The CCASAnet Data Audit Process

Stephany Duda MS, Catherine McGowan MD, Firas Wehbe MD, MS, Daniel Masys MD

The IeDEA Caribbean, Central, and South America Network for HIV Research (CCASAnet) has instituted a process of periodic data monitoring visits at its member sites, using audit principles adapted from the International Congress on Harmonization "Good Clinical Practice" guidelines [http://www.fda.gov/oc/gcp/guidance.html]. The CCASAnet Data Coordinating Center hosted at Vanderbilt University conducts project-driven audits when new data sets are submitted for proposed region-wide studies. Such audits help CCASAnet participants to identify sources of error in data collection, abstraction, and representation, and help the Vanderbilt Data Coordinating Center (VDCC) to determine the structure, quality, and reliability of the submitted data. The VDCC also conducts a second round of supportive quality improvement-focused audits at sites where significant challenges to data capture and quality have been identified.

For an audit, a team from the VDCC visits each of the CCASAnet participating sites and compares the contents of the electronic database the site has submitted to the VDCC to local source documents. The source documents available at CCASAnet sites are mainly paper clinical records, though some sites maintain electronic laboratory, pharmacy, and patient medical record systems. During the visit, the audit team consults as many data sources as the sites can make available.

The VDCC's current audit team consists of a physician specializing in HIV care and an informatics graduate student proficient in data management. The team usually begins a site audit on Monday morning and ends the visit on Wednesday afternoon with an exit interview with the site investigators.

Before the audit

The audit cycle begins when a CCASAnet site submits data to the Vanderbilt Coordinating Center for a region-wide project. The VDCC reviews the submitted data and checks for missing or outlying values, poor formatting, and logical inconsistencies. Record errors that are commonly encountered include: inconsistent dates (e.g. lab values occurring after death, regimen dates occurring after date of last visit), incomplete or malformed dates (e.g. missing month or day of month, or ambiguous date format such as xx/xx/yyyy), or lab values that are clearly out of range. The VDCC data manager communicates on an ongoing basis with the site's data manager to resolve any outstanding issues with the records. Corrections made to the data at this stage are not included in the data set used for auditing. The VDCC data manager randomly selects thirty to forty records for audit, and identifies an additional five records that have unresolved data inconsistencies. The research identification numbers (IDs) of the majority of randomly selected records are sent to the site at least one week prior to the data monitoring visit, so that site data personnel can pull the requested records in advance. The remaining randomly selected records plus the five "targeted" records are

given to the site investigators on the first audit day. These "day-of" record requests help to ensure no records are deliberately altered in advance of the data monitoring visit.

The VDCC data manager prepares audit forms for all selected records by merging all the available data for the specified patients (Appendix, Table 1). Data on the form are divided into categories such as demographics, clinical visit data, antiretroviral regimens, and laboratory results. The fields in the "from database" column of the audit form are automatically populated with values from the database. The audit team uses the corresponding "from source documents" column to record whether matching values are found in the site's source documentation. These pre-filled audit forms are checked manually, then printed and stapled.

At the audit site, data personnel pull the requested records – making note of any that are missing or unavailable – and print a look-up key that matches the de-identified research IDs stored in the database to the actual patient record numbers. The site staff also prepares a quiet space where the audit team can work uninterrupted.

During the audit

In the morning of the first visit day, the audit team meets with site investigators to review the audit process and answer questions posed by the local investigators. The audit team leader presents a list of the "day-of" and "targeted" audit record IDs and local staff arrange to have these records pulled immediately.

The team discusses local data collection and abstraction procedures with site investigators and requests copies of any data entry forms in use. Local clinical personnel present a sample patient record and explain the different components of the chart (flow sheets, laboratory reports, clinic intake and visit notes, pharmacy dispensing records, hospitalization records, death certificates, etc.) as well as any local coding schemes or shorthand commonly found in the clinicians' notes. The team schedules an exit interview with the site primary investigator (PI). If this is the first site visit, the team also takes a tour of the facilities.

The team from the VDCC then conducts the audit, making note of content in the database that does not exist in the clinical record, values that do not match between the database and the clinical record, and significant content in the clinical record that was not entered in the database. Audit results are recorded on the paper audit forms. The audit team also checks for well-maintained records, signed and dated consent-for-treatment forms when applicable, and erased or improperly edited content in the clinical chart. If other source documents are available, such as electronic pharmacy or laboratory systems, the audit team verifies database content using these sources also. The audit team generally works in private, but may consult with local personnel for clarifications.

On the final audit day, the audit team meets with the site PI and describes the preliminary findings of the data monitoring visit. The group discusses the site's data quality and the strengths and weaknesses of the current data collection approach. The audit team then

presents an initial set of recommendations and discusses their feasibility with the local PI and data personnel.

After the audit

After the audit team returns to the VDCC, team members review the audit records, classify data errors, omissions, and inconsistencies, and tabulate the results by data category (Appendix, Table 2). They also compile comparison charts for individual records that present side-by-side the database and source document content, so local data personnel can see specific discrepancies (Appendix, Table 3).

The members of the audit team discuss their findings with the VDCC data manager and biostatistician and the CCASAnet PI, and agree on several general recommendations to make to the site. The audit team then composes a data audit report that contains the error tables and comparison charts and describes in detail the audit team's findings and recommendations. The draft report is sent to the site PI for comment before it is finalized and the CCASAnet team solicits feedback from the sites about the data audit process.

The VDCC consults with site data personnel to adapt and implement the recommendations detailed in the audit report.

For further details, please contact the CCASAnet PI, Daniel Masys, MD (dan.masys@vanderbilt.edu), the CCASAnet data manager, Firas Wehbe, MD, MS (firas.wehbe@vanderbilt.edu) or the members of the audit team: Catherine McGowan, MD (cathy.mcgowan@vanderbilt.edu) and Stephany Duda, MS (stephany.duda@vanderbilt.edu).

Appendix

Table 1. Example of audit form drafted from database prior to data audit

ID: 52LN9
Demographics

Variable	Value	Audit Value	Notes
Gender	M		
Birthdate	1973-01-31		
Diagnosis date	2000-04-23		
Diagnosis place	123 Clinic		
Enrollment date	2004-08-23		
Risk MSM	Yes		
Risk heterosexual	No		
Risk IDU	Yes		
Risk vertical	No		
Risk blood	No		
Risk occupational	No		
Height	173		
Weight	64		
HAART naïve	Yes		
Education level	2		
Housing type	4		
Work status	2		

Clinical Data

Variable	Value	Audit Value	Notes
Date	2005-09-15		
Weight	58		
CDC stage	C3		
Date	2006-10-02		
Weight	64		
CDC stage	A1		
Lost to Follow-up / Death			
Last visit	2008-01-09		
Lost	No		
Death date			
Death place			

ARV-related Data

Variable	Value	Audit Value	Notes
Regimen	3TC/AZT/SAQ/RIT		
Start date	2004-08-23		
Stop date	2005-09-14		
Stop reason	Virologic Failure		
Regimen	D4T/ABC/EFV		
Start date	2005-09-15		
Stop date			
Stop reason			

Labs

Variable	Value	Audit Value	Notes
CD4			
CD4 date	2004-08-23		
CD4 abs. value	122		
CD4 percent	5		
CD4 date	2005-09-14		
CD4 abs. value	41		
CD4 percent	2		
Viral Load			
VL date	2004-08-23		
Copies	170803		
Log	5.23		
VL date	2005-12-11		
Copies	<50		
Log	<1.7		
Other Labs			
Test			
Test date			
Test result			

Table 2. Sample table of data audit findings of laboratory values at ABC Clinic

Data Type	Variables Reviewed	CR-DB Mismatches	Unverifiable DB Values	Total	Rate
Hepatitis					
Hep C test	25	3	4	7	**0.28**
Hep C test result	23	2	3	5	**0.22**
Hep C test date	25	3	0	3	**0.12**
CD4					
Date	102	8	2	10	0.10
Value	103	2	0	2	0.02
Percentage	99	2	1	3	0.03
Viral load					
Date	115	2	15	17	**0.15**
Copies	120	1	3	4	0.03
Log	118	3	7	10	0.08
Test method	32	1	6	7	**0.22**
Total	762	27	41	68	0.09

Table 3. Sample table detailing comparison of ARV regimen documentation in database with clinical records

#1830	Value in database	Value in clinical record	Comments
Regimen 1			
Regimen	COM NFV	COM NFV	
Start date	19 Apr 2000	19 Apr 2000	
Stop date	26 Jun 2002	26 Jun 2002	
Regimen 2			
Regimen	D4T DDI IDV RIT	D4T DDI IDV RIT	
Start date	25 Sept 2001	25 Sept 2001	
Stop date	28 Aug 2002	28 Aug 2002	
Regimen 3			
Regimen	AZT DDI RIT SAQ	AZT DDI RIT SAQ	
Start date	28 Aug 2002	28 Aug 2002	
Stop date	09 May 2006	06 Feb 2005	Notes from the I.D. clinic show that regimen 3 was stopped on 06 Feb 2005. The TB clinic continues to report regimen 3 active until 09 May 2006.

Regimen 4			
Regimen	---	AZT 3TC EFV	
Start date	---	06 Feb 2005	(documented in hospital orders and notes from I.D. clinic)
Stop date	---	01 Jan 2006	
Regimen 5			
Regimen	---	D4T 3TC EFV KAL	
Start date	---	14 Feb 2006	Regimen 5 began during hospitalization and is listed as dispensed in hospital orders
Stop date	---	21 Feb 2006	Regimen 5 stops appearing in hospital orders.
Regimen 6			
Regimen	D4T 3TC KAL	D4T 3TC KAL	
Start date	09 May 2006	22 Feb 2006	Regimen 6 was begun on 22 Feb 2006 during hospitalization and (according to the order sheets) the drugs were administered.
Stop date	19 Sept 2006	23 Jul 2006	I.D. notes show Regimen 6 was stopped on 23 Jul 2006.
Regimen 7			
Regimen	3TC AZT KAL	AZT 3TC KAL	
Start date	19 Sep 2006	19 Sep 2006	
Stop date	---	---	