HHS Public Access

Author manuscript

J Biomed Inform. Author manuscript; available in PMC 2016 August 16.

Published in final edited form as:

J Biomed Inform. 2015 August ; 56: 387–394. doi:10.1016/j.jbi.2015.07.009.

A structured approach to recording AIDS-defining illnesses in Kenya: A SNOMED CT based solution

Tom Oluoch[a,*], **Nicolette de Keizer**[b], **Patrick Langat**[c], **Irene Alaska**[c], **Kenneth Ochieng**[c], **Nicky Okeyo**[c], **Daniel Kwaro**[c], and **Ronald Cornet**[b,d]

[a]U.S. Centers for Disease Control and Prevention, Division of Global HIV/AIDS, Nairobi, Kenya [b]Department of Medical Informatics, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands [c]Kenya Medical Research Institute – CDC Collaborative Program, Kisumu, Kenya [d]Department of Biomedical Engineering, Linköping University, Linköping, Sweden

Abstract

Introduction—Several studies conducted in sub-Saharan Africa (SSA) have shown that routine clinical data in HIV clinics often have errors. Lack of structured and coded documentation of diagnosis of AIDS defining illnesses (ADIs) can compromise data quality and decisions made on clinical care.

Methods—We used a structured framework to derive a reference set of concepts and terms used to describe ADIs. The four sources used were: (i) CDC/Accenture list of opportunistic infections, (ii) SNOMED Clinical Terms (SNOMED CT), (iii) Focus Group Discussion (FGD) among clinicians and nurses attending to patients at a referral provincial hospital in western Kenya, and (iv) chart abstraction from the Maternal Child Health (MCH) and HIV clinics at the same hospital. Using the January 2014 release of SNOMED CT, concepts were retrieved that matched terms abstracted from approach iii & iv, and the content coverage assessed. Post-coordination matching was applied when needed.

Results—The final reference set had 1054 unique ADI concepts which were described by 1860 unique terms. Content coverage of SNOMED CT was high (99.9% with pre-coordinated concepts;

*Corresponding author at: U.S. Centers for Disease Control and Prevention, Division of Global HIV/AIDS, P.O. Box 606, 00621, Village Market, Nairobi, Kenya. Tel.: +254 20 286 7000; fax: +254 20 271 4745. toluoch@cdc.gov, oluoch.tom@gmail.com (T. Oluoch).

Disclaimer

The findings and conclusions in this report are those of the author(s) and do not necessarily represent the official position of the Centers for Disease Control and Prevention/the Agency for Toxic Substances and Disease Registry, or the Kenya Medical Research Institute.

Co-author contribution

Tom Oluoch, Nicolette de Keizer and Ronald Cornet conceptualized and designed the study. Irene Alaska and Daniel Kwaro contributed to the clinical components of the study and manuscript review. Patrick Langat and Nicky Okeyo provided oversight to the data management team that conducted the chart abstraction, FGD and overall data cleaning. Kenneth Ochieng did the programming in OpenMRS, including developing the user interface. All co-authors were involved in interpreting the results. TO, NdK and RC drafted and revised the manuscript. All authors edited and reviewed the manuscript and gave their final approval for submission to the journal.

Conflicts of interest

Authors are members of the Technical Committee (Ronald Cornet) and Quality Assurance Committee (Nicolette de Keizer) of the International Health Terminology Standards Development Organization (IHTSDO), which publishes SNOMED CT. Their positions at the IHTSDO, however, had no bearing on the research study or results.

100% with post-coordination). The resulting reference set for ADIs was implemented as the interface terminology on OpenMRS data entry forms.

Conclusion—Different sources demonstrate complementarity in the collection of concepts and terms for an interface terminology. SNOMED CT provides a high coverage in the domain of ADIs. Further work is needed to evaluate the effect of the interface terminology on data quality and quality of care.

Keywords

AIDS; SNOMED CT; AIDS-related opportunistic infections; Developing countries; Quality of healthcare

1. Introduction

AIDS-Defining opportunistic Illnesses (ADIs), defined as illnesses that occur with more severity and higher frequency among persons with HIV, remain the main course of morbidity and mortality among HIV-infected persons [1,2]. Although several studies show a decline in incidence of individual ADIs such as tuberculosis, Pneumocystis pneumonia and Kaposi sarcoma since the introduction of anti-retroviral therapy (ART) [3,4], mortality associated with ADIs remains high. Sub-Saharan Africa (SSA), which is home to nearly two-thirds of the world's 33.6 million people living with HIV, also has the highest rates of HIV-ADI co-infection [5]. The US Centers for Disease Control and Prevention (CDC) and the World Health Organisation (WHO) have released guidelines for prevention and treatment of ADIs, intended for use by healthcare providers and policy makers [6,7]. WHO classifies ADIs into four clinical stages indicative of HIV disease progression to help clinicians in resource-limited settings (mainly in Africa and Asia) with no immediate access to key laboratory testing capability to make decisions on ART eligibility [8].

Many studies conducted in SSA have shown that routine clinical data often have errors and in several cases missing key data elements such as diagnosis of ADIs [9–11]. Poor recording or incorrect inference of WHO clinical staging based on the ADIs can lead to under or over-treatment of patients. A majority of clinics providing HIV care and treatment in SSA use a fixed list of coarse-grained ADIs to perform clinical staging of HIV-infection as recommended by WHO [8]. Kiragga et al. estimate that ADIs in clinical settings are under-reported by up to 67% in HIV clinics in Uganda [10]. Classification and terminology systems are increasingly used in developed countries for clinical documentation [12]. Lack of a structured, uniquely coded and comprehensive set of ADIs based on a terminology standard, from which diagnoses can be documented to support clinical decisions such as eligibility for antiretroviral therapy can potentially compromise the quality of care [10,13]. Additionally, in busy clinics with over-worked health workers, ADIs are rarely recorded with a fine granularity including more specific disease sub-types. Lack of granular information on diagnosis may influence continuity of treatment given.

Jaramogi Oginga Odinga Teaching and Referral Hospital (JOOTRH) in Kisumu county, western Kenya has two busy HIV clinics where more than 7000 patients receive care and treatment. The hospital does not have a standardized way of documenting signs, symptoms

and disorders during patient visits and a previous data review showed that diagnoses may have been under-reported by up to 60% [14]. This potentially impacts the quality of data recorded, inference of WHO clinical staging as well as continuity of care for patients who transfer from one clinic to another. In order to address this problem, we propose to standardize the recording of ADIs using a standard terminology. A terminology system such as SNOMED Clinical Terms (SNOMED CT) can contribute to data accuracy and re-use through concept-based definition of diagnoses and coded storage of data [15].

This study describes a structured approach to derive a comprehensive set of ADI concepts based on SNOMED CT and evaluates SNOMED CT's content coverage for ADIs in a provincial referral hospital in Kenya. The ADI reference set is derived from various sources based on a framework developed by Bakhshi-Raiez et al. [16], which applies a multi-step approach. The resultant set of concepts will be integrated into Electronic Medical Record (EMR) systems and implemented as an interface terminology in OpenMRS EMR at the JOOTRH providing clinicians with a standardized tool for fine-grained recording of ADIs and automated inference of WHO clinical staging.

2. Background

2.1. WHO clinical staging and ART eligibility

The WHO clinical staging uses clinical presentation of patients to categorize HIV infection into 4 stages based on the severity of ADIs and prognosis. The stages reflect progression of HIV infection from asymptomatic to conditions where presumptive diagnoses can be made on the basis of clinical presentation or simple investigations [8]. Clinical staging is progressive, meaning that when a patient has recovered from a condition defined in a higher stage, they retain the higher clinical stage and can never be classified at a lower stage. Patients with stage I and II HIV infection present with (or have a history of) conditions such as minor skin diseases and upper respiratory tract infections, which are indicative of early HIV infection. WHO clinical stages III and IV, which are indicative of advanced HIV disease, may manifest as one or more illnesses such as severe bacterial infection, extra-pulmonary tuberculosis and non-Hodgkin lymphoma. WHO recommends initiation of ART for patients who have stage III or stage IV conditions [8].

The Kenyan Ministry of Health (MOH) HIV care and treatment guidelines [17], which are adapted from the WHO guidelines, recommend that all HIV-infected pregnant women receive ART, irrespective of their immunological or clinical indication of opportunistic infections associated with HIV [7].

2.2. SNOMED CT

SNOMED CT is an international logic-based, controlled vocabulary used in clinical care that amongst other areas includes terms for diagnostic concepts [18]. SNOMED CT enables meaning-based retrieval of clinical information from electronic health records (EHR) and is able to express finer-grained concepts than classification systems such as ICD-10. For this study, we used the January 2014 release of SNOMED CT which had 298,581 active concepts associated with 781,878 active descriptions. The concepts are interrelated by

896,942 active relationships which can be hierarchical (Is-A relationship) or non-hierarchical (e.g. “associated morphology” or “finding site”). A SNOMED CT reference set provides a mechanism to group concepts and/or descriptions for a common goal (such as diagnosis concepts for AIDS defining illnesses).

Rosenbloom et al. describe an interface terminology as a systematic collection of health-care phrases or terms that support clinicians’ entry of patient-related information into computer programs and decision tools [19]. Interface terminologies also support the display of computer-stored patient information to clinical users in simple human-readable format. The use of a SNOMED CT based terminology on the user interface of an EMR supports efficient selection of terms during data entry based on uniquely coded concepts and promotes data re-use in decision support, statistical analysis and reporting.

The Kenya Medical Research Institute (KEMRI), a collaborator in this study, applied for and obtained a free institutional SNOMED CT license from IHTSDO within the category of developing countries. Under this license, KEMRI can access the SNOMED CT database and use the concepts, descriptions and relationships in EMRs supporting clinical care and research. The license is renewed annually.

2.3. OpenMRS

OpenMRS is a free, collaborative, open source software that supports the delivery of health care and is widely used in developing countries. It is based on a “concept dictionary” that describes data items such as clinical findings and laboratory results. It provides a platform which can be customized through addition of new data items, forms and reports without programming. It is web based but can be deployed on a single standard-alone desktop computer, a laptop or a large server and runs on Linux, Windows or Mac OS X.

3. Methods

3.1. Study setting

The JOOTRH provides ambulatory HIV care and treatment services to over 7000 active patients, of whom about 4800 are receiving ART. The services are provided within the Maternal and Child Health (MCH), the HIV clinic/Patient Support Centre (PSC) and the Integrated Tuberculosis clinic. The three clinics in the hospital use a hybrid system of paper records and EMR to manage patient data; the EMR system is currently under staged deployment and is expected to replace the paper records by the end of 2014. The study was conducted at the HIV/PSC clinic and the MCH clinic which includes the Ante-Natal Care (ANC) clinic.

3.2. Study participants

We extracted free-text data used to describe diagnoses, signs and symptoms from paper-based clinical notes for all HIV-positive patients receiving care and treatment services at the HIV/PSC as well as HIV-positive pregnant women presenting for antenatal care or at the labor and delivery room during the months of February and March 2014, and receiving

Antiretroviral Therapy (ART). We also conducted a Focus Group Discussion (FGD) among clinicians and nurses who attend to patients visiting the MCH and HIV/PSC clinics.

3.3. Review of term-based ADI recording

A trained clinician (clinical officer) extracted terms for diagnoses, signs and symptoms recorded in the patient charts to determine the terms used to document ADIs at the ANC and HIV/PSC. Data were retrieved from the two clinics to ensure comprehensive coverage of terms used to describe ADIs. Although terms for diagnoses, signs and symptoms were retrieved from the charts, we only included in the analysis those that matched SNOMED CT "disorder" concepts as described later.

3.4. Deriving ADI reference set

Based on a framework developed by Bakhshi-Raiez et al. [16], we derived the reference set of ADI concepts through a multi-step approach. The first three steps describe how the ADI reference set was derived as shown in Fig. 1. In the first step, concept-based subsets of ADIs from SNOMED CT and Accenture/CDC (described below) were merged. The *CliniClue Xplore* browser [20] and *NICTIZ terminology explorer* [21] were used to explore the concepts. The second step entailed extraction of term-based concepts from patient charts and Focus Group Discussions. In step 3, the concept-based and term-based concepts were merged.

Step 1: Processing concept-based ADI subsets

a. ***Accenture/CDC subset:*** The US Centers for Disease Control and Prevention (CDC) derived a list of opportunistic infections considered to be AIDS defining illnesses [6]. CDC and Accenture mapped the *diagnosis* and *organism* concepts on to SNOMED CT to come up with a hierarchical set of SNOMED CT-based concepts stored in the Web Ontology Language (OWL) format [22]. The Accenture/CDC subset contains 907 concepts and 1528 associated descriptions for ADIs.

b. ***SNOMED CT subset***: We used structured query language (SQL) commands to select concepts and descriptions associated with ADIs from an instance of the January 2014 release of SNOMED CT which was stored in a MySQL database. The SQL commands were based on the conditions described below:

- Concepts described by terms including "assoc" and ("AIDS" or "HIV").
- Concepts linked by the relationship "*associated with*" to the concept "*HIV infection*".
- All subtypes (descendants) of the above selected concepts.
- The active descriptions of the resulting concepts in SNOMED CT.

The SQL queries used to retrieve the data are in Appendix A. Since all patients seen in the study clinics had HIV, those super-types of concepts were retrieved that did not include the phrase “Associated with AIDS”. For example: *Pneumococcal Pneumonia associated with AIDS* (concept ID = 420787001) was replaced with the parent concept *Pneumococcal Pneumonia* (concept ID = 233607000).

The two ADI subsets described in steps (a) and (b) above were merged into subset A as shown in Fig. 1, sorted by SNOMED CT concept ID and all duplicate concepts and terms deleted.

Step 2: Deriving term-based concepts

a. ***Free text abstraction:***

Two clinical officers reviewed free-text records from 583 initial patient visits and 762 follow-up visits for the same patients who visited the JOOTRH ANC and HIV/PSC clinics in the months of February and March 2014 and abstracted terms used to describe ADIs at the clinic. Due to the large number of patients seen at the clinics, and our interest to discover as many ADIs as possible, the two clinicians shared the data abstraction responsibility with each one reviewing a different set of patient charts. The free-text terms were extracted from paper-based patient records as well as those entered in the OpenMRS EMR. Terms and description IDs from SNOMED CT that matched the free-text terms were retrieved by a medical informatician with SNOMED CT expertise (TO). These concepts and the terms describing them were stored as subset B (Fig. 1).

b. ***Focus Group Discussion:***

In this step, a Focus Group Discussion (FGD) was held involving clinicians who attend to patients at the ANC and HIV/PSC clinics. A separate FGD was conducted among nurses working at the same clinics. All descriptions for the concepts from subset A were made available to the clinicians prior to the start of the FGD. Clinicians’ and nurses’ knowledge was sought on the most common ADIs in their setting (JOOTRH). Using local terms, they listed as many ADIs as they could in addition to the concepts and terms included in subset A. Clinicians and nurses were further asked what factors they consider in determining whether the symptoms and signs are suggestive of an ADI. Terms and description IDs from SNOMED CT that matched terms used for ADI diagnoses in the FGD were retrieved by the same medical informatician (TO).

The terms and concepts derived through abstraction of data from the MCH/PSC records and the FGD were merged into subset B and duplicates removed (see Fig. 1). Concepts were retrieved from SNOMED CT based on the “exact match” and if needed, by post-coordination. The example below shows how we derived a post-coordinated concept *oropharyngeal candidiasis* which is not in SNOMED CT. CDC describes *oropharyngeal*

candidiasis as candidiasis that develops in the throat or mouth [23]. Candidiasis (78048006): finding site (363698007) = Mouth and/or pharynx structures (312533001).

Step 3: Merging the subsets—In the final step, all derived concepts (subsets A and B) and their descriptions were merged into a single subset C and all duplicate concepts and descriptions deleted.

Step 4: SNOMED CT subset extension—If terms could not be matched with SNOMED CT concepts and descriptions, these terms were presented to a team of clinicians and informaticians with expertise in SNOMED CT for review. They determined whether these could be considered as local extensions of SNOMED CT through post-coordination of concepts or as entirely new concepts or descriptions.

Step 5: Constraining the ADI subset—Several non-ADI concepts were deleted from the term-based subsets. Although these non-ADIs were common co-infections that HIV patients attending the clinics at JOOTRH present with, they were not considered for implementation of the interface terminology. Such non-ADI concepts included *malaria* (concept ID = 61462000) and *urinary tract infection* (concept ID = 68566005).

3.5. Assessing content coverage

We matched each of the term-based concepts to SNOMED CT concepts in order to assess content coverage in the latter. Due to the small number of concepts, we did a manual matching and recorded the concepts that could not be found in SNOMED CT.

3.6. Implementation of the interface terminology

The resulting concept subset and description subset from the above steps were represented in the OpenMRS concept dictionary and implemented as the interface terminology on OpenMRS data entry forms to improve ease and accuracy of recording of ADIs. A separate set of non-ADI concepts representing common co-infections or signs, symptoms among patients seeking HIV treatment at JOOTRH was also created from the chart review and FGD for the purpose of presentation on the user interface. Subsetting the concepts into ADI and non-ADI disorders for the purpose of presentation on the user interface provides clinicians quick access to shorter lists of relevant diagnosis concepts to choose from. The ADI concepts were presented on the user interface as a dropdown menu with autocomplete/autosuggest textbox feature derived using a list builder. A ’more details’ button is displayed for each of the ADIs selected for a particular patient. Clicking on this button allows the selection of qualifiers of a particular diagnosis.

The patients’ WHO staging is automatically inferred based on the recorded ADI concept. To accomplish this, the ADI concept is matched with the corresponding WHO stage as defined in the HIV treatment guidelines [8,17]. The WHO clinical staging list showing the ADI concepts and corresponding WHO clinical stage is in Appendix B [8].

3.7. Ethical considerations

The study was approved by the Associate Director for Science at the Center for Global Health of the US Centers for Disease Control and Prevention (CDC) and the Kenya Medical Research Institute (KEMRI) institutional review board. The women whose data were included in the study consented to the use of their data for clinical care and research studies taking place at the ANC clinic. As the study subjects were considered a vulnerable population (HIV-positive pregnant women), the study team ensured that the individual patient data were de-identified and their privacy and confidentiality maintained in line with the KEMRI guidelines on research ethics.

4. Results

A total of 1104 ADI concepts were derived from the four data sources. The final reference set had 1054 unique concepts after deleting 50 duplicate concepts. The 1054 unique concepts were described by 1860 unique terms. Fig. 1 shows the number of terms and concepts used to describe ADIs in each subset.

4.1. Concept-based terms (CDC ADI and SNOMED CT subsets)

The CDC/Accenture ADI subset had 907 unique disorder concepts described by 1528 terms. For the SNOMED CT subset, the first query (*concepts described by terms including "assoc" and ("AIDS" or "HIV")*) yielded 102 concepts while the second query (*concepts linked by the relationship "associated with" (ID = 47429007) to the concept "HIV infection" (ID = 86406008)*) resulted in 105 unique concepts. Merging the two subsets described above and removing duplicates resulted in 109 unique concepts. Finally, the subtype children of the resulting concepts were derived using the fourth query. Where relevant, these concepts were replaced by their supertype concept for which the description did not contain "associated with AIDS". For the resulting 110 concepts, 254 terms together with their description ID were retrieved from SNOMED CT.

4.2. Term-based ADI subset (abstraction from MCH and PSC, and Focus Group Discussions)

A total of 87 ADI concepts were derived from the term-based subset i.e. those abstracted from the paper records of patients attending the MCH clinic and PSC, and from the FGD. The concepts were described using 157 terms.

4.3. Content coverage

The overall content coverage was 99.9% with pre-coordinated concepts and 100% with post-coordination. All but two term-based concepts from the chart abstraction at the MCH and PSC, as well as the FGDs matched the SNOMED CT concepts. The two concepts that required post-coordination were: *(i) Oropharyngeal candidiasis, and (ii) genital ulcer disease*. *Oropharyngeal candidiasis* was post-coordinated as: Candidiasis (78048006): finding site (363698007) = Mouth and/or pharynx structures (312533001). *Genital ulcer disease* was post-coordinated as: Ulcer (429040005): finding site (363698007) = Structure of genital organ (700037000).

4.4. Interface terminology implementation

The implementation of the interface terminology includes 1860 terms describing 1054 unique concepts. The data entry forms, developed using the HTML Form Entry module of OpenMRS, display ADI concepts from the reference set and include qualifiers such as severity (e.g. “mild”, “moderate” or “severe”), onset (e.g. “sudden” or “gradual”) and laterality which describes the side of the body that is infected (e.g. right middle zone pneumonia indicative that the infection is in the right middle lobe of the lung). Defining characteristics which include causative agent were implicit to the interface terminology, allowing more specific types of diagnosis to be selected (e.g. for “pneumococcal pneumonia” where “pneumococcus” is the causative agent). Additional context information that can be selected is the type of diagnosis, i.e. presumptive (working), differential or final diagnosis and status (“active” or “inactive”) (see Fig. 2). Once ADI diagnoses are captured, using the data entry form, the concept ID is stored and can be re-used for decision support, e.g. inferring ART eligibility and to generate statistical summaries. Statistical summaries are reported to the Kenyan Ministry of Health on a monthly basis.

5. Discussion

More than 1000 ADI concepts described by nearly 1900 terms were derived from four data sources using a well-structured approach based on an earlier published framework for developing interface terminology reference sets [16]. While most concepts were described by one term, some of the concepts were described using as many as 9 terms which provide a wide variety of relevant synonyms that include locally used terms from which clinicians can make selections to accurately describe diagnoses. Content coverage was high; consistent with the study by Rosenbloom et al. [24]. In our study, all but two concepts generated from the term-based sources matched SNOMED CT concepts.

Although signs and symptoms were included in the data abstracted from the MCH and PSC, as well as the FGD, we did not include them in the final reference set since they had a very broad scope and some were not associated with HIV infection. It is important to note that the WHO clinical staging list includes some signs and symptoms such as *fever (concept ID = 421154002*) together with ‘real’ ADIs used to determine the progression of HIV infection based on clinical presentation of the patient [8] (see Appendix B). These will be added to the local concept dictionary through dynamic updates to support automated inference of WHO clinical staging. The ADI list of concepts will be shared with Columbia University that maintains the concept dictionary for OpenMRS users. It will also be shared with the International Health Terminology Standards Development Organization (IHTSDO) that maintains SNOMED CT for distribution to its members and affiliates. Access to the reference set can be gained by contacting IHTSDO.

An observation of the term-based data abstracted from the MCH clinic and PSC, as well as data recorded from the FGDs showed a number of spelling mistakes. This shows that lack of a structured list of ADIs with a rich selection of terms contributes to poor data quality on diagnoses entered in EMR systems, as observed elsewhere [13]. The implementation of an interface terminology with the ADIs selected from a comprehensive reference set that includes child and parent concepts of key disorders will enhance the quality of data recorded

leading to appropriate decisions on clinical care. Additionally, enhanced decision support through automated inference of WHO clinical stage based on the recorded ADIs will ensure that patients get appropriate treatment determined by the accurate recording of progression of HIV infection. The study benefited from data collection from a busy referral hospital that provides HIV care and treatment to a large number of patients with diverse conditions hence a broad coverage of terms used to describe ADIs. Furthermore, the use of chart abstraction and FGDs enabled the study team to collect context-specific terms that are most appropriate for local use at JOOTRH but may also be broadly used in similar contexts in low-resource settings.

Our study had a few limitations. The chart abstraction was conducted over a two-month period and this may not be long enough to provide the breadth of data required. The effect of the short data collection period was minimized since this is a high-volume hospital with diverse patients. Incorrect spellings and use of ADI terms that were not specific enough, made matching SNOMED CT concepts difficult. Consultation with the clinicians helped resolve the errors and all the terms were matched. Additionally, the EMR implementation has a provision for dynamic update of concepts which may be missing from the current list or those that will be derived through post-coordination of new and existing terms.

6. Conclusion

We described a structured approach to deriving an ADI reference set based on SNOMED CT and the implementation of an interface terminology for OpenMRS at a busy provincial referral hospital in a resource-constrained setting. Use of different sources demonstrates complementarity in the collection of concepts and terms for an interface terminology. SNOMED CT provides a high coverage in the domain of ADIs. We anticipate that the context-specific reference set will support improved recording of high quality data ensuring completeness and reusability. Further work is needed to evaluate the effectiveness of the interface terminology on data quality and quality of care.

Acknowledgments

The authors would like to acknowledge the KEMRI Director and his staff for the approval of this study and for reviewing and clearing the manuscript. We would also like to thank Ms. Mevis Omolo and Philister Anyango for outstanding work in recording and transcribing the FGD, and all staff at the JOOTRH that supported the data collection for this study and implementation of OpenMRS.

Funding

This publication was made possible by support from the U.S. President’s Emergency Plan for AIDS Relief (PEPFAR) through the U.S. Centers for Disease Control and Prevention (CDC), Division of HIV/AIDS (DGHA), under the KEMRI/CDC Cooperative Agreement no. GH000048-04.

Appendix A. SQL queries for retrieval of ADI concept IDs for ADIs from SNOMED CT database

We used the queries below to retrieve the relevant ADI concepts and descriptions:

(i) ***Concepts described by terms including "assoc" and ("AIDS" or "HIV")***

```
CREATE VIEW c1 AS
SELECT c201401.Id, c201401.active, c201401.definitionStatusId, d201401.conceptID, d201401.typeId,
d201401.term
FROM c201401
JOIN d201401 ON c201401.Id = d201401.conceptId
WHERE c201401.active = 1
AND d201401.active = 1
AND (
d201401.term LIKE '%AIDS%'
OR d201401.term LIKE '%HIV%'
)
AND d201401.term LIKE '%assoc%'
```

(n = 102 unique concepts)

(ii) ***Concepts linked by the relationship "associated with" (Id = 47429007) to the concept "HIV Infection" (Id = 86406008)***

```
CREATE VIEW 'c2' AS
select 'sourceconcept'.'Id' AS 'conceptId'
FROM ((('c201401' 'sourceconcept' join 'r201401' on(('sourceconcept'.'Id' = 'r201401'.'sourceId')))
JOIN 'transitiveclosure_20140131' 't1' ON (('r201401'.'destinationId' = 't1'.'SubtypeId')))
JOIN 'transitiveclosure_20140131' 't2' ON (('t2'.'SubtypeId' = 'r201401'.'typeId')))
WHERE (('sourceconcept'.'active' = 1) and ('r201401'.'active' = 1)
AND ('t1'.'SupertypeId' = 86406008)
AND ('t2'.'SupertypeId' = 47429007))
```

(n = 105 unique concepts)

(iii) ***Merging the two views c1 and c2 into a new one c3***

```
CREATE VIEW 'c3' AS
SELECT 'c1'.'conceptID' AS 'conceptId'
FROM 'c1' union select 'c2'.'conceptId' AS 'conceptId' from 'c2'
```

(n = 109 unique concepts)

(iv) ***Concept Identifiers of all unique subtype concepts of AIDS-associated disorder concepts (above)***

```
CREATE VIEW c4 AS
SELECT DISTINCT subtypeId
FROM transitiveclosure_20140131
JOIN c3 ON supertypeId = conceptId
```

(n = 110 unique concepts)

c201401 is the concepts table
d201401 is the description table
r201401 is the relationship table

Appendix B. WHO clinical staging list used at the JOOTRH

	Description
	Clinical stage 1
1	Asymptomatic
2	Persistent generalized lymphadenopathy
	Clinical stage 2
1	Moderate unexplained weight loss (<10% of presumed or measured body weight)
2	Minor mucocutaneous manifestations (seborrheic dertmatitis, papular pruritic eruptions, fungal nail infections, recurrent oral ulcerations, angular cheilitis)
3	Herpes zoster
4	Recurrent upper respiratory tract infections (sinusitis, tonsillitis, bronchitis, otitis media, pharyngitis
	Clinical stage 3
1	Unexplained severe weight loss (over 10% of presumed or measured body weight)
2	Unexplained chronic diarrhea for longer than one month
3	Unexplained persistent fever (intermittent or constant for longer than one month)

	Description
4	Persistent oral candidiasis
5	Oral hairy leukoplakia
6	Pulmonary tuberculosis
7	Severe bacterial infections (e.g. pneumonia, empyema, pyomyositis, bone or joint infection, meningitis, bacteraemia)
8	Acute nectrotizing ulcerative stomatitis, gingivitis or periodontitis
9	Unexplained anemia (hemoglobin below 8 g/dl), neutropenia (below 0.5 109/l) and/or chronic thrombocytopenia (below 50 10^9/l)
	Clinical stage 4
	Conditions where a presumptive diagnosis can be made using clinical signs or simple investigations
1	HIV wasting syndrome
2	Pneumocystis jirovecii pneumonia (PCP)
3	Recurrent severe bacterial pneumonia (>2 episodes within 1 year)
4	Cryptococcal meningitis
5	Toxoplasmosis of the brain
6	Chronic orolabial, genital or ano-rectal herpes simplex infection for >1 month
7	Kaposi sarcoma
8	HIV encephalopathy
9	Extrapulmonary tuberculosis
	Conditions where confirmatory diagnosis testing is necessary:
1	Cryptosporidiosis, with diarrhea >1 month
2	Isosporiasis
3	Cryptococcosis (extrapulmonary)
4	Disseminated non-tuberculosis mycobacterial infection
5	Cytomegalovirus (CMV) retinitis or infection of the organs (other than liver, spleen, or lymph nodes)
6	Progressive multifocal leocoencephalopathy (PML)
7	Any disseminated mycosis (e.g. histoplasmosis, coccidiomycosis)
8	Candidiasis of the oesophagus or airways
9	Non-typhoid salmonella (NTS) septicaemia
10	Lymphoma cerebral or B cell Non Hogdkins's Lymphoma
11	Invasive cervical cancer
12	Visceral leishmaniasis
13	Symptomatic HIV-associated nephropathy or HIV-associated cardiomyopathy

References

1. Lawn SD, Gupta A, Wood R. Assessing the impact of prevalent tuberculosis on mortality among antiretroviral treatment initiators: accurate tuberculosis diagnosis is essential. AIDS. 2012; 26(13): 1730–1731. [PubMed: 22874482]
2. Toossi Z, Mayanja-Kizza H, Hirsch CS, Edmonds KL, Spahlinger T, Hom DL, et al. Impact of tuberculosis (TB) on HIV-1 activity in dually infected patients. Clin Exp Immunol. 2001; 123(2): 233–238. [PubMed: 11207653]
3. Grabar S, Lanoy E, Allavena C, Mary-Krause M, Bentata M, Fischer P, et al. Causes of the first AIDS-defining illness and subsequent survival before and after the advent of combined antiretroviral therapy. HIV Med. 2008; 9(4):246–256. [PubMed: 18366449]

4. Ledergerber B, Egger M, Erard V, Weber R, Hirschel B, Furrer H, et al. AIDS-related opportunistic illnesses occurring after initiation of potent antiretroviral therapy: the Swiss HIV Cohort Study. JAMA. 1999; 282(23):2220–2226. [PubMed: 10605973]
5. World Health Organization. Global Update on HIV Treatment 2013: Results, Impact and Opportunities. 2013 Jun 6.
6. Kaplan JE, Benson C, Holmes KK, Brooks JT, Pau A, Masur H. Guidelines for prevention and treatment of opportunistic infections in HIV-infected adults and adolescents: recommendations from CDC, the National Institutes of Health, and the HIV Medicine Association of the Infectious Diseases Society of America. MMWR Recomm Rep. 2009; 58(RR-4):1–207.
7. World Health Organizations. Consolidated Guidelines on the Use of Antiretroviral Drugs for Treating and Preventing HIV Infection. 2013 Jun 6.
8. World Health Organization. WHO Case Definitions of HIV for Surveillance and Revised Clinical Staging and Immunological Classification of HIV-related Disease in Adults and Children. 2006
9. Castelnuovo B, Kiragga A, Afayo V, Ncube M, Orama R, Magero S, et al. Implementation of provider-based electronic medical records and improvement of the quality of data in a large HIV program in Sub-Saharan Africa. PLoS ONE. 2012; 7(12):e51631. [PubMed: 23284728]
10. Kiragga AN, Castelnuovo B, Schaefer P, Muwonge T, Easterbrook PJ. Quality of data collection in a large HIV observational clinic database in sub-Saharan Africa: implications for clinical research and audit of care. J Int AIDS Soc. 2011; 14:3. [PubMed: 21251327]
11. Libamba E, Makombe S, Mhango E, de Ascurra TO, Limbambala E, Schouten EJ, et al. Supervision, monitoring and evaluation of nationwide scale-up of antiretroviral therapy in Malawi. Bull World Health Organ. 2006; 84(4):320–326. [PubMed: 16628306]
12. Hojen AR, Goeg KR, Elberg PB. Re-use of SNOMED CT subset in development of the Danish national standard for home care nursing problems. Stud Health Technol Inform. 2015; 210:140–144. [PubMed: 25991118]
13. Shah AD, Martinez C, Hemingway H. The freetext matching algorithm: a computer program to extract diagnoses and causes of death from unstructured text in electronic health records. BMC Med Inform Decis Mak. 2012; 12:88. [PubMed: 22870911]
14. Oluoch T, N de K, Kwaro D, Wattoyi I, Okeyo N, Cornet R. Inconsistencies between recorded opportunistic infections and WHO HIV staging in western Kenya. Stud Health Technol Inform. 2013; 192:1139. [PubMed: 23920913]
15. Liaw ST, Taggart J, Yu H, Lusignan SD, Kuziemsky C, Hayen A. Integrating electronic health record information to support integrated care: practical application of ontologies to improve the accuracy of diabetes disease registers. J Biomed Inform. 2014
16. Bakhshi-Raiez F, Ahmadian L, Cornet R, E de J, De Keizer NF. Construction of an interface terminology on SNOMED CT. Generic approach and its application in intensive care. Methods Inf Med. 2010; 49(4):349–359. [PubMed: 20582384]
17. Kenyan Ministry of Health. Guidelines for Antiretroviral Treatment in Kenya. 2014 Jun 6.
18. International Health Terminologies Standards Development Organization, SNOMED CT. <http://www.ihtsdo.org/snomed-ct/> (accessed 10.10.14)
19. Rosenbloom ST, Miller RA, Johnson KB, Elkin PL, Brown SH. Interface terminologies: facilitating direct entry of clinical data into electronic health record systems. J Am Med Inform Assoc. 2006; 13(3):277–288. [PubMed: 16501181]
20. Clinic Information Consultancy Ltd. CliniClue Xplore, CliniClue Xplore. <http://www.cliniclue.com/cliniclue_xplore> (10.10.14)
21. NICTIZ, Terminologiecentrum Nictiz, NICTIZ. <http://terminologie.nictiz.nl/art-decor/snomed-ct> (10.10.14).
22. US Centers for Disease Control and Prevention, Accenture, Unpublished List of AIDS Defining Illnesses, 6-6-2012.
23. US Centers for Disease Control and Prevention, Definition of Oropharyngeal Candidiasis. <http://www.cdc.gov/fungal/diseases/candidiasis/thrush/> (10.10.14)
24. Rosenbloom ST, Brown SH, Froehling D, Bauer BA, Wahner-Roedler DL, Gregg WM, et al. Using SNOMED CT to represent two interface terminologies. J Am Med Inform Assoc. 2009; 16(1):81–88. [PubMed: 18952944]

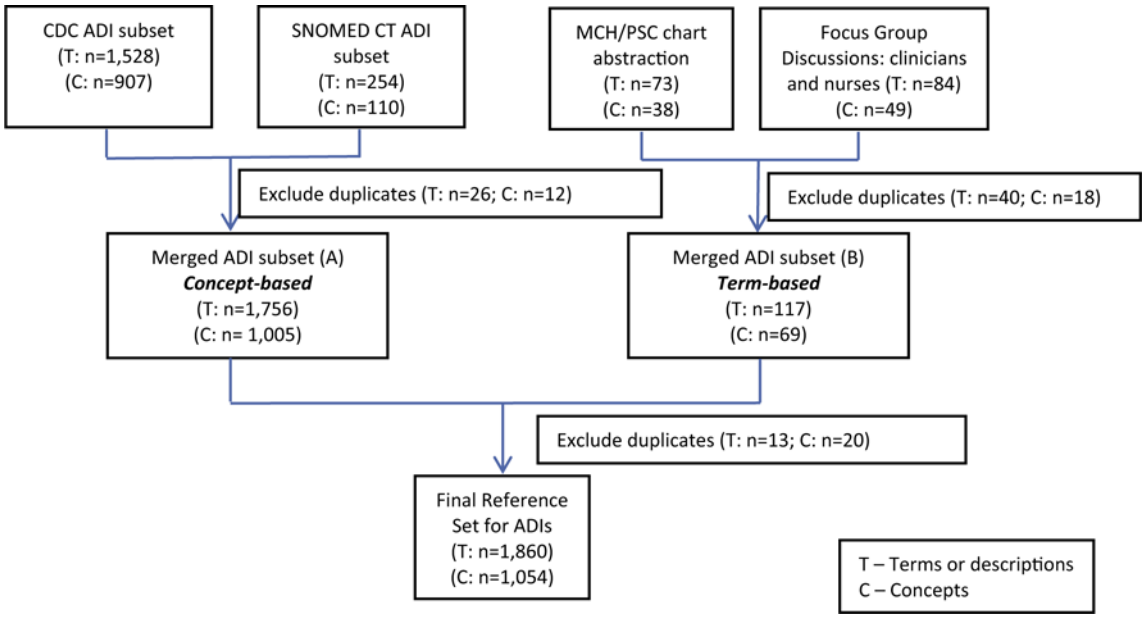

Fig. 1.

Derivation of terms and concepts used to describe ADI from concept-based and term-based sources.

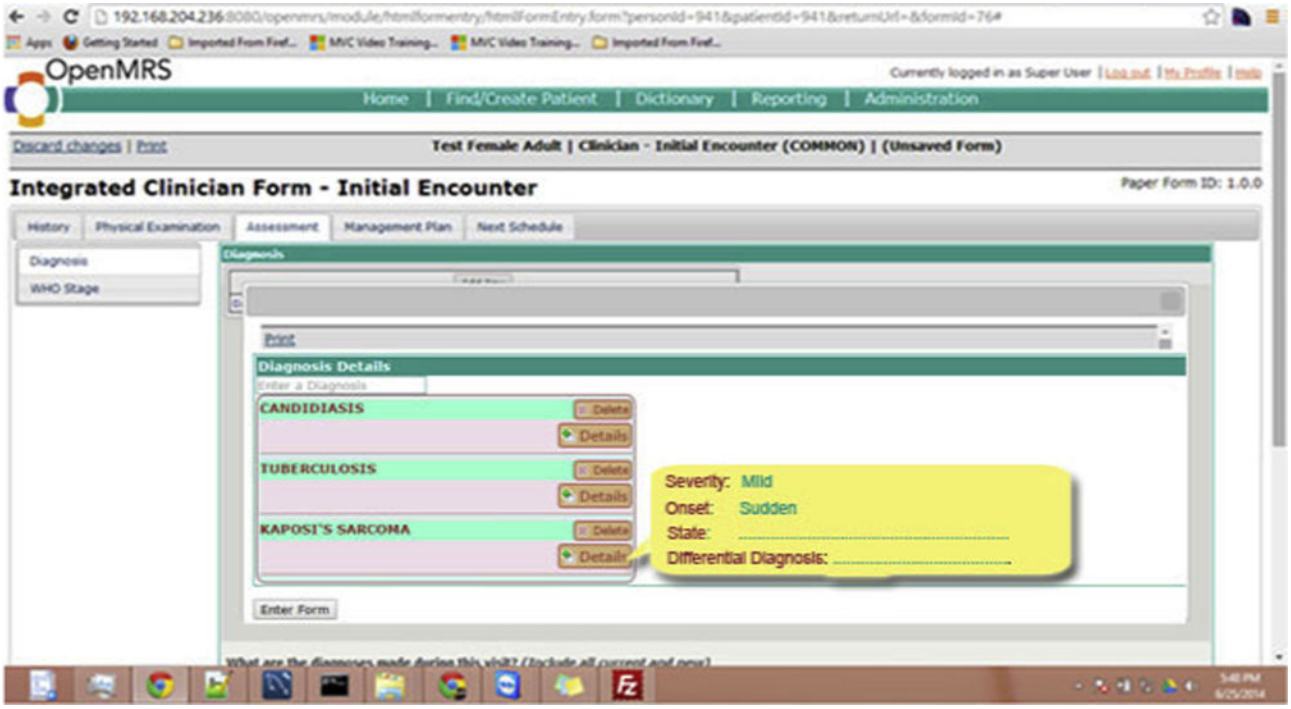

Fig. 2.
A screen-shot of a data entry form using ADI reference set.