

Electronic medical record systems, data quality and loss to follow-up: survey of antiretroviral therapy programmes in resource-limited settings

Mathieu Forster,^a Christopher Bailey,^b Martin WG Brinkhof,^a Claire Graber,^a Andrew Boulle,^c Mark Spohr,^b Eric Balestre,^d Margaret May,^e Olivia Keiser,^a Andreas Jahn^f & Matthias Egger^a for the ART-LINC collaboration of the International Epidemiological Databases to Evaluate AIDS

Objective To describe the electronic medical databases used in antiretroviral therapy (ART) programmes in lower-income countries and assess the measures such programmes employ to maintain and improve data quality and reduce the loss of patients to follow-up.

Methods In 15 countries of Africa, South America and Asia, a survey was conducted from December 2006 to February 2007 on the use of electronic medical record systems in ART programmes. Patients enrolled in the sites at the time of the survey but not seen during the previous 12 months were considered lost to follow-up. The quality of the data was assessed by computing the percentage of missing key variables (age, sex, clinical stage of HIV infection, CD4+ lymphocyte count and year of ART initiation). Associations between site characteristics (such as number of staff members dedicated to data management), measures to reduce loss to follow-up (such as the presence of staff dedicated to tracing patients) and data quality and loss to follow-up were analysed using multivariate logit models.

Findings Twenty-one sites that together provided ART to 50 060 patients were included (median number of patients per site: 1000; interquartile range, IQR: 72–19 320). Eighteen sites (86%) used an electronic database for medical record-keeping; 15 (83%) such sites relied on software intended for personal or small business use. The median percentage of missing data for key variables per site was 10.9% (IQR: 2.0–18.9%) and declined with training in data management (odds ratio, OR: 0.58; 95% confidence interval, CI: 0.37–0.90) and weekly hours spent by a clerk on the database per 100 patients on ART (OR: 0.95; 95% CI: 0.90–0.99). About 10 weekly hours per 100 patients on ART were required to reduce missing data for key variables to below 10%. The median percentage of patients lost to follow-up 1 year after starting ART was 8.5% (IQR: 4.2–19.7%). Strategies to reduce loss to follow-up included outreach teams, community-based organizations and checking death registry data. Implementation of all three strategies substantially reduced losses to follow-up (OR: 0.17; 95% CI: 0.15–0.20).

Conclusion The quality of the data collected and the retention of patients in ART treatment programmes are unsatisfactory for many sites involved in the scale-up of ART in resource-limited settings, mainly because of insufficient staff trained to manage data and trace patients lost to follow-up.

Bulletin of the World Health Organization 2008;86:939–947.

Une traduction en français de ce résumé figure à la fin de l'article. Al final del artículo se facilita una traducción al español. الترجمة العربية لهذه الخلاصة في نهاية النص الكامل لهذه المقالة.

Introduction

Access to antiretroviral therapy (ART) has improved in lower-income countries over the past 4 years as a result of an exceptional commitment by the international community and donor agencies. WHO estimates that about 3 million people were receiving ART in low- and middle-income countries at the end of 2007, a figure representing

a 7.5-fold increase over the previous 4 years.¹ The number of patients starting ART has increased exponentially since 2003 and must continue to do so if the goal of universal access to ART is to be achieved.²

In the absence of curative treatments, lifelong follow-up of patients on ART is required to monitor adherence, treatment response and adverse effects. A growing amount of increasingly com-

plex information needs to be reviewed at each visit, and new data must be added to the record. An important aspect is retention in care: a recent analysis of treatment programmes showed that losses to follow-up have become more common with the scale-up of ART.³ Programmes find it increasingly difficult to follow the growing population of patients and to trace those who do not return to the clinic. Electronic

^a Institute of Social and Preventive Medicine, University of Bern, Finkenhubelweg 11, 3012 Bern, Switzerland.

^b Department of Knowledge Management and Sharing, World Health Organization, Geneva, Switzerland.

^c School of Public Health and Family Medicine, University of Cape Town, Cape Town, South Africa.

^d Institut de Santé Publique, d'Epidémiologie et de Développement, Université Victor Segalen, Bordeaux, France.

^e Department of Social Medicine, University of Bristol, Bristol, England.

^f Lighthouse Clinic, Kamuzu Central Hospital, Lilongwe, Malawi.

Correspondence to Matthias Egger (egger@ispm.unibe.ch).

doi:10.2471/BLT.07.049908

(Submitted: 24 November 2007 – Revised version received: 1 May 2008 – Accepted: 7 May 2008 – Published online: 4 November 2008)

Medical Record (EMR) systems can improve health care by increasing adherence to therapeutic guidelines and protocols, informing clinical decisions and decreasing medication errors.^{4,5} EMR systems allow early identification of patients who miss appointments, thereby facilitating their timely tracing, and provide a platform for operational research.⁶ Little is known about the role of EMR systems in the context of the scale-up of ART in resource-limited settings. A recent review identified the need for studies on the best ways of using information systems to support the expansion of HIV care in such settings.⁷ The objective of this study was to describe the electronic medical databases used in ART programmes in lower-income countries and to assess the measures such programmes employ to maintain and improve data quality and reduce the loss of patients to follow-up.

Methods

Workshop and subsequent survey

In June 2006, representatives of 21 ART programmes from 15 countries (Benin, Brazil, Burundi, Côte d'Ivoire, the Gambia, India, Kenya, Malawi, Mali, Nigeria, South Africa, Thailand, Uganda, Zambia and Zimbabwe) attended a workshop on the use of EMR systems in ART programmes in resource-limited settings. Ten of the 21 programmes participated in the Antiretroviral Treatment in Lower Income Countries (ART-LINC) collaboration, a network of treatment sites of the International Epidemiological Databases to Evaluate AIDS (IeDEA, <http://www.iedea-hiv.org>).⁸⁻¹⁰ The workshop was jointly organized by IeDEA and the Knowledge Communities and Strategies (KCS) unit at WHO and hosted by the Centers for Disease Control and Prevention (CDC, United States Department of Health and Human Services) offices in Entebbe, Uganda.

Based on the workshop, an online questionnaire covering the EMR systems in place, human and electronic resources, reporting systems, data storage, quality control measures and the tracing of patients lost to follow-up was written in English, translated into French and revised after pilot testing in Bern and Bordeaux. The questionnaire is available from biblio@ispm.unibe.ch.

ART treatment programmes that participated in the Entebbe workshop were invited to complete the questionnaire. All sites ($n = 21$) agreed to participate in the survey. WHO's web-based Data Collector system¹¹ was used. The questionnaire was uploaded 20 December 2006, and all sites had responded by 12 February 2007.

Data quality in ART-LINC

Questions from the Entebbe survey were used to create indicators of data quality. First, for each programme, a computation was made of the number of hours dedicated to data entry per week divided by the number of ART patients enrolled. This was done separately for data entry clerks and medical staff. Second, whether or not staff had received training in data management and data quality control was determined. Third, a patient tracing indicator was calculated by adding up the measures implemented by the sites, including the availability of staff dedicated to tracing patients, the use of community-based organizations to trace patients, and the consultation of death registries to determine vital status.

ART-LINC sites that were not represented at the Entebbe workshop ($n = 7$) did not complete the questionnaire but were asked to provide information on these indicators. This allowed us to analyse their influence on the quality of the ART-LINC data. We also used an index of the burden of poverty [the proportion of the country's population living on less than 1 US dollar (US\$) a day]¹² to investigate the possible influence of background material deprivation on data quality.

The quality of ART-LINC data was assessed by defining a set of six key variables and calculating the proportion of missing data for each. The six variables were: age, sex, CDC or WHO clinical stage at baseline, baseline and follow-up, CD4+ lymphocyte (CD4) counts and year of ART initiation. An index was computed by determining, for each site, the median of the percentages missing data of all six variables. Sites were then ranked according to this global missing data index. Principal components analysis showed that all variables loaded heavily on a single component that accounted for 60.1% of the total variance. This suggests that most of the problem

of missing data for the six variables relates to a single phenomenon. A reliability analysis revealed appropriate internal consistency, with a Cronbach's α of 0.83. These analyses supported the use of a combined index rather than a separate analysis for each variable. A second indicator was defined for each site by computing the proportion of patients lost to follow-up at 1 year after starting ART. A patient was considered lost to follow-up if his or her last visit was recorded during the first year after starting ART and the patient had at least 1 year of additional potential follow-up until the closing date of the database. The closing date was defined for each cohort as the date of the most recent follow-up visit recorded in the database.

Statistical analysis

Associations between site characteristics and data quality were investigated using univariate and multivariate maximum likelihood logit models. The number of patients with missing data for key variables and the number lost to follow-up in each programme site was divided by the total number of patients enrolled in the site. Models for clustered data were then used to evaluate predictors. Robust standard errors to account for within-site correlation were used (blogit procedure in Stata version 9.2, Stata-Corp LP, College Station, TX, United States of America). These models assess the likelihood of a patient within a site having missing data or being lost to follow-up. Variables associated with outcomes in univariate models at $P < 0.25$ were entered into the multivariate model. Results were expressed as odds ratios (ORs) with 95% confidence intervals (CIs).

Results

Programme site characteristics

The 21 sites surveyed provided ART to a total of 50 060 patients. The median number of patients on ART per site was 1000 (interquartile range, IQR: 320 to 2398). Table 1 shows selected site characteristics. All programmes except one were in urban facilities; 11 (52%) were in public facilities; 9 (43%) were run by NGOs and one was a private, for-profit clinic. Eighteen sites (86%) received funding from a donor agency

Table 1. General characteristics and data entry staff of 21 ART programmes in low- and middle-income countries participating in a study correlating record-keeping practices with losses to follow-up

Site characteristics				Staff routinely involved in data entry				
Region	Clinic type	Sources of external funding	Patients currently on ART (n)	Weekly clerical staff hours	Weekly medical staff hours	Weekly clerical staff hours (per 100 patients currently on ART)	Weekly medical staff hours (per 100 patients currently on ART)	Training in data management or data quality control
Africa								
Central	NGO	b, e	1275	60	0	4.7	0.0	No
Eastern	Public	a, b, c, e	19320	980	0	5.1	0.0	Yes
	Public	a, e	302	45	45	14.9	14.9	Yes
	Public	None	4163	NA	NA	NA	NA	No
	Public	a, e	1204	40	0	3.3	0.0	Yes
Southern	NGO	a, b, e	3943	24	320	0.6	8.1	Yes
	NGO	a, c, d	814	40	40	4.9	4.9	Yes
	NGO	a, b, d, e	1250	NA	NA	NA	NA	No
	Private	None	100	NA	NA	NA	NA	No
	NGO	e	910	0	30	0.0	3.3	No
Western	Public	b, e	72	2	1	2.8	1.4	No
	Public	b, e	653	20	0	3.1	0.0	No
	NGO	None	2398	6	6	0.3	0.3	No
	NGO	b, e	200	5	23	2.5	11.5	Yes
	Public	e	1800	70	20	3.9	1.1	Yes
	Public	b	320	0	2	0.0	0.6	No
	Public	None	930	40	10	4.3	1.1	Yes
	Public	a	6113	400	350	6.5	5.7	Yes
Asia								
	NGO	b, d	3000	48	48	1.6	1.6	Yes
	NGO	e	293	80	40	27.3	13.7	Yes
South America								
	Public	d	1000	60	20	6.0	2.0	Yes

ART, antiretroviral therapy; NA, not applicable (sites not routinely using a computerized database); NGO, nongovernmental organization.

^a President's Emergency Plan for AIDS Relief.

^b The Global Fund to fight AIDS, Tuberculosis and Malaria.

^c Bill & Melinda Gates Foundation.

^d National Institutes of Health, Bethesda, MD, United States of America.

^e Other.

(mostly the Global Fund to fight AIDS, Tuberculosis and Malaria, the President's Emergency Plan for AIDS Relief and the Bill & Melinda Gates Foundation); 14 (67%) obtained funding from two sources and 5 (24%) from three. One site reported funding from seven different sources.

Overall, 18 (86%) sites routinely used an electronic database. The median number of weekly hours spent on the database per 100 patients on ART was 3.6 (IQR: 1.6–5.1) for clerks and 1.5 (IQR: 0.3–5.7) for medical personnel (physicians or nurses). Four of the 18 sites (22%) exclusively employed clerks for data entry and two (11%) used medical staff only. Thirteen of the 21 sites (62%) had personnel trained in data management or data quality

control. Sixteen of the 18 sites (89%) captured patient data by means of written charts during consultations. Three (14%) entered the data electronically during consultations.

Database characteristics and data quality measures

Among the 18 sites that had an electronic database, 11 (61%) used the same software for data collection and data management, six (33%) used two different packages and 1 site used as many as five because of various research and reporting requirements. Ten sites (56%) used generic software such as Microsoft Access (Microsoft Corporation, Seattle, WA, USA) or FileMaker Pro (FileMaker, Inc., Santa Clara, CA, USA), only 5 sites

(28%) used systems developed for this purpose, such as FUCHIA (Follow-Up of Clinical HIV Infection and AIDS, from Doctors without Borders and EpiCentre) or ESOPE (from *Ensemble pour une solidarité thérapeutique hospitalière en réseau, ESTHER*). A relational database with files for demographic data, clinical events, drugs and follow-up examinations was in place in 12 (67%) of the 18 sites using electronic databases. Only three such sites (17%) used solutions based on a Structured Query Language (SQL)^{13,14} server, which allows management of very large numbers of patients. Nine sites (50%) reported that the database could be linked to other data, including laboratory (8 sites) and pharmacy (7 sites) databases, the nutritional support unit

Table 2. Data quality control and means for patient follow-up observed in a 15-country study correlating medical record-keeping practices with losses to follow-up in ART programmes

Site	Purpose of database	Data quality control and safety					Patient follow-up		
		CD4 counts	Medication	Dates (of birth, laboratory measurements, follow-up visits, death)		Controlled medical vocabulary	Backup	Automatic alerts for scheduling tests and visits	Tracing of patients
				WORM (%)	Bounds (%)				
Africa									
Central	PM & R	D	None	0	0	None	Weekly	No	OR, CBO
Eastern	PM & R	D, B, W, N	FT, W	100	100	ICD-10	Daily	Yes	OR, DR
	R	None	FT	50	0	None	Weekly	No	OR, CBO, DR
	NA	NA	NA	NA	NA	None	NA	NA	CBO
	PM & R	N	None	0	100	None	Daily	No	OR
Southern	PM & R	D, B, W	FT, W	100	100	None	Daily	Yes	OR
	R	None	FT	0	25	ICD-10	Daily	Yes	OR, DR
	R	NA	NA	NA	NA	None	NA	NA	OR, DR
	R	NA	NA	NA	NA	None	NA	NA	None
	PM & R	None	None	0	25	None	Daily	No	CBO
Western	PM & R	D	FT	100	0	None	None	No	None
	PM & R	B	FT	100	0	None	Monthly	No	None
	PM & R	D	None	0	25	None	Daily	No	None
	PM	None	None	0	100	ICD-10	Daily	Yes	OR
	PM & R	W	None	100	0	None	Monthly	No	None
	PM & R	D	None	0	0	None	None	No	None
	PM & R	None	FT	0	0	None	Daily	Yes	OR, DR
	PM & R	None	None	0	0	None	Daily	No	OR
Asia									
PM & R	D, B, W	FT, W	100	100	None	Daily	No	DR	
PM & R	N, B	None	0	25	None	Daily	No	OR, CBO	
South America									
R	B	FT	100	0	None	Daily	No	DR	

ART, antiretroviral therapy; B, bounds system; CBO, community-based organizations; CD4, CD4+ lymphocytes; D, digit checks; DR, death registry; FT, fixed taxonomy; ICD-10, *International statistical classification of diseases and related health problems, 10th revision*; N, numerical alert; NA, not applicable (sites not routinely using a computerized database); OR, outreach team; PM, patient management; R, reporting; WORM and W, Write Once, Read Many.

(2 sites) or the socioeconomic support unit (3 sites). Standardized export formats were available at 6 sites (33%); five used Extensible Markup Language (XML) and one used the Health Level Seven (HL7) standard.

Table 2 summarizes the main purpose of each database. Fourteen sites (78%) stated that both patient management and reporting requirements were important reasons for having the database. Also shown are the measures in place to ensure the quality of the data for selected key variables: CD4 counts, drugs and important dates (date of birth and dates of laboratory measurements, follow-up visits and death). Box 1 defines commonly used measures to improve and ensure data quality, including bounds checking, digit checks, fixed taxonomies, numerical alerts and Write Once, Read Many (WORM)

computer data storage systems. For the variable CD4 cell counts, at least one of the measures listed in Box 1 was in place at 12 (67%) sites. Four (22%) sites reported two or more measures. Digit check and bounds checking were more frequently used ($n = 6$) than WORM systems ($n = 4$) and numerical alerts ($n = 3$). Fixed taxonomies of drugs were in use at 9 (50%) sites, and 3 (17%) sites additionally reported a WORM strategy. Regarding the four key dates, 7 (39%) reported WORM systems and 5 (28%) reported bounds checking for all four dates. Three (17%) reported both WORM and bounds for all four dates' records. Use of controlled medical vocabularies was reported by 3 sites; all used the *International statistical classification of diseases and related health problems, 10th revision* (ICD-10). Twelve sites (67%) performed a daily backup,

4 sites (22%) did a weekly or monthly backup and 2 sites (11%) had no backup strategy in place.

Tracing of patients and missing data

Fifteen (71%) of the 21 sites indicated that they traced patients lost to follow-up. This included outreach teams at 11 (52%) sites, collaboration with community-based organizations at 5 sites (24%) and checking death registry data at 7 (33%) sites. The majority of the registries consulted were local death registries, such as hospital registries. Fourteen sites (67%) indicated they recorded when patients moved to another clinic and transferred their records on these occasions. Among the 18 sites with electronic databases, 5 (28%) had automatic alerts for missed visits.

Analyses of missing data were based on 41 936 patients from 19 sites and analyses of loss to follow-up on 36 149 patients from 18 sites participating in the ART-LINC collaboration. Table 3 shows the proportion of missing values for the key variables that we used to create the missing data index, as well as the proportion of patients lost to follow-up. There was considerable variation across sites, as indicated by wide inter-quartile ranges. Missing data were more frequent for variables relating to laboratory measures than for demographic or clinical information. The median missing data index was 10.9% and the median proportion of patients lost to follow-up at 1 year was 8.5%.

As shown in Fig. 1 and Table 4, training of staff and clerk-hours spent per week per 100 patients on ART were associated with a decreased likelihood of there being missing data. The figure shows that the variance decreases as the amount of time spent on the database increases. About 10 hours per week per 100 patients on ART were required to lower the proportion of missing data for key variables to below 10%. Interestingly, the four programmes with lower clerk-hours spent on data and lower levels of missing data tended to have a strong research component. The amount of time spent by medical staff was only weakly associated with the missing data index (Table 4). The proportion of the population living on less than US\$ 1 per day was also positively associated with missing data. Loss to follow-up was negatively associated with the number of active tracing strategies in place. The effect of the individual

Box 1. Measures to improve the quality of the data collected in clinical databases

Bounds checking: Automated checking of whether or not a number lies within a pre-defined numeric range of possible or likely values.

Check digit: Additional number added to a unique identifier to check for errors when entering identification numbers. The check digit is calculated from the other digits in the identification number and is designed so that it will not match if any of the other digits is incorrect.

Fixed taxonomy: Predefined names assigned to a variable that prevent free text-related problems, including spelling mistakes and inconsistent terminology. Examples include the ATC code for drugs.

Numerical alert: System alerting the operator when a number is not expected (i.e. value out of range, or of a critical nature) which will prompt the operator to verify the number or take other appropriate actions.

WORM (Write Once, Read Many): Any type of data storage to which data can be written to only a single time, but can be read from any number of times. This prevents the user from accidentally or intentionally altering or erasing the data.

ATC, anatomical therapeutic chemical.

measures was similar: In the univariate logit models, the OR for loss to follow-up was 0.58 (95% CI: 0.25–1.34) for the presence of an outreach team; 0.54 (95% CI: 0.24–1.18) for collaboration with community-based organizations; and 0.41 (95% CI: 0.18–0.94) for collaboration with death registries. Finally, there was a positive correlation between the proportion of patients lost to follow-up and the proportion of data missing for key variables: the Spearman rank correlation coefficient (ρ) was 0.51 ($P = 0.031$).

Discussion

This survey indicates that EMR systems could play an important role in the scale-up of ART in lower-income countries. However, it also shows that the quality of the data collected and the retention of patients in treatment programmes are often unsatisfac-

tory, mainly because of staff that is insufficient or inadequately trained to manage data and trace patients lost to follow-up.

Antiretroviral database anarchy?

Most databases relied on software intended for personal or small business use, rather than on more sophisticated systems required for programmes with an exponential growth of patient numbers. Some sites used several software packages (as many as five in one clinic) to meet the requirements of different funding agencies, a practice leading to inefficiency and poor data quality. Our findings support the call for the development of affordable and sustainable solutions for medical record-keeping in ART scale-up sites.¹⁵ International agencies could prevent database anarchy in ART programmes by jointly developing systems able to support clinical care, monitoring and evaluation, and operational research. Suitable data exchange standards that allow for the transfer of patients from one system to another and facilitate the aggregation of data at the country level are also important. Open source databases, such as the Open Medical Record System (OpenMRS),¹⁶ can easily be adapted to local requirements and are increasingly used in the context of the scale-up of ART.^{17–20} OpenMRS is also an example of successful South-South and South-North collaborations.²¹

The price of quality

We examined factors influencing the completeness of the data in the ART-LINC database. The training of staff

Table 3. Percentage of missing data for key variables, missing data index and rate of loss to follow-up in a 15-country study correlating medical record-keeping practices with losses to follow-up in ART programmes

Type of missing data	Percentage Median (IQR)
Age at ART initiation	2.6 (0.0–9.9)
Sex	0.0 (0.0–0.3)
Year of ART initiation	1.0 (0.0–14.1)
Clinical stage (CDC or WHO)	3.5 (0.2–37.2)
Baseline CD4 count	19.8 (7.25–53.0)
Follow-up CD4 count	5.3 (0.7–27.1)
Missing data index ^a	10.9 (2.0–18.9)
Loss to follow-up	8.5 (4.2–19.7)

ART, antiretroviral therapy; CD4, CD4+ lymphocyte; CDC, Centers for Disease Control and Prevention, United States Department of Health and Human Services; IQR, interquartile range.

^a The missing data index was calculated as the median of the percentages of missing data for the six key variables.

was strongly associated with more complete data, but many programmes had no trained staff on site. The time spent on the database by data clerks was also positively associated with more complete data. About 10 hours per week per 100 patients on ART appear to be required for the proportion of missing data for key variables to drop below 10%. The use of medical staff was found to be less effective for improving data quality and is generally an inefficient use of resources. A few simple database features, including bounds checking, check digits, and numerical alert and WORM systems, can improve data quality. However, these measures have not been widely implemented. Similarly, standardized coding using fixed taxonomies or controlled medical vocabularies is not commonly used. Many sites are thus struggling with identifying and correcting data errors and inconsistencies after they occur, which is time-consuming and rarely fully successful. Data security is also of concern, with some sites not performing backups regularly.

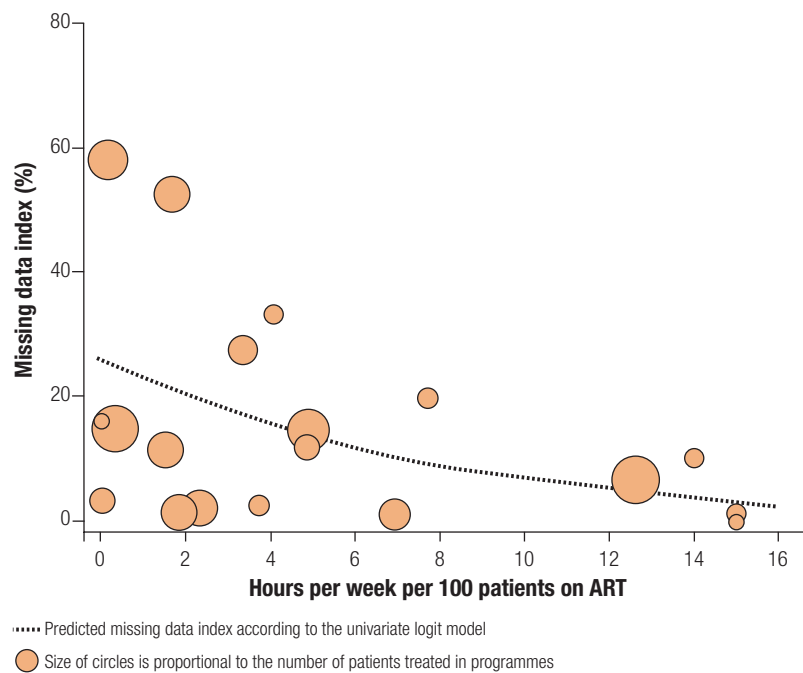
Minimizing losses to follow-up

The quality of both the health care provided and the data collected to evaluate it depends on a complete follow-up of patients and their retention in care. Loss to follow-up at 1 year was substantial, in line with the results of other studies.^{22,23} Most sites reported using some method for tracing patients, but many relied on only one approach, such as outreach teams or collaborations with community-based organizations. The concurrent use of several strategies appears to be most effective in reducing losses to follow-up. Fewer missing data correlated with fewer losses to follow-up, suggesting that better databases might contribute to retaining patients in programmes. Alternatively, limited resources may lead to both poor follow-up and poor data quality. Indeed, it was more difficult to achieve good data quality and to retain patients in more deprived settings. Previous analyses of the ART-LINC data showed that fees for services are associated with higher losses to follow-up.³

Strengths and weaknesses

Fraser et al. recently discussed five HIV treatment programmes using EMR systems to reduce loss to follow-up in

Fig. 1. Missing data index (median of percentage of data missing in six key variables) and hours spent by data clerks on the database each week



Haiti, Kenya, Malawi and Zambia.⁷ Three of the systems reviewed were also included in our survey. In Haiti, Partners In Health, which is Zanmi Lasante's flagship programme, implemented a web-based medical record system in several HIV treatment sites on the Central Plateau.²⁴ Automated e-mail alerts are used to promote timely initiation of treatment.⁷ The system is web-based, with the central server based in the United States and sites connected via satellite links. Data can be entered off line when the connection to the network is lost. The system has advantages over local servers in terms of stability and security but raises issues of data ownership. In Zambia, automated reports generated by the EMR systems supported the tracing of patients lost to follow-up by community health-care workers.²²

Our survey was conducted in 21 different clinics and cohorts providing ART to over 50 000 persons living on three continents. We could directly examine to what extent EMR systems and other programme characteristics correlated with data quality and rates of loss to follow-up. However, participating sites were heterogeneous and represented a convenience sample. The generalizability of our results is therefore uncertain, but clearly there are important problems in data col-

lection, data management and patient follow-up.

Implications for clinical care and operational research

The collection of incomplete and inaccurate data hampers the provision of high-quality care, the monitoring of patients over time, and their retention in treatment programmes. Failure to retain a high proportion of patients in care negates much of the potential benefit of ART treatment programme, since most patients start ART with advanced disease and are likely to die within weeks or months if therapy is discontinued. Although the scale-up of ART in these settings has been a formidable achievement and many deaths have been prevented, these issues need to be addressed to ensure and document the long-term success of ART in these settings.

Missing data are common in operational research. A popular approach is to restrict analyses to individuals with complete data. Although such "complete-case" analyses are often unbiased, they may be biased if patients with missing data are not typical of the whole sample. They are also inefficient because of the reduced sample size.²⁵ In the context of ART programme research, the inclusion of all patients is important; loss to follow-up is likely to

result in underestimated mortality.³ Although multiple imputation of missing values is increasingly used to overcome the problem of missing data,²⁶ the assumptions made in such analyses are difficult to verify.²⁷

Conclusions

Our study suggests that promoting appropriate and sustainable databases and systems to trace patients should be a priority in the context of scaling up ART. Patients could benefit both directly and indirectly from improved data quality, since accurate clinical data are a prerequisite for high standards of care and monitoring, which in turn supports patient retention in the programme. These issues may not have received sufficient attention from the governmental and nongovernmental organizations driving the scale-up of ART in resource limited settings. ■

Acknowledgements

We are grateful to WHO and CDC for supporting the Entebbe workshop and to all those who attended: Akum Aveika (Gambia); Abdoulaye Kalle (Mali); Alain Azondekon (Benin); Aluonzi Bosco (Uganda); Santhanam Anand (India); Andrew Mugisha (Uganda); Catherine Orrell (South Africa); Claudio Faulhaber (Brazil); Cyrille Franck Soppi (Côte d'Ivoire); Tim Meade (Zambia); Dembele Issiaka (Mali); Jean Marie Ntibigarura (Burundi); Jules Bashi Bagendabanga (Benin); Labake Akintunde (Nigeria); Margaret Pascoe (Zimbabwe); John Lelei (Kenya); Daouda Minta Kassoum (Mali); Monica Katyal (USA); Nicholas Musinguzi (Uganda); Tanakorn Apornpong (Thailand) and Karina Visser (South Africa).

Table 4. Probability of missing data for key variables and loss to follow-up according to characteristics of ART programmes in a 15-country study correlating medical record-keeping practices with losses to follow-up

	Number of sites	Crude OR (95% CI)	Adjusted OR (95% CI)
Missing data			
Training of data management staff			
No	5	1	1
Yes	14	0.61 (0.38–0.96)	0.58 (0.37–0.90)
Percentage of population below US\$ 1 a day			
Per 10% increase	19	1.19 (0.91–1.55)	1.25 (1.08–1.45)
Weekly clerk-hours per 100 ART patients			
Per 1 hour increase	19	0.86 (0.75–0.98)	0.95 (0.90–0.99)
Weekly medical staff hours per 100 ART patients			
Per 1 hour increase	19	0.91 (0.71–1.16)	
Loss to follow-up			
Active patient tracing strategies			
None (reference)	3	1	1
1	8	0.55 (0.31–0.95)	0.51 (0.25–1.01)
2	5	0.29 (0.15–0.56)	0.31 (0.15–0.61)
3	2	0.17 (0.14–0.20)	0.17 (0.15–0.20)
Percentage of population below US\$ 1 a day			
Per 10% increase	18	1.11 (0.95–1.31)	1.04 (0.87–1.25)

ART, antiretroviral therapy; CI, confidence interval; OR, odds ratio.

Results from univariate and multivariate (adjusted for all variables listed) logit models with outcomes missing data index and proportion lost to follow-up. Analyses of missing data were based on 41 936 patients from 19 sites and analyses of loss to follow-up on 36 149 patients from 18 sites participating in the ART-LINC collaboration.

We are also grateful to Mauro Schechter, Eduardo Sprinz and François Dabis for their helpful comments on an earlier draft of this paper.

Funding: The ART-LINC collaboration of the International Epidemiological Databases to Evaluate AIDS (IeDEA)

is funded by the United States National Institutes of Health (Office of AIDS Research and National Institute of Allergy and Infectious Diseases) and the French Agence nationale de recherches sur le sida et les hépatites virales.

Competing interests: None declared.

Résumé

Systèmes d'enregistrement électronique des données médicales, qualité des données et nombre de perdus de vue : enquête sur les programmes de traitement antirétroviral dans les pays à ressources limitées

Objectif Décrire les bases de données médicales électroniques utilisées par les programmes de traitement antirétroviral (ART) dans les pays à faible revenu et évaluer les mesures que ces programmes appliquent pour préserver et améliorer la qualité des données et réduire le nombre de patients perdus de vue.

Méthodes Une étude a été menée de décembre 2006 à février 2007 dans 15 pays d'Afrique, d'Amérique du Sud et d'Asie sur l'utilisation des systèmes d'enregistrement électronique des données médicales dans le cadre des programmes ART. Les patients recrutés sur les sites au moment de l'enquête qui n'avaient pas été vus pendant les 12 derniers mois ont été

considérés comme perdus de vue. On a évalué la qualité des données en calculant le pourcentage de variables clés manquantes (âge, sexe, stade clinique de l'infection à VIH, numération des lymphocytes CD4+ et année de mise en route du traitement ART). On a analysé les associations entre les caractéristiques des sites (telles que le nombre de membres du personnel affectés à la gestion des données) et les mesures pour réduire le nombre de perdus de vue (telles que la présence de personnel chargé de retrouver les patients) d'une part et la qualité des données et le nombre de perdus de vue, d'autre part, à l'aide de modèles logit multivariés.

Résultats Vingt-une sites, fournissant globalement un traitement ART à 50 060 patients, ont participé à l'étude (nombre médian de patients par site : 1000 ; intervalle interquartile, IIQ : 72 - 19320). Dix-huit de ces sites (86 %) ont utilisé une base de données électronique pour l'enregistrement des données médicales et 15 (83 %) d'entre eux ont fait appel à des logiciels destinés aux petites entreprises ou à un usage individuel. Le pourcentage médian de données manquantes pour les variables clés était de 10,9 % (IIQ : 2,0 - 18,9 %) et diminuait avec le degré de formation à la gestion des données (odds ratio, OR : 0,58 ; intervalle de confiance à 95 %, IC : 0,37-0,90) et avec le nombre d'heures par semaine consacrées par des employés à la base de données pour 100 patients sous traitement ART (OR : 0,95 ; IC à 95 % : 0,90-0,99). Il fallait environ 10 heures par semaine pour 100 patients sous traitement ART pour amener le pourcentage de données manquantes concernant

les variables clés à moins de 10 %. Le pourcentage médian de patients perdus de vue 1 an après la mise en route du traitement ART était de 8,5 % (IIQ : 4,2-19,7 %). Parmi les stratégies employées pour réduire le nombre de perdus de vue, figuraient le recours à des équipes de proximité et à des organisations communautaires, ainsi que le contrôle des données d'enregistrement des décès. La mise en œuvre de ces trois stratégies a permis de diminuer substantiellement les nombres de perdus de vue (OR : 0,17 ; IC à 95 % : 0,15-0,20).

Conclusion La qualité des données recueillies et la rétention des patients dans les programmes de traitement ART étaient insatisfaisantes pour de nombreux sites participant à l'extension du traitement ART dans les pays à ressources limitées, en raison principalement de la formation insuffisante du personnel à la gestion des données et au traçage des patients perdus de vue.

Resumen

Sistemas de historias clínicas electrónicas, calidad de los datos y pérdidas de seguimiento: encuesta sobre los programas de tratamiento antirretroviral en entornos con recursos limitados

Objetivo Describir las bases de datos clínicas electrónicas utilizadas en los programas de tratamiento antirretroviral (TAR) en los países de ingresos bajos y evaluar las medidas empleadas en esos programas para mantener y mejorar la calidad de los datos y reducir la pérdida de pacientes durante el seguimiento.

Métodos Entre diciembre de 2006 y febrero de 2007 se llevó a cabo en 15 países de África, América del Sur y Asia una encuesta sobre la utilización de sistemas electrónicos de historias clínicas en los programas de TAR. Se consideraron perdidos para el seguimiento los pacientes participantes en los centros estudiados a los que no se había vuelto a ver en los doce meses anteriores. La calidad de los datos se evaluó calculando el porcentaje de variables clave omitidas (edad, sexo, fase clínica de la infección por VIH, recuento de linfocitos CD4+ y año de inicio del TAR). Las relaciones entre las características del centro estudiado (como el número de trabajadores dedicados a la gestión de datos), las medidas tendientes a reducir las pérdidas de seguimiento (por ejemplo la existencia de personal dedicado a localizar a los pacientes) y la calidad de los datos y las pérdidas de seguimiento fueron analizadas mediante modelos logit de varias variables.

Resultados El estudio abarcó 21 centros que suministraron TAR a 50 060 pacientes en total (mediana de pacientes por centro: 1000; intervalo intercuartilico, IIC: 72–19 320). Dieciocho centros (86%) usaban bases de datos electrónicas para sus historias clínicas; 15 (83%) de esos centros utilizaban para ello software concebido para

uso personal o de pequeñas empresas. La mediana del porcentaje de datos omitidos sobre variables fundamentales por centro fue de 10,9% (IIC: 2,0–18,9%) y disminuía con la capacitación para la gestión de datos (razón de posibilidades, OR: 0,58; intervalo de confianza del 95%, IC95%: 0,37–0,90) y las horas semanales dedicadas por un auxiliar administrativo a la base de datos por 100 pacientes sometidos a TAR (OR: 0,95; IC95%: 0,90–0,99). Se requerían unas 10 horas semanales por 100 pacientes sometidos a TAR para reducir los datos faltantes para variables clave a niveles inferiores al 10%. La mediana del porcentaje de pacientes perdidos para el seguimiento al cabo de un año de iniciado el TAR fue de 8,5% (IIC: 4,2–19,7%). Entre las estrategias aplicadas para reducir las pérdidas de seguimiento cabe citar los equipos de extensión, las organizaciones comunitarias y la verificación de los datos de los registros de defunción. La aplicación de esas tres estrategias permitió reducir sustancialmente las pérdidas durante el seguimiento (OR: 0,17; IC95%: 0,15–0,20).

Conclusión La calidad de los datos reunidos y la retención de pacientes en los programas de TAR son insatisfactorias en muchos de los centros que participan en la expansión del TAR en entornos con recursos limitados, principalmente como consecuencia de la falta de personal suficiente con la preparación necesaria para gestionar los datos y localizar a los pacientes perdidos para el seguimiento.

ملخص

نظم السجلات الطبية الإلكترونية، وجودة البيانات وفقدان المتابعة: مسح لبرامج المعالجة بمضادات الفيروسات القهقرية في المواقع المحدودة الموارد

خلال الأشهر الاثني عشرة السابقة للمسح في عداد المفقودين من المتابعة. وتم تقييم جودة البيانات بحساب النسبة المئوية للمتغيرات الرئيسية المفقودة (العمر، الجنس، المرحلة السريرية (الإكلينيكية) للعدوى بفيروس الإيدز، وعدد اللمفاويات C44+ وسنة البدء بالمعالجة بمضادات الفيروسات القهقرية. وتم تحليل الترابط بين خصائص المواقع (مثل عدد العاملين المتخصصين بإدارة البيانات)، وإجراءات تقليل المفقودين من المتابعة (مثل وجود عامل متخصص باقتفاء المرضى) مع جودة البيانات وفقدان المتابعة وذلك باستخدام نماذج لوغاريتمية متعددة المتغيرات.

الهدف: وصف قواعد البيانات الإلكترونية الطبية التي تستخدم في برامج المعالجة بمضادات الفيروسات القهقرية في البلدان المنخفضة الدخل وتقييم الإجراءات التي تتبناها هذه البرامج لصيانة البيانات وتحسين جودتها وخفض فقدان متابعة المرضى.

الطريقة: أجري مسح حول استخدام نظم السجلات الطبية الإلكترونية في برامج المعالجة بمضادات الفيروسات القهقرية في 15 بلداً من أفريقيا وآسيا وأمريكا الجنوبية في الفترة بين كانون الأول/ديسمبر 2006 وشباط/فبراير 2007. وقد اعتبر المرضى الذين أدرجوا في المواقع وقت المسح ولم يشاهدوا

حول المتغيرات الرئيسية ما دون 10%. وبلغت النسبة المئوية الوسطية للمرضى الذين فقدوا في المتابعة بعد مرور سنة من البدء بالمعالجة بالأدوية المضادة للفيروسات القهقرية 8.5% (مدى الشريحة الربعية 4.2 – 19.7%). وتشمل استراتيجيات تقليل المفقود من المتابعة وجود فرق إيصال، ومنظمات مجتمعية المرتكز، والتحقق من البيانات في سجلات الموتى. وقد أدى تنفيذ جميع هذه الاستراتيجيات الثلاث إلى نقص ملحوظ في المفقودين من المتابعة (نسبة الأرجحية 0.17، وفاصلة ثقة 95% وتراوح نسبة الأرجحية بين 0.15 و 0.20).

الاستنتاج: إن جودة البيانات المجموعة واستبقاء المرضى في برامج المعالجة بالأدوية المضادة للفيروسات القهقرية لا تبعث الرضى في النفوس في الكثير من المواقع التي تشارك في الارتقاء بالمعالجة بالأدوية المضادة للفيروسات القهقرية في البلدان المحدودة الموارد، ويعود ذلك بشكل رئيسي إلى نقص العاملين المدربين على إدارة البيانات واقتفاء المرضى الذين فقدوا من المتابعة.

الموجودات: شملت الدراسة 21 موقعاً تقدّم مجموعها المعالجة بمضادات الفيروسات القهقرية إلى 50060 مريضاً (العدد الوسطي للمرضى في كل موقع 1000 ومدى الشريحة الربعية يتراوح بين 72 – 19 320). وكان 18 موقعاً (86%) يستخدم قاعدة بيانات إلكترونية لحفظ السجلات الطبية، فيما كان 15 موقعاً (83%) تعتمد على برمجيات تستهدف الاستخدام الشخصي أو الأعمال الصغيرة. وقد كان وسطى النسبة المئوية للمعطيات المفقودة للمتغيرات الرئيسية لكل موقع 10.9 (مدى الشريحة الربعية 2.0 و 18.9%)؛ ونقصت مع التدريب على إدارة البيانات (نسبة الأرجحية 0.58، وفاصلة ثقة 95%، إذ تراوحت نسبة الأرجحية 0.37 – 0.90)، وعدد الساعات التي قضاها احد الكتيبة في مد قاعدة البيانات بالمعلومات عن كل مئة مريض كل أسبوع حول المعالجة بالأدوية المضادة للفيروسات القهقرية (نسبة الأرجحية 0.95، وفاصلة ثقة 95% إذ تراوحت نسبة الأرجحية 0.90-0.99). وقد تطلب الأمر ما يقرب من 10 ساعات كل أسبوع لكل مئة مريض يتناول المعالجة بالأدوية المضادة للفيروسات القهقرية لتقليل البيانات المفقودة

References

- World Health Organization. *Towards universal access: scaling up priority HIV/AIDS interventions in the health sector: progress report 2008*. Available from: <http://www.who.int/hiv/pub.pdf> [accessed on 22 October 2008].
- World Health Organization. *Treating 3 million by 2005: making it happen: the WHO strategy: the WHO and UNAIDS global initiative to provide antiretroviral therapy to 3 million people with HIV/AIDS in developing countries by the end of 2005*. Available from: <http://www.who.int/3by5/publications.pdf> [accessed 25 April 2008].
- Brinkhof MW, Dabis F, Myer L, Bangsberg DR, Boule A, Nash D, et al. Early loss to program in HIV-infected patients starting potent antiretroviral therapy in lower-income countries. *Bull World Health Organ* 2008;86:559-67. PMID:18670668
- Handford CD, Tynan AM, Rackal JM, Glazier RH. Setting and organization of care for persons living with HIV/AIDS. *Cochrane Database of Systematic Reviews* 2006; issue 3. Available from: <http://mrw.interscience.wiley.com/cochrane/clsysrev/articles/CD004348/frame.html> [accessed on 25 April 2008].
- Rotich JK, Hannan TJ, Smith FE, Bii J, Odero WW, Vu N, et al. Installing and implementing a computer-based patient record system in sub-Saharan Africa: the Mosoriot Medical Record System. *J Am Med Inform Assoc* 2003; 10:295-303. PMID:12668697 doi:10.1197/jamia.M1301
- Simbini T. Computerised information management systems in HIV/AIDS care and outcomes research. *Cent Afr J Med* 2006; 52(5-6):65-7. PMID:18254459
- Fraser HS, Allen C, Bailey C, Douglas G, Shin S, Blaya J. Information systems for patient follow-up and chronic management of HIV and tuberculosis: a life-saving technology in resource-poor areas. *J Med Internet Res* 2007;9:e29. PMID:17951213 doi:10.2196/jmir.9.4.e29
- Dabis F, Balestre E, Braitstein P, Miotti P, Brinkhof WG, Schneider M, et al. Cohort profile: Antiretroviral Therapy in Lower Income Countries (ART-LINC): international collaboration of treatment cohorts. *Int J Epidemiol* 2005; 34:979-86. PMID:16157617 doi:10.1093/ije/dyi164
- Braitstein P, Brinkhof MW, Dabis F, Schechter M, Boule A, Miotti P, et al. Antiretroviral Therapy in Lower Income Countries (ART-LINC) collaboration. ART Cohort Collaboration (ART-CC) groups. Mortality of HIV-1-infected patients in the first year of antiretroviral therapy: comparison between low-income and high-income countries. *Lancet* 2006;367:817-24. PMID:16530575 doi:10.1016/S0140-6736(06)68337-2
- ART-LINC Collaboration of International Databases to Evaluate AIDS (leDEA): Keiser O, Anastos K, Schechter M, Balestre E, Myer L, Boule A, et al. Antiretroviral therapy in resource-limited settings 1996 to 2006: patient characteristics, treatment regimens and monitoring in sub-Saharan Africa, Asia and Latin America. *Trop Med Int Health* 2008; 13(7); 870-9 [Epub ahead of print].
- DataCol (Data Collector). Available from: <http://www.who.int/datacol/home.asp> [accessed on 25 April 2008].
- The state of the world's children 2006*. New York: United Nations Children's Fund; 2005.
- Delaney K. *Inside Microsoft® SQL Server™ 2005: the storage engine*. Microsoft Press; 2006.
- MySQL. Available from: <http://www.mysql.org> [accessed 25 April 2008].
- Lowrance D, Filler S, Makombe S, Harries A, Aberle-Grasse J, Hochgesang M, et al. Assessment of a national monitoring and evaluation system for rapid expansion of antiretroviral treatment in Malawi. *Trop Med Int Health* 2007;12:377-81. PMID:17313509
- OpenMRS (Open Medical Record System). Available from: <http://openmrs.org> [accessed on 25 April 2007].
- Mamlin BW, Biondich PG, Wolfe BA, Fraser H, Jazayeri D, Allen C, et al. Cooking up an open source EMR for developing countries: OpenMRS - a recipe for successful collaboration. *AMIA Annu Symp Proc* 2006;2006:529-33. PMID:17238397
- Allen C, Jazayeri D, Miranda J, Biondich PG, Mamlin BW, Wolfe BA, et al. Experience in implementing the OpenMRS medical record system to support HIV treatment in Rwanda. *Medinfo* 2007;12:382-6. In: Kuhn K, Warren JR, Leong TY, eds. *Medinfo 2007: proceedings of the 12th World Congress on Health (Medical) Informatics; building sustainable health systems*. Amsterdam: IOS Press, Inc.; 2007. pp. 382-6.
- Tierney WM, Rotich JK, Hannan TJ, Siika AM, Biondich PG, Mamlin BW, et al. The AMPATH medical record system: creating, implementing, and sustaining an electronic medical record system to support HIV/AIDS care in western Kenya. In: Kuhn K, Warren JR, Leong TY, eds. *Medinfo 2007: proceedings of the 12th World Congress on Health (Medical) Informatics; building sustainable health systems*. Amsterdam: IOS Press, Inc.; 2007. pp. 372-6.
- Fraser HS, Biondich P, Moodley D, Choi S, Mamlin BW, Szolovits P. Implementing electronic medical record systems in developing countries. *Inform Prim Care* 2005;13:83-95. PMID:15992493
- Chandrasekhar CP, Ghosh J. Information and communication technologies and health in low income countries: the potential and the constraints. *Bull World Health Organ* 2001;79:850-5. PMID:11584733
- Stringer JS, Zulu I, Levy J, Stringer EM, Mwango A, Chi BH, et al. Rapid scale-up of antiretroviral therapy at primary care sites in Zambia: feasibility and early outcomes. *JAMA* 2006;296:782-93. PMID:16905784 doi:10.1001/jama.296.7.782
- Rosen S, Fox MP, Gill CJ. Patient retention in antiretroviral therapy programs in sub-Saharan Africa: A systematic review. *PLoS Med* 2007;4:e298. PMID:17941716 doi:10.1371/journal.pmed.0040298
- Fraser HS, Jazayeri D, Nevil P, Karacaoglu Y, Farmer PE, Lyon E, et al. An information system and medical record to support HIV treatment in rural Haiti. *BMJ* 2004;329:1142-6. PMID:15539669 doi:10.1136/bmj.329.7475.1142
- Little RJ, Rubin DB. *A taxonomy of missing-data methods: statistical analysis with missing data*. New York: Wiley, 2002. pp. 19-23.
- Barnard J, Meng XL. Applications of multiple imputation in medical studies: from AIDS to NHANES. *Stat Methods Med Res* 1999;8:17-36. PMID:10347858 doi:10.1191/096228099666230705
- Vandenbroucke JP, von Elm E, Altman DG, Gotsche PC, Mulrow CD, Pocock SJ, et al. Strengthening the reporting of observational studies in epidemiology (STROBE): explanation and elaboration. *PLoS Med* 2007;4:e297. PMID:17941715 doi:10.1371/journal.pmed.0040297