A LONGITUDINAL PATIENT RECORD FOR PATIENTS RECEIVING ANTIRETROVIRAL TREATMENT

E. KOTZE AND T. MCDONALD

ABSTRACT

In response to the Human Immunodeficiency Virus (HIV) epidemic in the country, the South African Government started with the provisioning of Antiretroviral Therapy (ART) in the public health sector. Monitoring and evaluating the effectiveness of the ART programme is of the utmost importance. The current patient information system could not supply the required information to manage the rollout of the ART programme. A data warehouse, consisting of several data marts, was developed that integrated several disparate systems related to HIV/AIDS/ART into one system. It was, however, not possible to trace a patient across all the data marts in the data warehouse. No unique identifiers existed for the patient records in the different data marts and they also had different structures. Record linkage in conjunction with a mapping process was used to link all the data marts and in so doing identify the same patient in all the data marts. This resulted in a longitudinal patient record of an ART patient that displayed all the treatments received by the patient in all public health care facilities in the province.

Keywords: data warehouse, data mart, record linkage, antiretroviral therapy, longitudinal patient record.

1.   INTRODUCTION

The Acquired Immune Deficiency Syndrome (AIDS) epidemic, caused by the Human Immunodeficiency Virus (HIV), is a global crisis threatening development gains, economies, and societies. Within sub-Saharan Africa, where the epidemic first began and with the highest HIV prevalence, African countries have recorded death rates not seen since the 1950's or 1960's.   This region remains the worst affected by HIV, accounting for 68% (22.5 million) of all people living with HIV and 72% (1.3 million) of AIDS deaths in 2009 (UNAIDS, 2010a). With an estimated 5.6 million people living with HIV in 2009, South Africa's epidemic remains the largest in the world (UNAIDS, 2010a). The UNAIDS estimated that the number of AIDS related deaths in South Africa in 2009 ranged anywhere between 226 000 and 390 000 (UNAIDS, 2010b).

In response to this epidemic the South African Government created the HIV/AIDS and Sexually Transmitted Disease (STD) Strategic Plan (National Department of Health, 2003). The purpose of this plan is to provide a broad national framework around four priority areas: prevention; treatment, care and support; research, monitoring and evaluation; and human and legal rights. In November 2003, after considerable sustained pressure from advocacy groups, the government adopted the Operational Plan for Comprehensive HIV and

AIDS Treatment and Care, which included the provisioning of Antiretroviral Therapy (ART) in the public health sector. Following the South African National Policy, the Free State Provincial Government, and more specifically, the Free State Department of Health (FSDOH) established its Comprehensive Care, Management and Treatment of HIV and AIDS programme, which included the provisioning of highly active antiretroviral therapy (HAART) in May 2004 (National Department of Health, 2003). Monitoring and evaluating the effectiveness of the ART programme is of the utmost importance. Strategic information plays a major role in this endeavour.

The existing patient information system could not fulfil this role because it was a traditional online clinical system, dealing with the operational issues of accumulating patient data on a patient. Very little functionality was provided to deal with the complexities of managing the clinical outcomes of the ART programme. To add to the problem, other independent online transaction processing (OLTP) systems, all closely related to HIV/AIDS, had to be interrogated to gain an understanding of the impact the rollout of antiretroviral (ARV) drugs had. Examples of these systems were standalone human resource systems, information systems accumulating data on tuberculosis, patient admissions, notifiable diseases and blood tests. These systems were all incompatible with each other.

The problem of combining information has been under the microscope in the OLTP world. The outcome was an integrated Electronic Patient Record (EPR) that was captured in an online OLTP system. This OLTP system enabled real-time online electronic order entry, documentation, results review and clinical decision support (Ebidia, Mulder, Tripp and Morgan, 1999). Ebidia et al. (1999) proposed that a patient's EPR should contain the following information:

- Demographic Information (incl. Personal Details)
- Past Medical History (incl. Allergies and Blood group)
- Medication Profile and Prescriptions
- Blood Test Results
- Imaging Results

An Electronic Health Record (EHR) can be defined as an aggregation of patient-centric health data that originates in the patient record systems of multiple independent healthcare organizations for the purpose of facilitating care across multiple organizations (Rishel, Thomas, Handler and Edwards, 2005). In other words, the EHR is a long-term record of a patient, transcending his or her involvement with individual healthcare organizations and episodes of care. A clear distinction should be made between a Computer-Based Patient Record (CPR), Electronic Medical Record (EMR) and EHR. The CPR system provides support for all of the activities and processes involved in the delivery of clinical care, while the EMR is a CPR that is optimized to support ambulatory settings (Rishel et al., 2005). The EHR is the "system that integrates the other systems" and comprises of all the interconnected CRPs and EMRs (Rishel et al., 2005).

The EHR best facilitates the creation of a Longitudinal Patient Record (LPR) for a patient (Green and Bowie, 2004:95). The LPR contains records from different episodes of care, providers and facilities that are linked to form a view, over time, of a patient health care encounters (Green and Bowie, 2004:95). The LPR facilitates clinical decision support, analysis of diagnoses and treatments, and best practice multidisciplinary guidelines to the patient (Green and Bowie, 2004:95).

An LPR is more than a clinical repository and should also merge scheduling, billing, coverage and demographic data (Eichhorst, 2002). It is, however, important to note that most of the work done on the LPR is within the OLTP environment. The patient-centric LPR must be the result of an integrated approach with registration, scheduling, clinical and billing systems working together in a single-application framework (Eichhorst, 2002). Users should be able to, with a single logon to a computer workspace custom-tailored to their organizational role, access this wealth of discrete data. Individual organizations should be in a position to analyze their outcomes and begin to build their own best practice guidelines, contributing to the growing mass of medical knowledge in an actionable way (Eichhorst, 2002).

The FSDOH EPR (Electronic Patient Record) system is embedded within the Medical Record Interface (MRI) module of the hospital information system (supplied by Meditech) that is deployed in the major hospitals. This module provides patient demographic details and is referenced by all the Meditech modules, including the Medical Practice Management (MPM) Suite. This module is used for capturing all the information on the patients taking part in the ART programme. However, this EPR is OLTP based and is implemented on Massachusetts General Hospital Utility Multi-Programming System (MUMPS) technology. The problem with this OLTP based EPR is that it is not designed with decision-making in mind and lacks the necessary online analytical processing (OLAP) functionality (Ebidia et al., 1999). In addition to these shortcomings the OLTP based EPR also contains no information on TB encounters, NHLS blood results, hospital visits from regional hospital information systems and/or linkage with the Department of Home Affairs. The information contained in these independent systems can contribute to the understanding of the long term effects of ARV treatment on patients. The complete picture can assist in the study of interdependencies and effects of other clinical encounters on the effectiveness of ARV medication. Therefore, what was required by the FSDOH was an LPR for each ARV patient. The remainder of the paper will report on how this was achieved.

2.   METHODOLOGY

The first step was to follow a data warehousing approach that integrated all the disparate systems into one system, specifically focused on the ART programme. The data warehouse consisted of several data marts, each corresponding to one of the disparate systems (Kotze and McDonald, 2010b). This had the immediate advantage that health managers could extract strategic information from these systems, using a single business intelligence tool. The original design, however, did not allow all the data marts to be linked or to trace a specific patient across all the data marts. Even though each data mart had a patient dimension, the structure of the dimension was determined by the source system. A closer inspection of these dimensions revealed the following problems:

1. Each operational data source had different / unique patient identifiers for identifying a patient and in some of the operational data sources a unique patient identifier was totally absent.
2. Furthermore, the attributes for each patient were different among each operational data source.
3. Additionally, several of the attributes were spelled differently.

In order to meet the goal of an LPR, a uniformed (consolidated and conformed) patient dimension was required to link all the fact tables (in each data mart) and, in theory, produce a coherent and complete history of treatment or clinical encounters of each patient. A methodology was required to solve these problems.

Record linkage is a technique that is used to identify the same entity in one or more files when a single common identifier is absent (Cole, 2003). This is exactly what was required in order to identify the same patient in the different data marts. Record linkage has previously been applied successfully in health care (Bernillon, Lievre, Pillonel, Laporte and Costagliola, 2000; Grannis, Overhage, Hui and McDonald, 2003; Nitsch, Morton, DeStavola, Clark and Leon, 2006). In all of these cases record linkage were, however, used to *link files*. In this case, however, it was the first time that it was used to *link dimensions in a data warehouse*.

Currently, there are three commonly used computerized record linkage approaches: Match-Merge (Cole, 2003), Deterministic record linkage and Probabilistic record linkage (Winkler, 2005). Deterministic and Probabilistic record linkage will be discussed briefly in the next few paragraphs. The Match-Merge approach was irrelevant to this study and subsequently omitted.

Deterministic record linkage requires an exact match of identifying information, but uses *multiple* criteria to establish a match (Cole, 2003). This method generates links or a set of rules on the basis of a full agreement of a

unique identifier or a set of common identifiers (Li, Quan, Fong and Lu, 2006; Grannis, Overhage and McDonald, 2004). This minimizes the uncertainties in the match between two databases since only a complete match on a set of personal variables is accepted at the cost of lowering the linkage rate (Cole, 2003).

Probabilistic record linkage identifies a match between records based on a formal statistical model and a linkage is made based on a calculated statistical probability (Cole, 2003). A probability is used to determine whether a pair of records approximately refers to the same individual (Winkler, 1999) and when the probability of a match exceeds a certain threshold, the linkage is made (Cole, 2003). The advantage of probabilistic record linkage is that it uses all available identifiers to establish a match (e.g., name, sex, date of birth, social security number, race, address, phone number, etc.) and does not require identifiers to match exactly. Identifiers that do not match exactly are assigned a "distance" measure to express the degree of difference between files. Each identifier is assigned a weight and the total weighted comparison yields a score, that is used to classify records as linked, not linked, or uncertainly linked according to whether the statistical probability of a match exceeds a certain threshold (Winkler, 1999).

Both probabilistic and deterministic approaches were used to link the patient dimensions of the different data marts. For a detailed description of how this was achieved, the reader is referred to research done by Kotze and McDonald (2010a).

Once the patient dimension tables were linked, the development of the LPR algorithm followed. The LPR algorithm's main outcome was to bring all the information of the following data marts together in a cohesive view:

- Clinical encounters from the Hospital Information System implemented in regional and district hospitals
- Clinical encounters from the Hospital Information System implemented in hospitals in the main centers Meditech System
- ARV clinical encounters
- Tuberculosis encounters
- Notifiable Diseases encounters
- Blood test results from NHLS national database

The LPR algorithm was developed in Oracle 10 release 2 using native SQL and PL/SQL. A decision was taken to develop a web-based application on the existing Oracle Portal platform that is being used in the FSDOH. Oracle PL/SQL and Oracle Web Toolkit were used for this development to ensure a low-bandwidth solution accessible to any user.

Access to the web-based LPR was established by requesting a URL from

within the Clinic Information System when the user selects the Link LPR button. The parameters Username and Patient's were passed to the LPR web application (see Figure 1) as hidden parameters to enable automatic user login.
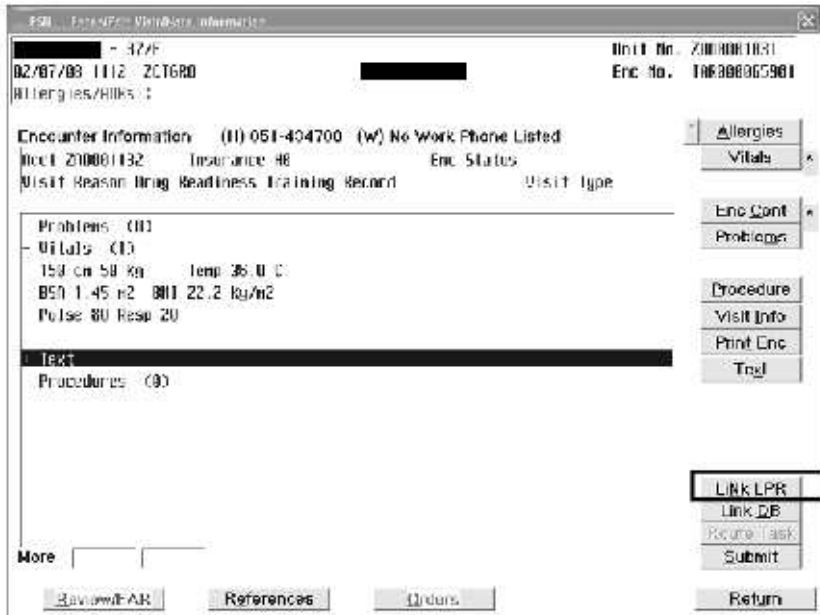

Figure 1: Screenshot of accessing the LPR from an online system.

Patient confidentiality is an important issue when dealing with sensitive topics such as HIV and AIDS. In South Africa the issue is complicated further by the fact that HIV is not a notifiable disease and therefore the HIV status of a patient is confidential. In an attempt to secure the data, additional security measures were implemented by testing a user's occupation before displaying a patient's record. Only nurses or clinical staff would be allowed to view the patient's record. Anyone else would be blocked and a security message would be sent to the data warehouse to record a 'possible' bridge of security. All these nurses or clinical staff must adhere to the medical ethical standards as employees of the FSDOH.

3.   RESULTS

The following series of screenshots (see figures 2 to 6) are based on the results of the web-based LPR application. It will demonstrate how all the incidents were combined into a single screen, easy-to-use by a clinician or nurse.

All the screens are in respect of the same patient. Figure 2 first of all displays the demographics of the patient and then shows all the ARV incidents. The FSDOH ARV programme is a nurse-driven model and all visits to clinics and hospitals are recorded in the Clinic Information System (CIS). All the visits and dates are now combined and shown together. If there were visits belonging to the same patient, but with a different patient identifier, it will also be shown here, because the linkage exercise combined the patient records.

Figure 2: ARV Incidents from the LPR

Figure 3 shows that there were no records from the Notifiable Diseases data mart for this particular patient. It then continues with the results of the patient's blood tests, showing the viral load (VL) and the cluster of differentiation 4 (CD4) count. In a healthy adult, a normal CD4 count can vary a great deal, but is typically between 600 to 1200 cells per cubic millimeter of blood. Values between 600 and 350 in an HIV+ person are considered "very good". Values between 350 and 200 indicate that the immune system is weakened and the HIV+ person may therefore be at increased risk of infection and illness. The clinicians may consider ART. If the values are less than 200, the immune system is weakened severely and the HIV+ person is at a much greater risk of opportunistic infections. Figure 3 shows that the CD4 went from 232 to 136 while the VL went from 210000 to 520000.

Figure 3: Notifiable Diseases, NHLS and MPM Blood Results from the LPR

Figure 4 displays the ARV drug received by the patient. It shows that the patient is receiving Regimen 1a which is the first line regimen for ART. All three of the regimens were received when visiting a clinic and the patient was seen by a professional nurse. The patient's weight was also recorded and Figure 4 shows that the patient started on 41 kg and was weighing 49 kg at the last visit recorded.

Figure 4: Regimen, Drugs and Weight Records from the LPR

The patient's hospital visits can be seen in Figure 5. It indicates that the patient has visited both the Pelonomi Regional Hospital and the Diamant District Hospital. The FSDOH uses a patient referral model and one can see that the patient visited the closest District Hospital on the 7th March and the Regional Hospital on the 15th March. The patient was diagnosed with bacterial meningitis.

Figure 5: Meditech and PADS Hospital Visits from the LPR

Information in respect of the patient's TB treatment is shown in Figure 6. This is very important information, because patients that have enrolled in the TB treatment programme are not allowed to receive ART at the same time. Figure 6 shows that this patient was indeed diagnosed with TB in September 2007 and the TB classification was TB Pulmonary.

Finally, all patients receiving ART are verified against the National Birth and Death register to confirm the status of these patients. In figure 6 one can see that this patient was still alive when the verification was done on the 20th October 2008.

Figure 6: TB Register and Home Affairs from the LPR

4.   DISCUSSION

When this article was written, insufficient data (< 25 rows) was available to allow for a comprehensive pattern analysis on the usage of the LPR based on scientific research methods. Patient confidentiality was another obstacle, and the LPR usage could not be studied in detail. Finally, the FSDOH Top Management only started the process of approving the LPR, and it was therefore not implemented at every hospital or clinic providing ARV treatment.

However, in order to test its basic functionality, the LPR was made available to the FSDOH research partners at the University of Cape Town (UCT) Lung Institute and the Medical Research Council of South Africa (MRC).

A Senior Manager of Biomedical Informatics Research at the MRC had the following to say: "Thank you very much for developing the Longitudinal Patient Record (LPR) application and making it available to us. This is a major achievement and a substantial help to our efforts to strengthen service delivery for patients accessing the Free State public sector antiretroviral treatment program". He concluded by stating that this data resource is unsurpassed in terms of its ability to support cohort analysis and that the MRC hopes to also develop a patient failure summary sheet on top of this data. He also pointed out that he believes the integration of ART, TB and the Home Affairs vital registration system to be a first in South Africa. A Senior Researcher of the Free State Antiretroviral Therapy programme at the UCT Lung Institute had the following to say: "I think you have really done something incredible innovative - a first for South Africa or Africa for that matter". It has to be pointed out that several key features of the LPR were used to extract information for research in evaluating the Free State ARV programme (Ingle, May, Eubel, Timmerman, Kotze, Bachmann, Sterne, Egger and Fairral, 2010a; Ingle, May, Eubel, Timmerman, Kotze, Bachmann, Sterne, Egger and Fairral: 2010b; May, Ingle, Timmerman, **Kotze,** Uebel, Bachmann, Sterne, Egger, Fairral, 2009)

From this study it is clear that the scope of the proposed LPR is focussed around patients enrolled for ARV treatment. The algorithm therefore first focussed on ARV patients and was then linked to other data marts' patient dimensions. This could be seen as a limit to this approach, but within the context of the study and the available resources usually associated with a constraint environment, it is still quite applicable and produced the expected result. By adopting a web-based application to deploy the algorithm, users from the FSDOH were able to make enquiries on patients from within the Meditech system as well as external systems. This approach can be considered as a possible way of merging an LPR with an existing operational system. Both systems work towards a common goal of improving clinical decision making for patients.

5.   CONCLUSION

Managing the ARV treatment programme thoroughly is of the utmost importance. In this endeavor strategic information plays a major role. The data warehouse that was developed for this purpose consisted of several independent data marts, each representing a system closely related to HIV/AIDS/ARV. Even though valuable information could be extracted from these data marts using business intelligence tools, it was impossible to link the different data marts and identify the same patient across the different data

marts. In effect an LPR was hidden in the different data marts. The challenge was to extract it.

Each of the patient dimensions of the data marts had different structures and different identifiers. Record linkage has previously been used in similar situations and, therefore, it was applied to link the patient dimensions of the different data marts to produce a coherent and complete history of treatment or clinical encounters of the ARV patients. An algorithm was subsequently developed and implemented in a web based application to display LPRs. To our knowledge this is a new and unique approach to solve the challenge that was posed.

This application is an invaluable tool for researchers trying to determine the long term effects of ARV drugs. It captures a patient from his first visit to a clinic, all follow-up visits, all blood test results, all ARV regimens received, visits to regional and district hospitals, whether the patient is receiving TB treatment, if he/she is still alive and numerous other pieces of important information. It was previously impossible to obtain this information.

Although no scientific evaluation was performed on the LPR due to the unavailability of sufficient data for analysis, impressions were provided by the FSDOH external research partners. Both these partners' (MRC and UCT Lung Institute) impressions about the usability and functionality of the LPR were very positive.

6.   REFERENCES

Bernillon, P., Lievre, L., Pillonel, J., Laporte, A. and Costagliola, D. 2000. Record-linkage between two anonymous databases for a capture-recapture estimation of underreporting of AIDS cases: France 1990-1993. *Journal of Epidemiology*, 29(1), 168-174.

Cole, N. 2003. Feasibility and Accuracy of Record Linkage to Estimate Multiple Program Participation: Record Linkage Issues and Results of the Survey of Food Assistance Information Systems, E-FAN-03-008-1, Volume 4, 1-78. United States Department of Agriculture, Economic Research Service. Available from:
    http://www.ers.usda.gov/publications/efan03008/efan03008-1/efan03008-1b.pdf [accessed on 21 August 2006].

Ebidia, A., Mulder, C., Tripp, B. and Morgan, M.W. 1999. Getting data out of the electronic patient record: critical steps in building a data warehouse for decision support. Proceedings of the 1999 AMIA Annual Symposium, 745-749. Available from: http://www.amia.org/pubs/symposia/D005548.PDF [accessed on 27 July 2006].

Eichhorst, B. 2002. Patient-centric HIS - A healthcare information system based on a longitudinal patient record provides benefits to patients and

clinicians, administrators and IT staff as well. *Health Management Technology*, 23(4), 40-42.

Grannis, S., Overhage, J., Hui, S. and McDonald, C. 2003. Analysis of a Probabilistic Record Linkage Technique without Human Review. Proceedings of the 2003 AMIA Annual Symposium. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?tool=pubmed&pubmed id=14728174 [accessed on 26 January 2007].

Grannis, S., Overhage, J. and McDonald, C. 2004. Real World Performance of Approximate String Comparators for use in Patient Matching. *11*(1), 43-47.

Green, M.A. and Bowie, M.J. 2004. Essentials of Health Information Management. Thomson Delmar Learning. United States of America, New York.

Ingle, S. M., May, M., Uebel, K., Timmerman, V., Kotze, E., Bachmann, M., Sterne, J. A., Egger, M. and Fairral, L.; for the International epidemiologic Databases to Evalute AIDS Southern Africa (IeDEA-SA). 2010a. Differences in access and patient outcomes across antiretroviral treatment clinics in the Free State province: A prospective cohort study. *SAMJ*, 100(10), 675-681.

Ingle, S. M., May, M., Uebel, K., Timmerman, V., Kotze, E., Bachmann, M., Sterne, J. A., Egger, M. and Fairral, L.; for IeDEA-Southern Africa. 2010b. Outcomes in patients waiting for antiretroviral treatment in the Free State Province, South Africa: prospective linkage study. *AIDS 2010*; 24(17), 2717-2725.

Kotze, E. and McDonald, T. 2010a. Using record linkage to conform the patient dimension of a data warehouse to manage antiretroviral therapy. Proceedings of the IASTED International Conference on Health Informatics (AfricaHI 2010). Gaborone, Botswana, 6-8 September 2010.

Kotze, E. and McDonald, T. 2010b. Effectiveness of a Business Intelligence Solution to Manage the Antiretroviral Theraphy Programme. Proceedings of the 4th International Joint Conference on Biomedical Engineering Systems and Technologies (HEALTHINF 2011). Rome, Italy, 26-29 January 2011.

Li, B., Quan, H., Fong, A. and Lu, M. 2006. Assessing record linkage between health care and Vital Statistics databases using deterministic methods. *BMC Health Services Research*, *6*, 48-58.

May, M., Ingle, S., Timmerman, V., Kotze, J.E., Uebel, K., Bachmann, M., Sterne, J., Egger, M., Fairral, L. 2009. Determinants of waiting times for ART in the Free State Province, South Africa: prospective cohort study with

retrospective database linkage. 5th IAS Conference on HIV Pathogenesis, Treatment and Prevention, Cape Town, 19-22 July 2009

National Department of Health, South Africa. 2003. Cabinet's decision on the operational plan for comprehensive care and treatment of people living with HIV and AIDS, Pretoria, South Africa. Available from: http://www.doh.gov.za/docs/pr/2003/pr1119.html [accessed on 16 July 2008].

Nitsch, D., Morton, S., DeStavola, B.L., Clark, H. and Leon, D.A. 2006. How good is probabilistic record linkage to reconstruct reproductive histories? Results from the Aberdeen Children of the 1950s study. *BMC Medical Research Methodology*, 6, 15-24.

Rishel, W., Thomas, J., Handler, M.D. and Edwards, J.A. 2005. Clear Definition of the Electronic Health Record. Gartner document G00130927. Available from: http://www.gartner.com/ [accessed on 8 November 2008].

UNAIDS. 2010a. UNAIDS Report on the Global AIDS Epidemic. Available from:
http://www.unaids.org/en/media/unaids/contentassets/documents/unaidspublication/2010/20101123_globalreport_en.pdf [accessed on 21 November 2011].

UNAIDS. 2010b. South Africa. Available from:
http://www.unaids.org/en/regionscountries/countries/southafrica/ [accessed on 21 November 2011].

Winkler, W.E. 1999. The State of Record Linkage and Current Research Problems. *Technical Report, Statistical Research Division*, U.S. Bureau of the Census. Available from: http://www.census.gov/srd/papers/pdf/rr99-04.pdf [accessed on 26 January 2007].

Winkler, W.E. 2005. Overview of Record Linkage and Current Research Directions. *Research Report, Statistical Research Division*, U.S. Bureau of the Census. Available from: http://www.census.gov/srd/papers/pdf/rrs2006-02.pdf [accessed on 12 February 2007].