consequences for demographic analysis*. – Washington D.C., National Academy Press, 112 p. (Committee on population and demography, Report n° 4).
- Gage Anastasia (1995). *An assessment of the quality of data on age at first union, first birth and first sexual intercourse for phase II of the demographic and health surveys program*. - Occasional paper, Calverton, Maryland, ORC Macro, 53 p. [DHS. Demographic and health surveys. Occasional paper 4].
- HERTRICH Véronique, 1992. – Apport des sources existantes à la datation des événements. Une enquête en pays bwa au Mali, *Population*, n°5, p. 1263-1292.
- Johnson, K., Grant, M., Khan, S., Moore, Z., Armstrong, A., and Sa, Z., (2009). *Fieldwork-related factors and data quality in the demographic and health surveys program*. Calverton, Maryland, USA: ICF Macro. (DHS Analytical Studies No. 19). <http://www.measuredhs.com/pubs/pdf/AS19/AS19.pdf>.
- Lohle-Tart Louis et Rémy Clairin (éds.), 1988. – *De l'homme au chiffre. Réflexions sur l'observation démographique en Afrique*. – Paris, CEPED-UIESP-IFORD, 329 p.
- Lohle-Tart Louis et Michel François, 1999. – *État civil et recensements en Afrique francophone*. – Ceped, Paris, 555 p. [Les documents et manuels du CEPED, n° 10]
- Moultrie, T.A. and Dorrington, R. (2008). Sources of error and bias in methods of fertility estimation contingent on the P/F Ratio in a time of declining fertility and rising mortality. *Demographic Research* 19(46): 1635–1662. doi:10.4054/DemRes.2008.19.46.
- Pison G. (1979), Âge déclaré et âge réel: une mesure des erreurs sur l'âge en l'absence d'état civil, *Population*, N° 3, pp. 637-648.
- Pullum, T.W. (2006). *An assessment of age and date in the DHS surveys, 1985 - 2003*. (DHS Methodological Reports 5). http://www.measuredhs.com/pubs/pub_details.cfm?ID=664&srchTp=type.
- Pullum, T.W. and Stokes, S.L. (1997). Identifying and adjusting for recall error, with application to fertility surveys. In: Lyberg, L. and Biemer, P. (eds.). *Survey Measurement and Process Quality*. New York: John Wiley and Sons: 711–732.
- Roger Gilles, Waltisperger Dominique et Corbille-Guitton Christine, 1981. – *Les structures par sexe et âge en Afrique*. – Paris, Groupe de Démographie Africaine, IDP/INED/INSEE/MICOOP/ORSTOM, 556 p.
- Rutstein, Shea O., and George T. Bicego. (1990). *Assessment of the quality of data used to ascertain eligibility and age in the Demographic and Health Surveys*. In *An assessment of DHS-I data quality*. DHS Methodological Reports No.1. Columbia, Maryland: Institute for Resource Development/Macro Systems Inc.
- Schoumaker, Bruno (2009). *Stalls in fertility transitions in sub-Saharan Africa : real or spurious ?*; Louvain-la-Neuve : Université catholique de Louvain, Département des sciences de la population et du développement (SPED), 2009 - 1 vol. (46 p.) ; 21 cm. - (Document de travail ; 30)
- Som, R.K. (1973). *Recall Lapse in Demographic Enquiries*. New York: Asia Publishing House.
- Waltisperger Dominique (1988). Les structures par âge, in : Lohle-Tart Louis et Clairin Rémy (éds.), *De l'homme au chiffre. Réflexions sur l'observation démographique en Afrique*, p. 203-216– Paris, CEPED-UIESP-IFORD, 329 p.

Quelle fiabilité accorder aux déclarations sur les âges en Afrique ?

Les enseignements d'un appariement de recensements au Mali.

Assa Gakou Doumbia (INSTAT, Bamako, Mali) et Véronique Hertrich (INED, Paris, France)

Résumé court

L'estimation des âges et des dates reste un problème important en Afrique subsaharienne, en termes de collecte comme d'analyse. Dans des populations où ces notions sont peu utilisées, les estimations sont imprécises et suivent des schémas-types d'erreurs qui biaisent les données de structure mais aussi les tendances démographiques. Différents indicateurs ont été conçus pour évaluer ces données : ils permettent de mettre en évidence des irrégularités et de poser des hypothèses sur leurs causes mais pas de les vérifier ni de les documenter. Dans cette communication nous utilisons un corpus inédit de données longitudinales sur une population rurale du Mali, qui fournit un appariement individuel d'enregistrements collectés de façon totalement indépendante par 4 opérations : les recensements nationaux de 1976, 1987, 1998 et 2009. On a ainsi la possibilité d'examiner la dynamique des déclarations sur l'âge, à mesure que les individus avancent dans leur cycle de vie (âge, statut matrimonial, descendance...). La communication aura deux objectifs : 1) évaluer la fiabilité des déclarations sur les âges et son évolution ; 2) documenter les mécanismes d'erreurs de déclaration associés au statut des individus en évaluant les incohérences de déclarations entre recensements selon le cycle de vie parcouru par l'individu dans l'intervalle.

Résumé long

L'estimation de l'âge et la datation des événements restent problématiques en Afrique subsaharienne. Côté collecte, l'enregistrement de ces informations s'avère souvent fastidieuse et requiert des efforts particuliers pour estimer des âges et des dates que les populations, peu familières de ces notions, connaissent mal et que l'état civil, généralement déficient, ne couvre que très partiellement (Ewbank, 1981 ; Lohle-Tart et Clairin, 1988 ; Hertrich, 1992 ; Lohle-Tart et François, 1999 ; Johnson et al., 2009). Côté analyse, les résultats rendent compte bien souvent de biais et de distorsions importantes qui pèsent sur l'interprétation (Pison, 1979 ; Blanc and Rutenberg, 1990 ; Rutstein et al., 1990 ; Gage, 1995 ; Pullum and Stokes, 1997 ; Pullum, 2006). En effet, imprécision et erreurs d'estimation ne sont pas aléatoires mais suivent des schémas-type assez marqués, variant avec le sexe, l'âge et les attributs sociaux qui leur sont associés. Ces limites sont manifestes sur les données de structure, notamment avec des déséquilibres d'effectifs à certains âges (attraction ou répulsion des âges ronds) ou groupe d'âges (sur ou sous-représentation au profit des groupes encadrant), et des variations problématiques des rapports de masculinité (Roger et al., 1981 ; Waltisperger, 1988). Elles ont des retombées sur les indicateurs démographiques et pèsent aussi sur l'approche des tendances démographiques. Des recherches récentes montrent ainsi que, du fait de ces distorsions, la baisse de la fécondité en Afrique subsaharienne a sans doute été surestimée (Moultrie and Dorrington, 2008 ; Schoumaker, 2009) et les tendances de la mortalité des enfants sont probablement déformées. En dépit d'efforts engagés au niveau des procédures de collecte et de redressement des données, le problème est loin d'être réglé et reste un défi non seulement pour la production de données mais aussi pour une approche fiable des changements démographiques en Afrique subsaharienne.

Dans cette communication, nous proposons d'analyser les erreurs de déclarations des âges avec une perspective longitudinale, à l'échelle d'une population rurale du Mali, objet d'un suivi démographique depuis 25 ans.

Les indicateurs usuels d'évaluation des données sur les âges sont des indicateurs du moment : ils permettent de mettre en évidence des irrégularités et de poser des hypothèses sur leurs causes mais pas de les vérifier ni de les documenter. Pour la population étudiée, nous disposons d'un corpus inédit appariant pour chaque individu, les enregistrements collectés de façon totalement indépendante par 4 opérations nationales : les recensements généraux de la population et de l'habitat de 1976, 1987, 1998 et 2009. On a ainsi la possibilité d'examiner la dynamique des déclarations sur l'âge, à mesure que les individus avancent dans leur cycle de vie (âge, statut matrimonial, constitution de la descendance, mariage des enfants...).

L'objectif de notre communication sera double : d'une part évaluer la fiabilité des déclarations sur les âges et son évolution dans le temps ; d'autre part documenter les mécanismes d'erreurs de déclaration associés au statut des individus. Le premier point sera traité en construisant des indicateurs sur la cohérence des déclarations de l'âge, pour les individus ayant été enregistrés par plusieurs des recensements. Le second point mettra l'accent sur la variation des écarts de déclaration selon les étapes du cycle de vie parcourus par les individus d'un recensement à l'autre, et, bien sûr, leur sexe. On s'attend ainsi à ce que deux individus de même âge à un recensement donné aient une probabilité d'autant plus importante d'être déclarés d'âges différents au recensement suivant, qu'ils n'affichent pas les mêmes marqueurs statutaires (par exemple : l'un marié, l'autre célibataire ; l'un chef de famille, l'autre non ; l'une ayant eu des naissances l'autre non...).

Population et données

La population étudiée se situe au Sud-Est du Mali (450 km environ de Bamako) et appartient à l'ethnie des Bwa. Elle présente les traits classiques d'une population d'agriculteurs sahéliens, avec une économie d'autosubsistance centrée sur la culture du mil, une fécondité toujours élevée (8 enfants par femme), des groupes domestiques de structure complexe... Cependant depuis les années 1990, des changements se produisent sur de multiples fronts : développement de la scolarisation (environ 50% des enfants scolarisés aujourd'hui), augmentation de l'enregistrement des naissances à l'état civil (environ 60% aujourd'hui), changements matrimoniaux, essor des migrations de travail des jeunes vers la capitale Bamako...

Cette population fait l'objet d'un suivi de population depuis 25 ans (projet Slam – Suivi longitudinal au Mali). Un dispositif de collecte comportant plusieurs enquêtes (enquête biographique, enquête renouvelée...) y a été mis en place en 1987-88, avec un nouveau passage tous les 5 ans (le dernier en 2009-2010). Pour cette communication nous utiliserons principalement les données de l'enquête renouvelée, menée dans 7 villages (4300 habitants en 2009).

L'enquête renouvelée réunit les données individuelles issues de différents recensements. À chaque nouveau passage, un nouveau recensement est réalisé, dont les données sont appariées avec celles des recensements antérieurs. Des entretiens sont réalisés auprès des familles de telle sorte que l'état (vivant, décédé ou non-né) et la situation de résidence de tout individu saisi (par l'une au moins des opérations) soient connus au moment de chacune des observations. Grâce à un partenariat entre l'Ined et l'Instat, les données des recensements nationaux portant sur les villages étudiés ont pu être intégrées à la démarche. On se trouve ainsi dans une situation exceptionnelle d'appariement de données individuelles collectées de façon totalement indépendante et dans le cadre du dispositif de collecte national du recensement. Très concrètement, on dispose pour chaque individu d'autant de « versions » de l'âge que de recensements auxquels il a été recensé.

La base de données actuelle comprend 9 recensements : les recensements nationaux (1976, 1987, 1998, 2009) et 5 recensements locaux¹ (1988, 1994, 1999, 2004, 2009). Elle compte environ 8300 enregistrements d'individus recensés comme résident à l'un au moins des 9 recensements.

Nos analyses porteront sur la comparaison des déclarations sur les âges faites aux différents recensements nationaux pour les individus recensés à deux d'entre eux au moins, soit 4853 observations (appariement d'information sur l'âge fournies par 2 recensements consécutifs).

Analyses et plan pressenti de la communication

Introduction.

Positionnement du problème et des enjeux de l'estimation des dates et âges en Af sub-saharienne.
Revue bibliographique.

Partie 1. Qualité des données sur les âges dans les recensements maliens

Point sur la qualité des données sur les âges au Mali et son évolution au travers des 4 recensements nationaux. Il s'agit ici de faire un point à partir des données nationales et en utilisant les indicateurs classiques de mesure de la qualité et des distorsions sur les âges (pyramides, rap de masculinité, indices de whipple & co).

Partie 2. Focus sur un suivi longitudinal à petit échelle

- Présentation de l'enquête renouvelée et de l'appariement des RGPH
- Présentation rapide de la population

Partie 3. Fréquence des discordances.

- Examen de la qualité et de la cohérence des données dans la population étudiée au travers :
 - 1) des indicateurs classiques (reprise de certains indicateurs utilisés pour évaluer la situation nationale et comparaison)
 - 2) Indicateurs longitudinaux (% de cohérences/incohérences selon sexe et âge, comparaison entre les estimations de 2 ou 3 recensements)

Partie 4. Les mécanismes à l'œuvre.

- Mise en évidence des déformations en se focalisant sur le passage à des « âges critiques » : autour de l'adolescence/âge adulte, autour de l'entrée en vieillesse, Et autour de schémas type (exemple : plus fort vieillissement des ho par rap aux femmes)
- Probabilité de discordance, à génération égale et sexe égal, selon le fait que l'individu ait connu ou non un événement porteur de changement de statut (mariage, veuvage, fécondité, accès aux responsabilités de chef de famille...) entre les deux recensements

Conclusion : bilan, perspectives

¹ L'âge n'a pas fait l'objet d'un enregistrement indépendant par les recensements locaux ; ceux-ci ne seront pas utilisés dans le cadre de cet article.

Références

- Blanc Ann et Naomi Rutenberg, (1990). Assessment of the quality of data on age at first sexual intercourse, age at first marriage, and age at first birth in the demographic and health surveys, in: Institute for Resource Development, 1990. - *An Assessment of DHS-I Data Quality*. - Maryland, Macro Systems Inc. (DHS Methodological Report, No.1)
- Ewbank Douglas, (1981). *Age misreporting an age-selective underenumeration : sources, patterns, and consequences for demographic analysis*. – Washington D.C., National Academy Press, 112 p. (Committee on population and demography, Report n° 4).
- Gage Anastasia (1995). *An assessment of the quality of data on age at first union, first birth and first sexual intercourse for phase II of the demographic and health surveys program*. - Occasional paper, Calverton, Maryland, ORC Macro, 53 p. [DHS. Demographic and health surveys. Occasional paper 4].
- HERTRICH Véronique, 1992. – Apport des sources existantes à la datation des événements. Une enquête en pays bwa au Mali, *Population*, n°5, p. 1263-1292.
- Johnson, K., Grant, M., Khan, S., Moore, Z., Armstrong, A., and Sa, Z., (2009). *Fieldwork-related factors and data quality in the demographic and health surveys program*. Calverton, Maryland, USA: ICF Macro. (DHS Analytical Studies No. 19). <http://www.measuredhs.com/pubs/pdf/AS19/AS19.pdf>.
- Lohle-Tart Louis et Rémy Clairin (éds.), 1988. – *De l'homme au chiffre. Réflexions sur l'observation démographique en Afrique*. – Paris, CEPED-UIESP-IFORD, 329 p.
- Lohle-Tart Louis et Michel François, 1999. – *État civil et recensements en Afrique francophone*. – Ceped, Paris, 555 p. [Les documents et manuels du CEPED, n° 10]
- Moultrie, T.A. and Dorrington, R. (2008). Sources of error and bias in methods of fertility estimation contingent on the P/F Ratio in a time of declining fertility and rising mortality. *Demographic Research* 19(46): 1635–1662. doi:10.4054/DemRes.2008.19.46.
- Pison G. (1979), Âge déclaré et âge réel: une mesure des erreurs sur l'âge en l'absence d'état civil, *Population*, N° 3, pp. 637-648.
- Pullum, T.W. (2006). *An assessment of age and date in the DHS surveys, 1985 - 2003*. (DHS Methodological Reports 5). http://www.measuredhs.com/pubs/pub_details.cfm?ID=664&srchTp=type.
- Pullum, T.W. and Stokes, S.L. (1997). Identifying and adjusting for recall error, with application to fertility surveys. In: Lyberg, L. and Biemer, P. (eds.). *Survey Measurement and Process Quality*. New York: John Wiley and Sons: 711–732.
- Roger Gilles, Waltisperger Dominique et Corbille-Guitton Christine, 1981. – *Les structures par sexe et âge en Afrique*. – Paris, Groupe de Démographie Africaine, IDP/INED/INSEE/MICOOP/ORSTOM, 556 p.
- Rutstein, Shea O., and George T. Bicego. (1990). *Assessment of the quality of data used to ascertain eligibility and age in the Demographic and Health Surveys*. In *An assessment of DHS-I data quality*. DHS Methodological Reports No.1. Columbia, Maryland: Institute for Resource Development/Macro Systems Inc.
- Schoumaker, Bruno (2009). *Stalls in fertility transitions in sub-Saharan Africa : real or spurious ?*; Louvain-la-Neuve : Université catholique de Louvain, Département des sciences de la population et du développement (SPED), 2009 - 1 vol. (46 p.) ; 21 cm. - (Document de travail ; 30)
- Som, R.K. (1973). *Recall Lapse in Demographic Enquiries*. New York: Asia Publishing House.
- Waltisperger Dominique (1988). Les structures par âge, in : Lohle-Tart Louis et Clairin Rémy (éds.), *De l'homme au chiffre. Réflexions sur l'observation démographique en Afrique*, p. 203-216– Paris, CEPED-UIESP-IFORD, 329 p.