PHARMACOEPIDEMIOLOGY AND DRUG SAFETY 2012; **21**: 407–414
Published online 19 October 2011 in Wiley Online Library (wileyonlinelibrary.com) **DOI**: 10.1002/pds.2252

ORIGINAL REPORT

Records linkage of electronic databases for the assessment of adverse effects of antiretroviral therapy in sub-Saharan Africa

Catherine Corbell[1]*, Ishmael Katjitae[2], Assegid Mengistu[3], Francis Kalemeera[3], Evans Sagwa[4], David Mabirizi[4], Jennie Lates[3], Jude Nwokike[5], Sherrilyne Fuller[6] and Andy Stergachis[7,8]

[1]*Department of Pharmacy, School of Pharmacy, University of Washington, Seattle, WA, USA*
[2]*Ministry of Health and Social Services, Internal Medicine, Windhoek, Namibia*
[3]*Ministry of Health and Social Services, Pharmaceutical Services, Windhoek, Namibia*
[4]*Management Sciences for Health/Center for Pharmaceutical Management/Strengthening Pharmaceutical Systems, Windhoek, Namibia*
[5]*Management Sciences for Health/Center for Pharmaceutical Management/Strengthening Pharmaceutical Systems, Arlington, VA, USA*
[6]*Division of Biomedical and Health Informatics, School of Medicine, University of Washington, Seattle, WA, USA*
[7]*Departments of Epidemiology and Global Health, School of Public Health, University of Washington, Seattle, WA, USA*
[8]*Global Medicines Program, University of Washington, Seattle, WA, USA*

ABSTRACT

Purpose In 2009, the Ministry of Health and Social Services in Namibia decided to conduct a confirmatory assessment of the risk of anemia associated with zidovudine (AZT)-based highly active antiretroviral therapy (HAART) using records contained in three electronic databases. These records did not share a unique identifying number. The first step was to apply probabilistic record linkage methods to link records in the three databases.

Methods Records of persons, aged 19–65 years, newly initiated on HAART between January 2007 and June 2008, were selected from a pharmacy electronic dispensing tool (EDT) and linked to an electronic medical records database (ePMS) and a laboratory database (MEDITECH). Using the paper-based clinical record as the gold standard, we measured the sensitivity of the starting HAART regimen, that is, proportion of AZT users in the clinical record correctly identified in electronic record, and specificity of severe anemia, that is, proportion of non-cases of severe anemia in the clinical records correctly identified in the electronic record. Kappa and intraclass correlation coefficients were used to determine reliability.

Results A total of 12 358 records were selected from EDT. Seventy-six percent and 58% of EDT records were linked to ePMS and MEDITECH, respectively. The sensitivity of the starting HAART regimen was 98%, whereas specificity of severe anemia was 100%. The reliability scores for variables including weight, hemoglobin, and CD4 counts were moderate to perfect and ranged from 0.59 to 0.99.

Conclusion Probabilistic record linkage methods were effective for records linkage in this sub-Saharan African setting. Copyright © 2011 John Wiley & Sons, Ltd.

KEY WORDS—records linkage; pharmacovigilance; Namibia; antiretroviral therapy; electronic databases

Received 16 February 2011; Revised 21 July 2011; Accepted 17 August 2011

INTRODUCTION

The adverse effects of antiretroviral therapy (ART) are well enumerated in clinical trials and observational studies conducted in Europe, North America, and Australia. Such studies may not be generalizable to African settings where special challenges may exist such as malnutrition and co-infections with tuberculosis and malaria. In 2009, the Ministry of Health and Social Services (MoHSS) of Namibia partnered with the US Agency for International Development-funded Strengthening Pharmaceutical Systems (SPS) program to investigate the risk of anemia associated with zidovudine (AZT). AZT is a first-line antiretroviral drug and is part of highly active antiretroviral treatment (HAART). The study was proposed as a retrospective cohort study using information obtained from existing electronic and manually maintained data sources. Record linkage was required, in which clinical and medicines information belonging to the same individual would be linked across databases and over time.

Methods for record linkage of electronic databases are either deterministic or probabilistic. Deterministic

*Correspondence to: C. Corbell, Department of Pharmacy, University of Washington, Box 357630, Seattle, WA 98195, USA. E-mail: cww@u.washington.edu

record linkage requires that matching variables be identical for two records to be considered linked and is most effective where databases share a common unique identifying number.[1] Probabilistic record linkage (PRL) does not require matching variables to be identical and uses statistical methods involving probabilities and maximum likelihoods to identify matched from non-matched records.[2] Because deterministic record linkage methods require that matching variables be identical, there tends to be high specificity (i.e., high proportion of true matches) but lower sensitivity (i.e., lower number of linked records), as slight variations in matching variables can result in false non-matches. PRL has high sensitivity but lower specificity, although the linkage algorithm can be modified to trade lower sensitivity for higher specificity.[3,4] PRL is a valid approach for linking records that do not share a common unique identifier and has been utilized in postmarketing surveillance studies assessing the effects of drugs in large patient populations.[5–7]

In Namibia, three databases exist containing clinical, pharmaceutical, and laboratory information for persons with human immunodeficiency virus (HIV) who receive health care through public institutions. A common unique identifying number is not available to link records in these databases. The aim of our study was to apply PRL methods in record linkage of the three electronic databases and assess data quality and completeness and the validity of record linkage.

METHODS

Study objectives

Our study objectives were as follows: (i) to develop an algorithm that would be used for record linkage using PRL methods; and (ii) assess the accuracy of the linkage algorithm by testing completeness, validity (i.e., sensitivity and specificity), and reliability (i.e., kappa statistics and intraclass correlation coefficients (ICCs) of select variables in the linked data file).

Study population

Our study population consisted of persons, aged 19–65 years, who initiated HAART using either one of the first-line regimens: AZT or stavudine (d4T). Only persons who initiated HAART between January 2007 and June 2008 were included. The year 2007 was used as the starting point because that was the year when AZT-based HAART became a first-line therapy in Namibia, in addition to d4T.[8] The period of 2007–2008 was selected to provide at least 1 year follow-up of time from the start of HAART to the end of the study period in 30 June 2009.

Data sources

We used four data sources, consisting of three electronic databases and manually maintained clinical records. The electronic dispensing tool (EDT), an archived, pharmacy-based tool available through SPS and MoHSS, was used to obtain information on HAART exposures. The EDT contains information on the date of HAART initiation, types of HAART regimens dispensed, quantity, strength and dosage form, and HAART dispensing dates. In addition, EDT contains information on select demographic characteristics such as age, gender, and follow-up status, that is, death or lost to follow-up. EDT was first implemented in public ART facilities in 2005,[9] and by 2009, dispensing records for approximately 60 000 HAART users seen at 35 major ART facilities countrywide were contained in EDT.[10] However, EDT was not implemented uniformly across ART facilities, and therefore, an individual's first dispensing record in the EDT did not always correspond with the recorded date of starting HAART. Subsequently, only records with complete dispensing information were included.

The electronic Patient Management System (electronic Patient Management System) repository available through IntraHealth International (Namibia) and MoHSS contains electronically stored medical records for 60 000 HAART users. Implemented in 2007, ePMS captures clinical variables such as weight, World Health Organization (WHO)-defined clinical stage, and CD4 counts. ePMS was used to collect clinical characteristics and demographic data for persons during the study period. MEDITECH is a national laboratory database implemented in 2004 primarily as a billing tool for the Namibia Institute of Pathology (NIP). The NIP is a countrywide public institution that conducts laboratory tests for persons receiving care at public health care facilities. MEDITECH data encompasses all laboratory tests performed through NIP, including full blood counts, CD4 counts, and HIV viral loads. It is reported that NIP performs tests on 3000 laboratory samples per day.[11] MEDITECH was used to obtain laboratory values of hemoglobin (Hb) and CD4 counts. From MEDITECH, records of all persons aged 19–65 years, who received HIV care and who were tested through NIP from July 2006 to June 2009 were identified. This extended period was selected to obtain baseline measurements of Hb and CD4 counts prior to starting HAART for persons who initiated treatment in early 2007 and to obtain follow-up measurements of Hb that were used to define anemia. Records in the three electronic databases did not share a unique identifying number. However, EDT records contained an ART number, which formed the last four digits of the 12-digit ART number in ePMS.

DOI: 10.1002/pds

The Patient Care Booklet (PCB) is a manually maintained structured clinical record containing comprehensive clinical, pharmaceutical, and laboratory information for persons receiving HIV care through public health facilities. Data abstracted from the PCBs were used as the gold standard for validation of record linkage of the three electronic databases. Study approval was received from the University of Washington's institutional review board (IRB) and local IRB in Namibia. All files with personal identifying information were encrypted and stored in password-protected computer files.

Record linkage

Prior to record linkage, certain procedures were carried out on the data files to remove extraneous information and minimize errors. These procedures included deletion of special characters, such as commas and hyphens; parsing of names into last and first name; and standardizing age, gender, and date variables so that were presented in the same way in all the data files. Person names, dates of birth, and ART numbers were examined for data completeness to assess their suitability as matching variables.

PRL was used to link the EDT cohorts to ePMS and MEDITECH. PRL generates two probabilities to link records belonging to the same person: the probability that matching variables agree, if the linked record truly belongs to the same person (m), and the probability that matching variables agree if the linked record belongs to different persons (u).[12] For each linked record, a positive weight is assigned to variables that match using the formula $\log_2 m/u$, whereas a negative weight is assigned to variables that do not match using the formula $\log_2 1-m/1-u$. The generated weights are then summed up to generate linkage scores, which have a bimodal distribution as illustrated in Figure 1.

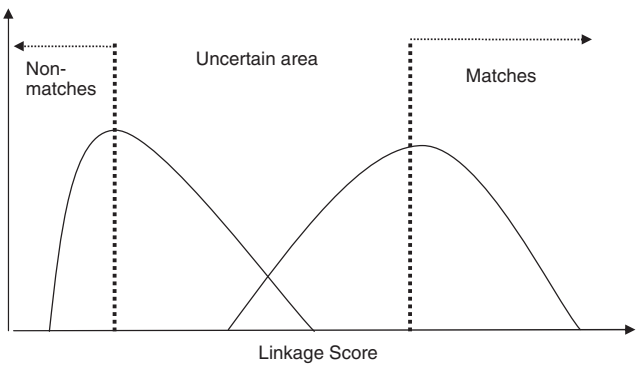

Figure 1. Bimodal distribution of linkage scores following probabilistic record linkage. Matches are record pairs with high linkage scores, whereas non-matches are record pairs with low linkage scores. The uncertain area represents an area of ambiguity as to whether record pairs can be considered as matches or non-matches. Dotted line represents threshold value used to discriminate linked records

Records that are true matches will have higher linkage scores than non-matched records. The bimodal distribution of linkage scores is used to determine a threshold value above which record pairs will be labeled as true matches and a threshold below which record pairs are labeled as non-matches. In between the two thresholds is the uncertain area, which is an area of ambiguity as to whether record pairs can be considered matched or non-matched.

Considering that person names were likely to have inconsistencies in spelling, we used the Soundex coding system, which is a system of coding phonetically similar names.[13] To reduce computational complexity and linkage time, we employed "blocking" in which comparisons of matching variables were performed only within groups that had similar first or last names.[14] Link Plus software was used for record linkage.[15] Microsoft Excel and Stata version 11.0[16] were used to determine the thresholds for discriminating between record matches and non-matches and for reclassifying linked records in the uncertain area as matches or non-matches.

Table 1. Completeness (%) of variables in the data files extracted from electronic databases*

Variables	EDT ($n=28\,775$)	ePMS ($n=43\,637$)	MEDITECH ($n=286\,752$)
Personal identifiers			
ART number	90	99	—
First name	100	100	100
Last name	100	100	100
Date of birth	100	95	97
Gender	100	100	97
Other identifiers			
Facility name	100	93	100
Specimen number	—	—	100
HAART regimen			
Regimens dispensed and dates	59	—	—
Clinical information			
Baseline weight	30	94	—
Baseline clinical status	—	97	—
Baseline WHO clinical stage	—	95	—
Laboratory information			
CD4 count†	—	96	53
Hemoglobin	—	—	71
Date variables			
Date of starting HAART	100	100	—
Date of laboratory measurement‡	—	—	78
Follow-up status	100	100	—

ART, antiretroviral therapy; EDT, electronic dispensing tool; ePMS, electronic Patient Management System; WHO, World Health Organization.
n = number of records.
—, not available.
*The assessment of data completeness was performed for all records ($n=359\,164$) in the three data files.
†Refers to CD4 count at start of HAART.
‡Refers to records that had either a hemoglobin or a CD4 count measurement date.

DOI: 10.1002/pds

Record linkage validation

An external validation process involved a random selection of approximately 100 linked records for comparison of their clinical and diagnostic automated information with paper-based clinical records. Four health facilities located in the capital city and in one peri-urban town were sampled, from which 25 patient records were randomly selected per facility using a random number generator. Information from the clinical records was abstracted using a standardized abstraction form designed specifically for this study. All information contained in the manual abstractions was entered into an electronic database using Epi Info version 3.5.[17]

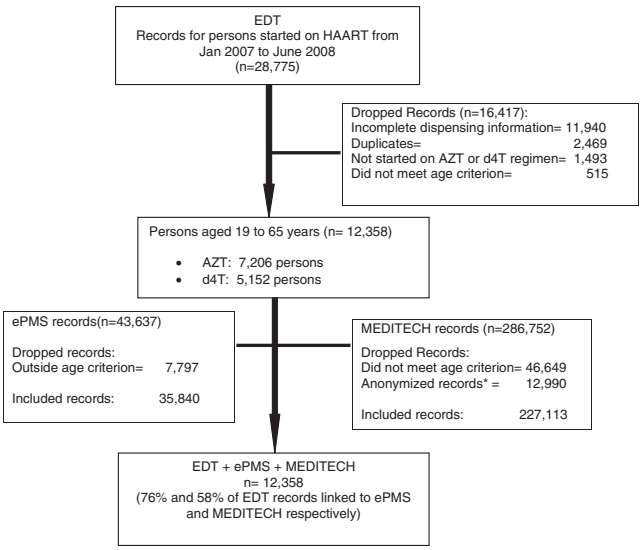

Figure 2. Flowchart of records linkage results for the electronic databases. Abbreviations: AZT, zidovudine; d4T, stavudine; EDT, electronic dispensing tool; ePMS, electronic Patient Management System; HAART, highly active antiretroviral therapy; n, number of records. *Anonymized records were records in which person names were represented as alphanumeric codes.

Using the data abstracted from the manually maintained PCB as the gold standard, we assessed the validity of the starting HAART regimen by measuring sensitivity, that is, proportion of AZT users in the PCB correctly identified in electronic record and specificity, that is, proportion of non-AZT users in the PCB correctly identified in the electronic record. For the outcome variable, severe anemia, we measured specificity, that is, proportion of non-cases of severe anemia in the PCB correctly identified in the electronic record. We focused only on specificity of the outcome variable because poor specificity will result in an underestimate of relative risk.[18] The reliability of variables, such as AZT duration of use, that is, AZT use from start of HAART to 30 June 2009, baseline weight, and CD4 counts was analyzed using Cohen's kappa coefficient for binary variables and the ICC for continuous variables.[19] Variables with an ICC or kappa score of 0.0–0.40 were considered to have poor to slight agreement, 0.41–0.59 had moderate agreement, 0.60–0.79 had substantial agreement, ≥ 0.80 had outstanding agreement, and 1.0 had perfect agreement.[20]

RESULTS

All data files had at least 90% completeness on last and first name, date of birth, and ART number, making these variables suitable as matching variables for record linkage (Table 1). Figure 2 illustrates the results of records linkage. Forty-one percent of EDT records ($n = 11\,940$) had missing dispensing information between the recorded date of starting HAART and the first dispensing record and were subsequently excluded. Excluded records were comparable to included and total records on age, gender and weight at start of HAART (Table 2). In addition, all ART sites were represented in the included records.

Table 2. Characteristics of persons excluded from electronic dispensing tool file for incomplete dispensing information

Variable	Total records ($n = 24\,298$)*	Excluded records ($n = 11\,940$)	Included records ($n = 12\,358$)
Mean age at start HAART, years (SD)	37.3 (8.9)	37.3 (9.0)	37.3 (8.8)
% Female	61.2	61.0	61.3
Mean weight at start of ART	55.5	55.0	55.5
Facilities with highest patient numbers (%)	Onandjokwe (11.6), Oshakati (10.9), Katutura IH (7.3), Katutura HC (5.8)	Oshakati (12.5) Engela (13.5) Katima Mulilo (11.7) Katutura IH (11.2)	Onandjokwe (22.3%) Katutura HC (9.2%) Walvis Bay (8.1%) Oshakati (6.0)
Total number of regions	13	13	13
Total number of ART facilities	35	35	35

ART, Antiretroviral Therapy; HAART, Highly Active Antiretroviral Therapy; HC, Health Center; IH, Intermediate Hospital; SD, Standard Deviation.
n = records of persons started on zidovudine-based or stavudine-based HAART between January 2007 and June 2008.
*There were a total of 28 775 records but 4477 records were excluded because of duplicities or because the HAART regimen at start of HAART was neither zidovudine nor stavudine based or because they did not meet the age criteria of 19–65 years at start of HAART.

DOI: 10.1002/pds

Records with a linkage score less than five were considered as non-matches, whereas those with a linkage score above 11 were considered true matches. Records whose linkage scores fell between 5 and 11 were considered to be uncertain. For record pairs that were considered to be uncertain matches, a manual review revealed that the uncertainty resulted from disparities in dates of birth and/or missing first or last name in one record. Subsequently, we developed STATA code for reclassifying records in the uncertain area as true linkages using the following algorithm: (i) matched on first name or last name, and with dates of birth not more than 6 months apart; (ii) matched on first name and date of birth for records with missing last name; (iii) matched on last name and date of birth for records with missing first name; and (iv) matched on the last four digits of the ART numbers (only for records linkage of EDT and ePMS). This matching algorithm resulted in 76–58% of EDT records finding linkages with ePMS and MEDITECH, respectively.

Hemoglobin values in the linked data file were classified as baseline values if measurements were made up to 6 months prior to starting HAART or up to 1 month after starting HAART. Hb measurements taken more than 1 month after starting HAART were re-categorized into anemia based on the WHO classification of hematological toxicities in HAART.[21]

A summary of data completeness of clinical, laboratory, and HAART variables in the linked data file is given in Table 3. Data completeness for risk factors of anemia such as weight and WHO clinical stage ranged from 70 to 73%. ePMS provided more information on CD4 counts than MEDITECH (69% versus 18%), and was used as the data source for CD4 counts. Forty percent of records in the linked data file had a Hb value at start HAART and 51% of records in the linked data file had a follow-up Hb value. A descriptive analysis of subject characteristics stratified by the availability of baseline Hb values found that 63.5% of persons with baseline Hb values were started on AZT compared with 55.3% of persons with missing baseline Hb values (Table 4).

Table 3. Completeness of variables in the linked data file ($n = 12\,358$)

Variable	Source	Completeness, %
HAART regimens dispensed	EDT	100
Baseline weight	ePMS	73
Baseline WHO clinical stage	ePMS	70
Baseline clinical status	ePMS	70
Baseline CD4 count	ePMS	69
Baseline CD4 count	MEDITECH	18
Baseline hemoglobin*	MEDITECH	40
Hemoglobin at follow-up	MEDITECH	51

AZT, Zidovudine; EDT, Electronic Dispensing Tool; ePMS, electronic Patient Management System; HAART, Highly Active Antiretroviral Therapy; WHO, World Health Organization.
n = Records of persons started on AZT-based or stavudine-based HAART between January 2007 and June 2008.
*Baseline hemoglobin was measured at start HAART up to 6 months prior to start of HAART or up to 1 month after start of HAART.

Table 4. Descriptive analysis of persons with and without baseline hemoglobin values

Variable	Baseline Hb values available ($n = 4746$)	Missing baseline Hb values ($n = 7506$)
% started on AZT	63.5	55.3
Mean age at start of HAART, years (SD)	37.6 (8.6)	37.3 (8.7)
% Female	62.4	61.6
Mean baseline weight, kg (SD)	55.6 (11.6)	56.5 (12.0)
Baseline CD4 count (cells/μL), median (IQR)	149 (100,193)	148 (90,189)
% with available Hb value at follow-up	87	30

AZT, Zidovudine; HAART, Highly Active Antiretroviral Therapy; Hb, Hemoglobin; IQR, Interquartile Range; SD, Standard Deviation.

Table 5. Validity and reliability measures for select variables in linked data file ($n = 94$)

Variable	Database	Measure	95%CI
Gender	EDT	Sensitivity: 92%	81–98%
Starting HAART regimen	EDT	Sensitivity: 98%	92–99%
AZT duration of use*	EDT	ICC: 0.86	0.78–0.93
Severe anemia†	MEDITECH	Specificity: 100%	93–100%
Date of start HAART	EDT	ICC: 0.99	0.99–1.00
Date of birth	EDT	ICC: 0.99	0.90–1.00
Baseline hemoglobin	MEDITECH	ICC: 0.81	0.61–1.00
Baseline CD4	ePMS	ICC: 0.92	0.85–1.00
Baseline weight	ePMS	ICC: 0.96	0.94–1.00
Baseline WHO clinical stage	ePMS	kappa: 0.59	0.46–0.72
Baseline clinical status	ePMS	kappa: 0.74	0.51–0.97

AZT, Zidovudine; CI, Confidence Interval; EDT, Electronic Dispensing Tool; ePMS, electronic Patient Management System; HAART, Highly Active Antiretroviral Therapy; Hb, Hemoglobin; ICC, Intra Class Correlation; WHO, World Health Organization.
n = number of clinical records abstracted for record linkage validation.
*Duration from date of starting AZT-based HAART to last recorded dispensing up to 30 June 2009.
†Severe anemia defined as grade 3 or 4 anemia and is based on WHO categorization of hematological toxicities in HAART.[21]

DOI: 10.1002/pds

In addition, 87% of persons with baseline Hb values had follow-up Hb values available compared with 30% of persons with missing baseline Hb values. Persons were nonetheless comparable on other select clinical and demographic characteristics.

Of the 100 records randomly selected for external validation, 94 records were available at the health facilities for manual review. The validity and reliability assessments for key data variables are summarized in Table 5. The sensitivity of the variable of HAART regimen at start of treatment and specificity of the severe anemia variable were 98 and 100%, respectively. The reliability scores for variables including AZT duration of use, weight, Hb, and CD4 counts were moderate to perfect and ranged from 0.59 to 0.99.

Of 12 records that were missing baseline Hb values in the electronic data file, four records were missing baseline Hb values from the clinical record. Furthermore, of 13 records that were missing follow-up Hb values in the electronic data file, three records were missing follow-up Hb values in the clinical record.

DISCUSSION

To our knowledge, our study marked the first time record linkage of electronic databases was being attempted in this sub-Saharan African setting. The linked data file was used in a retrospective cohort assessment of the incidence and risk of anemia associated with AZT-based HAART, the results of which will be reported elsewhere. Although 41% of EDT records were excluded because of missing HAART data, we did not identify any systematic differences between included and excluded records on age, gender, and weight. In addition, all 35 ART facilities were represented in the included records, enhancing generalizability. Nevertheless, patients with excluded records may have differed from those included based on other unmeasured characteristics.

Twenty-four percent of EDT records had missing clinical information as a result of not being linked with ePMS, whereas 42% lacked laboratory information from not being linked with MEDITECH. Missing data for variables obtained from ePMS, such as CD4 count, weight, and WHO clinical stage, was 27–31%, which could bias our relative risk estimates toward the null. To mitigate any potential biases arising from missing data, we planned a nested case–control study using a subsample of anemia cases and controls identified from the linked data file.

We found that persons with available baseline Hb values were more likely to have been started on AZT than d4T and to have Hb values at follow-up. This finding was not surprising because baseline Hb values are required for making clinical decisions on whether to initiate treatment with AZT.[8] Also, it is possible that healthcare providers judiciously monitored Hb levels in AZT users as a result of prior knowledge of the anemia risk associated with AZT. An analysis of the incidence of anemia using the linked data file would thus need to be stratified by the availability of baseline Hb values. Still, missing follow-up Hb values in our data set might result in an underestimate of the incidence of anemia among AZT users. We plan to compare our incidence rates of anemia in AZT users with findings reported in other African settings.[22,23]

We compared our study findings with others that have used PRL methods to link pharmaceutical and clinical databases. Herk-sukel *et al*.[6] used PRL to link a cancer registry to a database containing pharmaceutical and laboratory information in the Netherlands. Although 85% of cancer records were linked, 80 and 45% of linked records contained laboratory and pharmaceutical information, respectively. This finding was reflected in our study, in which data incompleteness was a limitation of record linkage. Another Dutch study linking pharmaceutical to morbidity data was able to link 93% of the records through PRL methods, although there was no information provided on data completeness and accuracy.[5] PRL methods can be complex and involve several time-consuming steps. Record linkage would have been simpler if databases shared a unique identifying number that could be used for deterministic record linkage. However, deterministic record linkage can result in few linkages if there are errors in the identifying number or if there are missing data. In one UK study for example, 24% of pharmacy records could not be linked with morbidity records because of missing unique identifying numbers.[24]

Through concerted efforts by ministries of health and non-governmental organizations, countries in sub-Saharan Africa have begun using electronic databases to monitor the performance of antiretroviral programs.[25,26] These databases are used to provide aggregate data on key indicators such as the number of people on ART, persons lost to follow-up and deaths. Countries such as Kenya, Rwanda, Malawi, Zambia, and South Africa are utilizing electronic medical records to record clinical information of persons receiving HIV care.[27–30] The EDT pharmacy database has been implemented in the Ivory Coast, Ethiopia, Kenya, Rwanda, Tanzania, and Zambia.[31] The EDT is a valuable tool that can be used to provide estimates of HIV prevalence, ART exposure, and the cost of HIV care and, as shown here, for pharmacoepidemiological studies. In future, these methods of record linkage will be useful for assessing possible renal toxicities from tenofovir, the newly recommended first-line ART in the 2010 WHO treatment guidelines.[32] Data quality control checks should be routinely conducted

DOI: 10.1002/pds

to avoid data incompleteness and inaccuracies that can undermine the usefulness of databases for pharmacoepidemiological studies.

KEY POINTS

- Record linkage of automated databases was conducted for the first time in Namibia.
- Records of persons, aged 19 to 65 years, newly initiated on highly active antiretroviral therapy (HAART) were identified from a pharmacy electronic dispensing tool (EDT), and linked to an electronic medical records database (ePMS) and the national laboratory database (MEDITECH).
- Record linkage of automated databases provided a platform for the assessment of adverse events of HAART.

CONFLICT OF INTEREST

The sponsor of this project had the right of commenting, but the authors retained the right to accept or reject comments or suggestions.

ACKNOWLEDGEMENTS

This study was funded by the US Agency for International Development (USAID) through Management Sciences for Health's Strengthening Pharmaceutical Systems program, under the terms of Cooperative Agreement no. GHN-A-00-07-00002-00. The contents are the responsibility of the authors and do not necessarily reflect the views of USAID. The authors would like to offer their sincere gratitude to the following people/organizations for assistance and support in this work: The management of the Ministry of Health and Social Services (MoHSS); the Technical Advisory Committee (TAC), MoHSS; Directorate of Special Programmes, MoHSS; Therapeutics Information and Pharmacovigilance Center (TIPC), MoHSS; USAID/ Namibia; Management Sciences for Health (MSH) and the Strengthening Pharmaceutical Systems program (SPS); Namibia Institute of Pathology (NIP); Intrahealth International; the staff of Katutura Health Center, Windhoek Central Hospital, Intermediate Hospital Katutura, and Saint Mary's Catholic Health Services Hospital, Rehoboth; and Professors David Blough and Louis Garrison from the University of Washington.

REFERENCES

1. Pacheco AG, Saraceni V, Tuboi SH, Moulton LH, Chaisson RE, Cavalcante SC, *et al*. Validation of a hierarchical deterministic record-linkage algorithm using data from 2 different cohorts of human immunodeficiency virus-infected persons and mortality databases in Brazil. *Am J Epidemiol* 2008; **168**(11): 1326–32.
2. Meray N, Reitsma JB, Ravelli AC, Bonsel GJ. Probabilistic record linkage is a valid and transparent tool to combine databases without a patient identification number. *J Clin Epidemiol* 2007; **60**(9): 883–91.
3. Fonseca MG, Coeli CM, de Fatima de Araujo Lucena F, Veloso VG, Carvalho MS. Accuracy of a probabilistic record linkage strategy applied to identify deaths among cases reported to the Brazilian AIDS surveillance database. *Cad Saude Publica* 2010; **26**(7): 1431–8.
4. Nakhaee F, McDonald A, Black D, Law M. A feasible method for linkage studies avoiding clerical review: linkage of the national HIV/AIDS surveillance databases with the National Death Index in Australia. *Aust N Z J Public Health* 2007; **31**(4): 308–12.
5. Herings RM, Bakker A, Stricker BH, Nap G. Pharmaco-morbidity linkage: a feasibility study comparing morbidity in two pharmacy based exposure cohorts. *J Epidemiol Community Health* 1992; **46**(2): 136–40.
6. van Herk-Sukel MP, van de Poll-Franse LV, Lemmens VE, Vreugdenhil G, Pruijt JF, Coebergh JW, *et al*. New opportunities for drug outcomes research in cancer patients: the linkage of the Eindhoven Cancer Registry and the PHARMO Record Linkage System. *Eur J Cancer* 2009; **46**(2): 395–404.
7. van Leeuwen MT, Grulich AE, McDonald SP, McCredie MR, Amin J, Stewart JH, *et al*. Immunosuppression and other risk factors for lip cancer after kidney transplantation. *Cancer Epidemiol Biomarkers Prev* 2009; **18**(2): 561–9.
8. Republic of Namibia. *National Guidelines for Antiretroviral Therapy* (2nd edn). Ministry of Health and Social Services: Windhoek, Namibia, 2007.
9. Rational Pharmaceutical Management Plus. *ART Dispensing Tool: User's Guide, Version 1.3*. Management Sciences for Health: Arlington, Virginia, 2007.
10. Hong SY, Jonas A, Dumeni E, Badi A, Pereko D, Blom A, *et al*. Population-based monitoring of HIV drug resistance in Namibia with early warning indicators. *J Acquir Immune Defic Syndr* 2010; **55**(4): 27–31.
11. Namibia Instititute of Pathology. Annual report 2006/2007. Windhoek: Namibia Institute of Pathology Ltd.; 2007.
12. Fellegi I, Sunter A. A theory for record linkage. *J Am Stat Assoc* 1969; **64**: 1183–1210.
13. Smith ME, Newcombe HB. Automated follow-up facilities in Canada for monitoring delayed health effects. *Am J Public Health* 1980; **70**(12): 1261–8.
14. Kelley RP. Blocking considerations for record linkage under conditions of uncertainty. Proc. Social Statistics Section American Statistical Association; 1984.
15. Centers for Disease Control. *Link Plus version 2.0* (2.0 edn). Centers for Disease Control and Prevention: Atlanta, USA.
16. Stata version 11. In. 11.0 ed. Texas, USA: StataCorp.
17. Epi Info version 3.5.1. In. 3.5.1 ed. Atlanta, USA: Centers for Disease Control and Prevention.
18. Thomas K, Noel W. *Epidemiologic methods: studying the occurence of illness*. Oxford University Press Inc: New York, 2003.
19. Strom B (ed.). *Pharmacoepidemiology* (2nd edn). John Wiley & Sons Ltd: Chichester, 1994.
20. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; **33**(1): 159–74.
21. World Health Organization. Antiretroviral therapy for HIV infection in adults and adolescents in resource-limited settings: Recommendations for a public health approach. 2006. [cited 2009 May]; Available from: http://www.who.int/hiv/pub/guidelines/artadultguidelines.pdf
22. Hoffmann CJ, Fielding KL, Charalambous S, Sulkowski MS, Innes C, Thio CL, *et al*. Antiretroviral therapy using zidovudine, lamivudine, and efavirenz in South Africa: tolerability and clinical events. *AIDS* 2008; **22**(1): 67–74.
23. Moh R, Danel C, Sorho S, Sauvageot D, Anzian A, Minga A, *et al*. Haematological changes in adults receiving a zidovudine-containing HAART regimen in combination with cotrimoxazole in Cote d'Ivoire. *Antivir Ther* 2005; **10**(5): 615–24.
24. Crombie IK, Brown SV, Hamley JG. Postmarketing drug surveillance by record linkage in Tayside. *J Epidemiol Community Health* 1984; **38**(3): 226–31.
25. Nash D, Elul B, Rabkin M, Tun M, Saito S, Becker M, *et al*. Strategies for more effective monitoring and evaluation systems in HIV programmatic scale-up in resource-limited settings: Implications for health systems strengthening. *J Acquir Immune Defic Syndr* 2009; **52** (Suppl 1): S58-62.
26. Ekouevi DK, Karcher S, Coffie PA. Strengthening health systems through HIV Monitoring and Evaluation in Sub-Saharan Africa. *Curr Opin HIV AIDS* 2011; **6**(4): 245–250.
27. Douglas GP, Gadabu OJ, Joukes S, Mumba S, McKay MV, Ben-Smith A, *et al*. Using touchscreen electronic medical record systems to support and monitor national scale-up of antiretroviral therapy in Malawi. *PLoS Med* 2010; **7**(8): 1–6.
28. Fraser HS, Blaya J. Implementing medical information systems in developing countries, what works and what doesn't. *AMIA Annu Symp Proc* 2010: 232–6.

29. Hannan TJ, Rotich JK, Odero WW, Menya D, Esamai F, Einterz RM, *et al*. The Mosoriot medical record system: design and initial implementation of an outpatient electronic record system in rural Kenya. *Int J Med Inform* 2000; **60**(1): 21–8.
30. Siika AM, Rotich JK, Simiyu CJ, Kigotho EM, Smith FE, Sidle JE, *et al*. An electronic medical record system for ambulatory care of HIV-infected patients in Kenya. *Int J Med Inform* 2005; **74**(5): 345–55.
31. Strengthening Pharmaceutical Systems Program. Electronic Dispensing Tool. [cited 2011 May]; Available from: http://www.msh.org/projects/sps/SPS-Documents/upload/edt_flyer_english.pdf
32. World Health Organization. Antiretroviral therapy for HIV infection in adults and adolescents: Recommendations for a public health approach. 2010 [cited 2011 May]; Available from: http://whqlibdoc.who.int/publications/2010/9789241599764_eng.pdf

DOI: 10.1002/pds