REVIEW

An ontology for factors affecting tuberculosis treatment adherence behavior in sub-Saharan Africa

Olukunle Ayodeji
Ogundele[1]
Deshendran Moodley[1]
Anban W Pillay[1]
Christopher J Seebregts[1,2]

[1]UKZN/CSIR Meraka Centre for
Artificial Intelligence Research and
Health Architecture Laboratory,
School of Mathematics, Statistics
and Computer Science, University
of KwaZulu-Natal, Durban,
KwaZulu-Natal, [2]Jembi Health
Systems NPC, Cape Town,
South Africa

Correspondence: Olukunle Ayodeji
Ogundele
UKZN/CSIR Meraka Centre for Artificial
Intelligence Research and Health
Architecture Laboratory, School of
Mathematics, Statistics and Computer
Science, University of KwaZulu-Natal,
H1 Block, Westville Campus, University
Road, Durban 3629, South Africa
Tel +27 78 617 0144
Email zinmanship@yahoo.co.uk

Purpose: Adherence behavior is a complex phenomenon influenced by diverse personal, cultural, and socioeconomic factors that may vary between communities in different regions. Understanding the factors that influence adherence behavior is essential in predicting which individuals and communities are at risk of nonadherence. This is necessary for supporting resource allocation and intervention planning in disease control programs. Currently, there is no known concrete and unambiguous computational representation of factors that influence tuberculosis (TB) treatment adherence behavior that is useful for prediction. This study developed a computer-based conceptual model for capturing and structuring knowledge about the factors that influence TB treatment adherence behavior in sub-Saharan Africa (SSA).

Methods: An extensive review of existing categorization systems in the literature was used to develop a conceptual model that captured scientific knowledge about TB adherence behavior in SSA. The model was formalized as an ontology using the web ontology language. The ontology was then evaluated for its comprehensiveness and applicability in building predictive models.

Conclusion: The outcome of the study is a novel ontology-based approach for curating and structuring scientific knowledge of adherence behavior in patients with TB in SSA. The ontology takes an evidence-based approach by explicitly linking factors to published clinical studies. Factors are structured around five dimensions: factor type, type of effect, regional variation, cross-dependencies between factors, and treatment phase. The ontology is flexible and extendable and provides new insights into the nature of and interrelationship between factors that influence TB adherence.

Keywords: tuberculosis, treatment adherence behavior, influencing factor, conceptual model, ontology

Introduction

Poor adherence or nonadherence of patients with tuberculosis (TB) to prescribed treatment is a major contributor to treatment failure.[1–3] Treatment adherence behavior (TAB) is defined as the extent to which a person's practice of taking medication, following a diet, and/or executing lifestyle changes corresponds with agreed recommendations from a health care provider.[4] Thus, poor adherence is the failure of patients with TB to take medication or follow a diet and lifestyle in accordance with the prescription given by a health worker.[5] Patients with TB who exhibit poor adherence to treatment over a period of time have a high risk of becoming resistant to prescribed drugs that may eventually become life-threatening.

Adherence is a complex and dynamic phenomenon with a wide range of interacting socioeconomic factors impacting on a patient's adherence behavior.[6] These factors vary in both granularity and the extent of their effects on adherence behavior across different

socioeconomic statuses and geographical regions and can sometimes have opposite effects under different circumstances. Understanding adherence behavior in patients with TB is important for effective and efficient treatment planning, and improved understanding of fluctuations in treatment outcomes in disease-monitoring programs.[4] Knowledge of the pattern of influencing factors and adherence behavior is also useful for decision support in TB disease control programs. Taking patients' subjective treatment experiences into consideration can facilitate patient-centered interventions and become a tool to better promote treatment adherence.[6] Structured and systematic synthesis of qualitative research can contribute to improved understanding, interpretation, and comparison of the growing volume of studies about patients' adherence to treatment.[6]

A number of systems have emerged to analyze, structure, and compare knowledge about influencing factors contributing to adherence in patients with TB.[4,6,7] These systems are limited in terms of their comprehensiveness and representational support; some categories are vague and ambiguous, and there are fundamental semantic differences between the classification systems, which make them incompatible with each other. Transforming the current systems into a holistic formal and computational model is a step toward specifying a common, consistent, and unambiguous vocabulary and structure for consolidating the current knowledge around TB adherence behavior. This consolidated knowledge, or knowledge repository, can form the basis for building adherence risk predication models for specific communities to identify knowledge gaps and inform further research studies into adherence behavior.

To this end, we developed a conceptual model for structuring, curating, and uncovering scientific knowledge about factors influencing TB adherence. The study presents an evidence-based model that is essential for clear identification and understanding of community-specific factors that influence TB patients' adherence and identify communities at risk.

The conceptual model is formalized as an ontology and expressed in the web ontology language[8] (OWL). An ontology is an explicit specification of a conceptualization.[9] Ontologies have been used successfully to represent concepts in the public health domain.[10–12] OWL is the most widely used language for expressing and sharing ontologies. It is designed to represent rich and complex knowledge about things, groups of things, and relations between things (http://www.w3.org/2001/sw/wiki/OWL). SNOMED CT (Systematized Nomenclature of Medicine – Clinical Terms) is represented as an ontology with OWL.[13]

The ontology is extendable, can be navigated and queried, and is useful for computer-based prediction. The ontology was evaluated for its effectiveness in representing and classifying factors associated with adherence to TB treatment in sub-Saharan African (SSA) countries.

The remainder of the paper is organized as follows: the methods followed in the study are detailed in the "Methodology" section. Existing categorization models are reviewed in the "Review of existing categorization models" section. The conceptual model is presented in the "Development of a conceptual model and ontology" section, and the ontology that is evaluated is described in the "Evaluation of the conceptual model (ontology)" section. Finally, the "Discussion and conclusion" section is given.

Methodology

Three process steps were used to develop the TB TAB ontology. The first step, knowledge acquisition, entailed a review of the literature on treatment adherence of patients with TB to identify existing dimensions for classifying influencing factors. The second step, model development, involved the development of a conceptual model using the information extracted from the literature review and expressing this as an OWL ontology. The ontology was then evaluated in the final step.

A review of the literature was conducted to provide background knowledge required for the ontology development process. The repositories searched included Google Scholar, Science Direct (Elsevier), SCOPUS, Web of Science, EBSCO, and PubMed. Keywords such as "Tuberculosis Treatment Adherence Predictors" OR "Tuberculosis Medication Adherence Factors" were used to carry out searches for related literature. The word "treatment" was also substituted with "drugs" and "medication". The word "adherence" was substituted for "compliance", and the word "factor" was substituted for "predictor". Some of the search phrases used for the search include the following:

- Factors influencing (medication/treatment) (compliance/adherence) behavior of tuberculosis patient
- Factors influencing tuberculosis patient (poor/non) (compliance/adherence) with prescribed (medication/drug)
- Predictors of (drug/medication/treatment) (compliance/adherence) behavior of tuberculosis patient
- Predictors of tuberculosis patient (poor/non) (compliance/adherence) to prescribed (drug/medication/treatment)

Scientific papers were collated and analyzed iteratively as base knowledge for developing the model. A total of 66 papers were initially identified in the review. Twenty-one of these were excluded because they did not focus on determining the influencing factors (predictors) of TB TAB. The remaining 45 papers were classified into clinical studies or review papers.

Eight review papers were selected and used as a basis for formulating the classification dimensions. Five papers explicitly proposed categorization systems or identified categories while the remaining three papers supplemented the general formulation of the final categories.

Thirty-seven papers that reported on clinical studies were used to identify factors that influence adherence for specific communities that can be included in the model. Six of these papers were excluded because they did not focus on factors that influence TB patients' adherence. Of the remaining 28 papers, only 14 focused on patients with TB in SSA countries. These 14 were used to evaluate the model.

The development of the conceptual model involved the consolidation of the existing categorization systems and identification of dimensions for representing and structuring factors. Categorization dimensions were extracted from published papers through a manual process. A conceptual model that effectively represents the complexity of factors and objectively captures existing domain knowledge (from the literature) was developed using an iterative process. The conceptual model was formalized into an OWL ontology by following a rigorous ontology engineering method that was adapted from the Unified Process for Ontology Building[14] methodology.

The correctness and comprehensiveness of the ontology in capturing and extending knowledge of factors that influence treatment adherence of patients with TB in SSA were evaluated and validated. First, a comparative analysis with the existing categorization was carried out to verify the representativeness of the model. Second, we validated the effectiveness of the model in representing the nuances of the influencing factors by using the model to capture scientific publications that provide information about patients with TB in SSA. Finally, we validated the use of the ontology for building predictive models by using it to construct a Bayesian decision network model for SSA TB communities.

Review of existing categorization models

Several categorizations of factors contributing to adherence behavior have been published.[4,6,15] These earlier studies carried out an assessment of these factors for the purpose of providing a better understanding of the relationship between the factors and patients' adherence, and for proposing appropriate intervention strategies. These studies include the World Health Organization (WHO) study,[4] a systematic review and study by Munro et al,[6] and a quantitative literature review by Jin et al.[15] These three studies presented dimensions for categorizing influencing factors. Additional categorization concepts that are not evidence based but, nonetheless, are useful for categorizing influencing factors have been proposed, eg, temporal variation proposed by Castelnuovo[16] and Kruk et al.[17]

The WHO model

A study by the WHO was aimed at structuring appropriate intervention plans for several infectious and chronic diseases.[4] This is the earliest known attempt to consolidate knowledge about influencing factors for comprehensive intervention plans for different types of diseases. The study draws on several qualitative and quantitative studies to present a categorization with five major categories: patient-related, socioeconomic, health system, therapy-related, and condition-related. Second, two categories were presented based on the type of effect: positive factors that stimulate patients to adhere more and negative factors that cause a decrease in adherence.[4]

Munro et al's model

Munro et al[6] conducted a systematic review of the literature from 1999 to 2005 and developed a model for categorizing TB influencing factors. The review was aimed at understanding which factors are considered important by patients with TB, caregivers, and health care providers. A total of 44 articles drawn from different regions of the world were reviewed. From the study, four main categorization themes were developed. The four themes are as follows: structural factors, personal factors, social context factors, and health service factors.

Jin et al's model

Jin et al[15] identified some categorizations for representing influencing factors through a systematic review of 102 articles that focused on all types of therapy for several chronic and infectious diseases. The study examined common factors causing therapeutic nonadherence from the patient's perspective and identified three dimensions for classifying these factors.

First, they presented five categories based on factor type: patient-centered, therapy-related, health care system, social and economic, and disease-related. Second, they presented three categories based on the type of effect: compliance increment, compliance decrement, and no effect. Third, they presented three categories based on difficulties encountered in measuring the effect and counter intervention of the factors. They are hard factors, whose impacts are more quantifiable, and soft factors, whose effects are difficult to measure and counter.

Temporal concept

Two categories were identified through a review of six studies carried out by Castelnuovo[16] to depict the period of effect of factors. The categories relate to the treatment phases of an anti-TB treatment plan. The first is the "intensive phase", which is the first 2 months of anti-TB treatment after the patients are diagnosed with TB. The second is the "continuation phase", which starts immediately after the intensive phase and continues for 4–6 months.[16] Other temporal representations are the weekly and monthly categorizations introduced by Kruk et al.[17] They reviewed 14 studies that focused on the timing of default in low-income countries' TB treatment.

Challenges of the existing categorization

Variations in the models presented in existing studies pose challenges for the common and shareable representation of factors. For instance, the factor type categories identified across the papers may appear similar, but the description of the categories and the factors belonging to each category vary. There are variations in the number of categories presented under the same dimensions. The WHO[4] study proposed five categories, Munro et al[6] developed a model of four categories, and Jin et al[15] identified five categories, which are similar to the WHO's categories. Similarly, the type of effect proposed by the WHO and Jin et al is different. Although the WHO proposed three categories, Jin et al proposed two categories. A comparison of the different categorization systems is given in Table 1.

Additionally, the naming and definition of existing categories are inconsistent. There are no generally accepted names for the categories. For instance, patient-related factors have different names and meanings across the three models. They are named as personal factors in Munro et al and patient-related factors in the WHO and Jin et al. The WHO's patient-related-factor category focuses on patient demographic information and excludes certain lifestyle and psychological attributes included in Jin et al's category.

There is also no uniformity in the classification hierarchy; some of the existing models introduce subcategories, while others do not. In the absence of subcategories, factors are directly grouped under the main categories. Jin et al introduced two subcategories in their classification only for the "patient-centered" category, and they are the demographic and psychological factor categories. Munro et al used the eight themes as the intermediate groups, but the relationships with the four themes are not clearly defined. The WHO report did not provide any subcategories in its classification.

Finally, none of the categorization systems represent all the categorization dimensions identified in Table 1. While some represent more than one dimension in their studies, others concentrate only on one dimension. Three of the five studies, WHO,[4] Munro et al,[6] and Jin et al,[15] focused on categorizing factors, ie, the factor type dimension. Two studies classified factors according to the type of effect. Two studies focused solely on the period of effect.

Table 1 Existing influencing factor categorizations

Dimension	WHO[4]	Munro et al[6]	Jin et al[15]	Castelnuovo[16]	Kruk et al[17]
Factor type	Patient-related factors	Personal factors	Patient-centered factors		
	Therapy-related factors	Health service factors	Therapy-related factors		
	Health system factors	Social context factors	Health care system factors		
	Socioeconomic factors	Structural factors	Social and economic factors		
	Condition-related factors		Disease-related factors		
Type of effect	Positive factors		Compliance increment factors		
	Negative factors		Compliance decrement factors		
			No-effect factors		
Measurement			Hard factors		
			Soft factors		
Temporal				Intensive phase	Weekly/monthly
				Continuation phase	

Abbreviation: WHO, World Health Organization.

Some dimensions are not incorporated across all categorizations. One of these is the cross-dependency between influencing factors. Some clinical studies have established cross-dependencies among factors, ie, a factor's influence is dependent on another factor.[18]

Development of a conceptual model and ontology
Restructuring existing categorizations into a common conceptual model

The proposed conceptual model is aimed at representing, collating, and structuring knowledge found in the literature in a consistent manner for clear understanding and classification of the factors. The model is intended to be used as a formal basis to develop the ontology. Five dimensions were identified from the review of existing categorizations. They have been restructured in order to have a complete and unique representation of the influencing factors and their application to patients with TB in SSA. The key elements of the classification as drawn from the review are factor type, type of effect, treatment phase, region, and cross-dependency.

Factor type

Factor type represents the grouping of influencing factors according to the similarity of common terms as presented in the literature. This type of grouping enables the creation of a category, sometimes in a hierarchy, to assist in distinguishing terms. It is a common dimension for categorizing influencing factors.

We used the classifications found in the three existing studies to develop unique and specific factor type categories. The existing categories were restructured to eliminate concept overlaps and misrepresented factors. They were iteratively checked in terms of their effectiveness to classify factors found in scientific publications.

The process of restructuring the categories involves matching of existing categories based on the similarity of names and meaning. Similar factor type categories were merged to produce a comprehensive category. In addition, some of the broad categories that represent heterogeneous factors were split to produce unique categories without unnecessary overlap. Through this process, seven factor types were defined and their boundaries were set to facilitate the inclusion of factors from scientific evidence. They are patient-centered, social, economic, therapy-related, health system, lifestyle, and geographical access.

A hierarchical model was introduced to capture the factor type in a consistent manner. The top level of the hierarchy includes the main categories, while the second level represents subgroups of factors. This second level is generated from some ad hoc groupings found in existing studies. The lowest level in the hierarchy will represent concrete and measurable influencing factors. Table 2 shows the proposed model with new categories developed from the existing models.

The patient-centered category was created by merging related categories and was redefined. The term patient-centered was taken from the study by Jin et al[15] as against "patient-related" in the WHO[4] and Munro et al's[6] "personal character". The category also reflects the definition given

Table 2 Three-level hierarchy of factors based on the factor type

Top level	Middle level	Bottom level
Patient-centered	Demographic	Age group
		Sex
		Marital status
	Knowledge	Knowledge of TB
		Education level
	Psychology	Emotional state
		Psychiatric condition
		Depression
Economic	Finance	Income class
		Poverty
	Employment	Job class
		Employment status
	Basic amenities	Lack of food
		Homelessness
Social	Social network	Family support
		Community network
	Stigma-related	Perceived stigma
		Experienced stigma
	Belief	Wellness perceived as cured
		Treatment efficacy belief
Therapy	Therapy effect	Drug adverse effect
		Symptoms persistence
	Comorbidity	HIV coinfection
	Treatment	Defaulting history
		Treatment alternative
Health system	Health care facility	Opening hour favorability
		Drug availability
	Health care staff	Staff friendliness
		Communication
		Gap experience
Lifestyle	Substance abuse	Alcoholism
		Smoking/tobacco usage
		Hard drug usage
	Healthy living	Diet
		Exercise
Geographical access	Location	Distance to facility
		Dwelling region
	Transportation	Travel time
		Transportation cost

Abbreviations: HIV, human immunodeficiency virus; TB, tuberculosis.

by Munro et al. The new patient-centered category is defined as the category of influencing factors based on the demographic attribute of patients and the attitude that defines the characteristics of the patients. Our patient-centered category excludes social-related factors from the definitions presented by Jin et al and Munro et al, interpretation of wellness and illness, motivation, and beliefs.[6,15] In addition, compliance history and substance abuse included in Jin et al were excluded, because they are therapy- and lifestyle-related factors, respectively.

Economic factors were separated from the social factors following Munro et al's classification to create two categories. This will allow for a unique representation of the factors in a specific category and reveal the potential of a factor to belong to more than one category. The "social factor" category represents the social context and situation of a patient while the "economic factor" category relates to the economic status and condition of the patient.

The "therapy-related" factor was adopted from WHO and Jin et al. It represents the category of influencing factors that relate to therapy difficulty faced by patients and clinical procedures that facilitate or hinder patients from adhering to treatment. It also consists of part of the disease-related factor presented by Jin et al and part of the "health service" category of Munro et al.

The "health system" category consists of influencing factors that relate to the performance of health care providers and accessibility of patients to health care service at the health facilities. The health system category is directly represented in categorizations by Jin et al and the WHO.

The "lifestyle" factor is a new category that is introduced to distinctly cover those factors related to a patient's lifestyle that are circumstantial habits developed by patients and are subject to change, eg, substance abuse, diet, and exercise. Jin et al classified some of these factors as patient-centered, and the WHO classified them as "condition-related" factors. Separating these factors into different categories will allow for a clear identification of the unhealthy lifestyle-related factors.

A "geographical access" category was also introduced to represent the category of influencing factors that relate to the location of health care facilities and the house/workplace of the patients, and accessibility costs in terms of distance, time, effort, and financial expenses. This will help in understanding both the financial and nonfinancial burden that relate to a patient's geographic access to health facilities.

Type of effect

This category represents the type of effect a factor has on patients' TB adherence, and the degree of effect represents the intensity of influence on a patient with TB. The type of effect is based on that of the WHO study. Another type was included based on the "no-effect" type identified in the study by Jin et al. The three types of effect included in this model are positive, negative, and neutral effects.

"Positive influencing factor" represents a group of factors that show significant motivating influence in the improvement of good adherence behavior. These factors are known to encourage patients to adhere to medication as prescribed by a health care officer. This category corresponds to the positive effect[4] and compliance increment.[15]

"Negative influencing factor" represents factors that show significant demoralizing influence on patients' attitudes and cause poor adherence behavior. This category corresponds to the negative effect (WHO) and compliance decrement.[15]

"Neutral influencing factors" are a group of factors that show no significant effect or correlation on patients' attitude toward adhering to treatment. This category corresponds to the no-effect category in the WHO study.

The patients' state, perception, or experience in relation to these factors makes the factors negative or positive. The sex-related factor is based on whether being a male is a negative influencing factor or being a female is a positive influencing factor. Therapy-related factors are mostly based on patient experience. Drug adverse effect, eg, is based on the treatment experience of the patients receiving TB treatment and is seen to cause poor adherence. Belief-related factors are based on the perception of patients about circumstances or conditions. An example is a patient who has a strong belief in treatment efficacy (positive influencing factor) and the lack of this is regarded as a negative influencing factor.

Treatment phase

The treatment phase factor refers to the stage during which a factor is influential during treatment. The SSA clinical cohort studies have considered measuring adherence and the defaulting rate over different treatment phases. For example, the two main TB treatment phases are the intensive and continuation phases of treatment. Previous studies have concluded that there is an increasing trend of poor adherence as patients go into the continuation treatment phase, and that more patients tend to default at the continuation than the intensive phase.[17,19,20]

Other treatment phases can be included, eg, the "drug resistance phase factor" represents the category of factors that are influential during a drug resistance treatment phase for the treatment of patients resistant to first-line regimen drugs and can be as long as 2 years.

Region

The regional variation of the influencing factors describes the existence of a factor with a significant influence, in particular, on socioeconomic regions. Although, there is no existing regional model for influencing factor classification, several studies have used geographic regions for their classification. The result of several clinical and review studies revealed that influencing factors can vary across regions. Regions can be delineated based on socioeconomic or geographic similarities. The administrative area is commonly used for classification and represents geographical regions with internationally recognized administrative boundaries and governance, eg, country and provinces. The geographical region is a representation of regions with physical boundaries or common geographical/physical features. The region does not have recognized political boundaries and governance and represents the communities where the clinical studies were carried out. Finally, the socioeconomic region is a collection of regions with social and economic similarities.

Cross-dependency

Although, cross-dependency relationships between influencing factors are not represented in current categorizations, they are common in the findings of clinical studies focusing on influencing factors. A cross-dependency relationship implies that a certain factor was found to only influence adherence behavior if another factor was present. Cross-dependency relationships are represented in a way that they link the "trigger factor" to the factors that are dependent on the trigger caused by the trigger factor. A "dependent factor" is only triggered when another factor is present.

For example, suppose some study found that being male contributes to negative adherence behavior only when there are unfavorable conditions at work,[18] then male sex is represented as a factor that is triggered by unfavorable working conditions.

An ontology for TB TAB

The section "Restructuring existing categorizations into a common conceptual model" presents an abstract conceptual model for structuring knowledge around adherence. This subsection describes the TB adherence behavior ontology, which provides a concrete, formal, and computer-accessible representation of the conceptual model.

An ontology is a specification of a conceptualization, provides an unambiguous logic-based model of some domain of reality, and allows for the representation of rich and complex knowledge about things, groups of things, and relations between things.[21] Ontologies not only allow for explicitly capturing, storing, and sharing expert knowledge but also enable computers to perform automatic reasoning, consistency checks, data analysis, and decision support.[12]

Figure 1 provides an overview of the ontology in the OWL. Key concepts of the model are represented as classes in the ontology, eg, influencing factor is represented as a class. The ontology also incorporates a class for evidence to represent and link published clinical studies that assert different adherence factors. Relationships between concepts (classes) are represented as class properties (the arcs between the nodes in Figure 1). For instance, the "evidence" class is linked with "influencing factor" by asserts-influencing factor "object property".

The factor type dimension is represented as influencing factor and is a hierarchy of categories of influencing factors. The type of effect and treatment phases are represented as a hierarchical object property that links the evidence with the
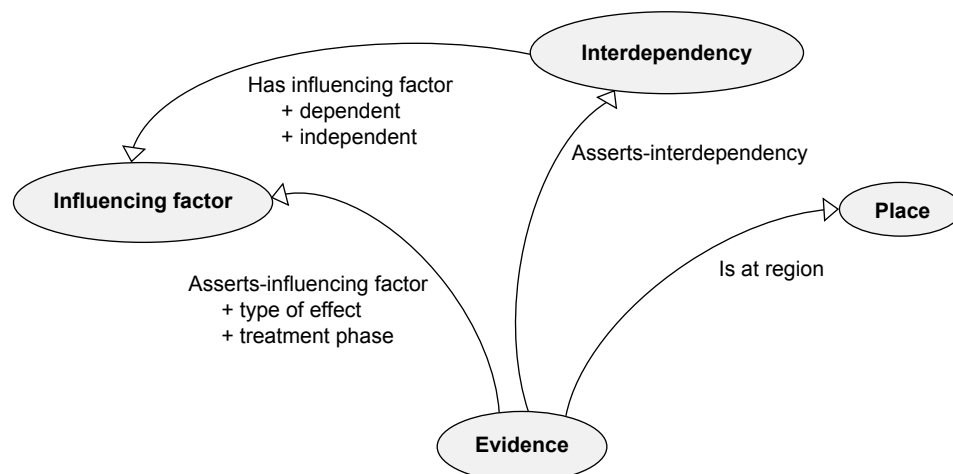
Figure 1 Overview of the key concepts and relations in the ontology.

influencing factor class. The region dimension is represented using the "place" class that is linked to evidence in order to connect each factor to a specific location. Finally, cross-dependency is represented as an interdependency class and linked to the evidence and influencing factor.

The developed ontology facilitates the categorization of influencing factors. It can be applied in the structuring of influencing factors of adherence behavior for patients with TB in SSA. It provides links to the information source, ie, scientific publications, by representing the type and period of effect as the object property and linking this to the evidence class. The ontology also maintains knowledge about where and when these studies were performed, allowing users to classify factors that fit the profile of their community.

In order to integrate the knowledge adherence with other knowledge sources, existing ontologies were incorporated and reused where possible. The evidence class is based on the evidence ontology[22] and the place class is based on the Geonames[23] ontology.

Using the ontology

The rich computational representation of the ontology is ideally suited to provide a sound basis for developing tools useful for clinicians and researchers. The ontology was used to develop a prototype web-based knowledge repository that allows users to update, navigate, and query the knowledge. The interface (Figure 2) currently allows users to navigate, filter, and search for classes and properties in the ontology.

To use the ontology, users navigate or search through the ontology to discover and select potential factors that are appropriate for a specific community. A complex search for an influencing factor can be carried out using a combination of the classes and class properties in the ontology. Categories can be navigated to find specific factors that have been identified by the published literature. Factor properties can also be filtered, eg, the type of effect can be used to identify factors that have a specific type of effect, by specifying, eg, negative influencing factors.

Community-specific influencing factors can be identified by either specifying a region of interest or describing the characteristics of the region. Search results will include factors directly associated with the specified region as well as those factors that are associated with communities that are contained within the specified region. For instance, a user may request for negative factors that can be found in Africa. By specifying Africa, the repository will include factors pertaining to communities within countries and geographical regions within Africa.
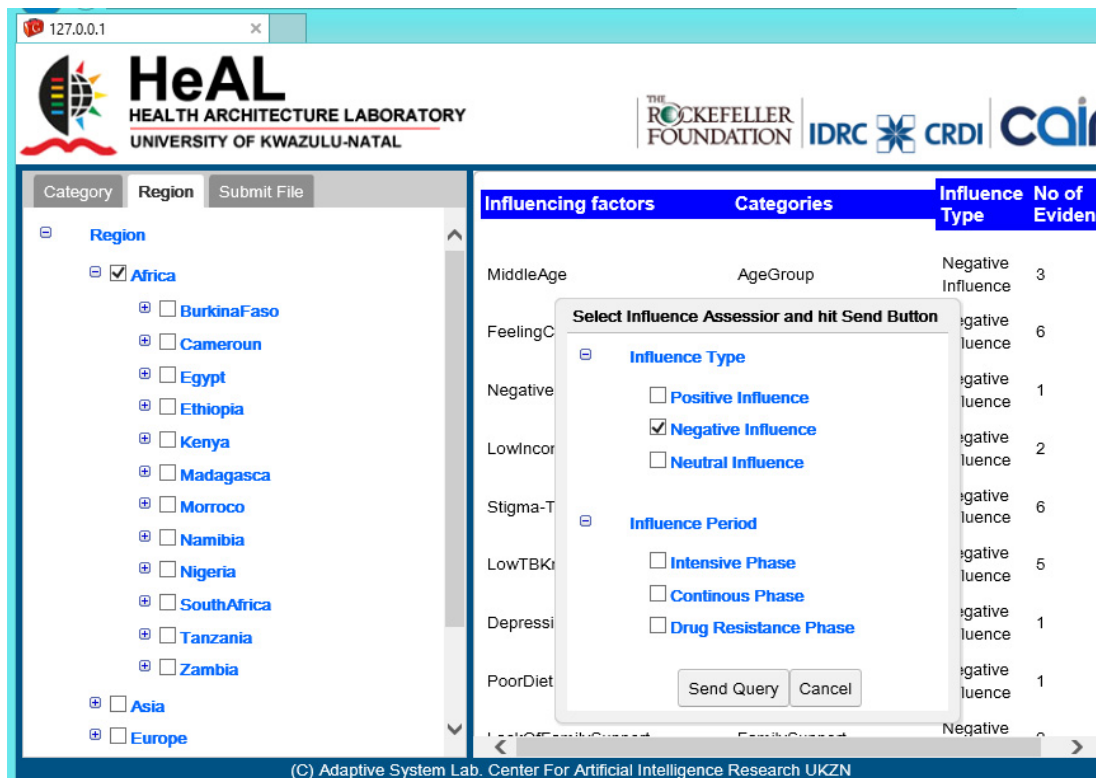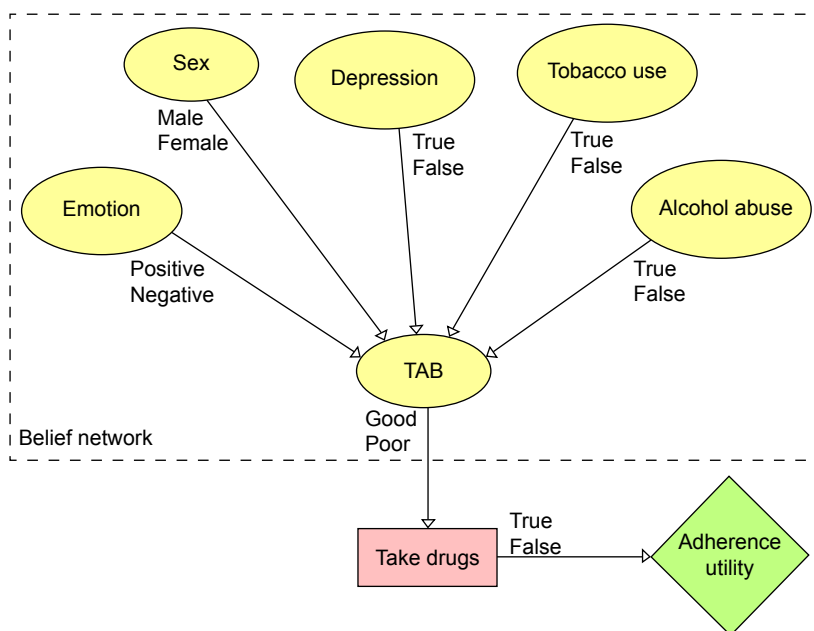

Figure 2 The interface for ontology repository.

Figure 3 A Bayesian decision network for predicting TB TAB.
Abbreviations: TAB, treatment adherence behavior; TB, tuberculosis.

The ontology was designed to be extendable. The categories in the ontology can be extended, a new category can be defined as equivalent to the collection of existing categories or factors. This is useful for those users who want to represent a different classification mechanism or introduce new categories that are not currently in the ontology. Users can easily add additional factors and associated scientific papers to the repository.

Support for Bayesian decision network construction
One of the design goals of the ontology is to aid in the building of predictive models for specific communities. The ontology allows for automating the construction of a Bayesian decision network. A Bayesian network is an annotated directed graph that encodes probabilistic relationships among distinctions of interest in an uncertain-reasoning problem.[24] In a typical usage scenario, the modeler would search the repository for and identify factors that are likely to impact on adherence in

a target community. These factors will then be used to automatically generate the causal structure of a decision network with default conditional probabilities for that community (Figure 3). The modeler must still use his/her expertise to refine and set the weightings of the conditional probabilities, or degree of effect of each factor. The resultant Bayesian decision network represents the adherence profile applicable to that community and may even be used to predict adherence behavior for individual patients in that community.

Evaluation of the conceptual model (ontology)
Comparative analysis with existing categorizations
Table 3 compares the adherence ontology in terms of its coverage with existing categorizations. The developed ontology is more comprehensive than the existing categorizations. It

Table 3 Coverage of the ontology compared with existing categorizations

Dimensions	WHO[4]	Munro et al[6]	Jin et al[15]	Castelnuovo[16]	The ontology
Factor type	✓	✓	✓		✓
Type of effect	✓		✓		✓
Treatment phase				✓	✓
Region		✓ (gp)			✓ (exp)
Difficulty of measurement			✓ (imp)		
Cross-dependency			✓ (imp)		✓ (exp)
Total dimensions covered	2	2	4	1	5

Abbreviations: exp, explicit; gp, geopolitical; imp, implicit; WHO, World Health Organization.

includes five out of six identified dimensions for influencing factor categorization extracted from the extensive literature review. Jin et al's categorization covers four of the dimensions. Both the WHO's categorization and Munro et al's models cover two dimensions. Castelnuovo's categorization only covers the treatment phase category.

One important feature that makes the ontology more comprehensive than the existing categorizations is the explicit representation of the region and the cross-dependency dimensions. Both the geographical region and the interdependency between factors have not been explicitly modeled by existing categorizations.

Representing findings in SSA communities

We tested the comprehensiveness and effectiveness of the conceptual model in representing the "nuances" of factors found in communities in SSA. Factors and their characteristics were extracted from clinical cohort studies that focused on adherence in TB communities in SSA. A total of 14 clinical studies found in the SSA region were used in the identification of factors, which were then classified and captured in the ontology. The coverage of these factors by the model was analyzed.

Factor type

The new categories provide a comprehensive range of factors identified in SSA. First, the newly created patient-centered category covers ten (71%) of the factors identified in relation to patients with TB in SSA. This matches the personal character category defined by Munro et al, although named differently. This is because our definition of patient-centered factors is similar to personal character as it includes demographic and psychological factors. Patient-centered[6] covers 86% of the factors, which is higher than the new category. Patient-related[4] categories only cover 43% and show a very narrow representation of the category (Table 4).

The new economic and social categories have a wider coverage than the "socioeconomic" category presented by the WHO and Jin et al. Eighty-six percent of the studies identified factors belonging to these classes. Economic-related factors are identified in six studies, even with the exclusion of transportation-related factors. The socioeconomic[4] covers 64%, the social and economic[15] covers 71% of the factors, while the social context[6] covers 14%. The newly created social category covers 43% of the factors. Similarly, the newly created "economic" category covers 43%, which makes it lower than condition-related[4] (71%) and

Table 4 Analysis of existing and new factor type categories

Influencing factor classifications	No of studies (14) in sub-Saharan Africa, n (%)
Patient-related[a]	6 (43)
Personal factor[b]	10 (71)
Patient-centered[c]	12 (86)
Patient-centered[d]	10 (71)
Socioeconomic[a]	9 (64)
Social context[b]	2 (14)
Social and economic[c]	10 (71)
Social[d]	6 (43)
Condition-related[a]	10 (71)
Structural[b]	9 (64)
Economic[d]	6 (43)
Therapy-related[a]	8 (57)
Therapy-related[c]	8 (57)
Clinical-related[d]	8 (57)
Health system[a]	3 (21)
Health service[b]	2 (14)
Health care system[c]	6 (43)
Health system[d]	4 (26)
Disease-related[c]	2 (14)
Lifestyle[d]	6 (43)
Geographic access[d]	5 (36)

Notes: [a]WHO;[4] [b]Munro et al;[6] [c]Jin et al';[5] [d]the ontology.
Abbreviation: WHO, World Health Organization.

"structural"[6] factors (64%). This is due to the fact that most factors in the structural- and condition-related[4] categories are incorporated into the two new categories: geographic access and lifestyle.

The new health system category covers 26% of the factors. It covers less than Jin et al's "health care system"[15] which is 43%. This is because not all factors in Jin et al's category are represented in the new category. For instance, lack of accessibility to a health care facility was included under health care system[15] and under geographic access but was excluded from the new category. The new category covers more factors than both the health system[4] (21%) and health service[6] (14%) categories.

The coverage of therapy-related matches those from the two studies, which cover 57% of the factors extracted from SSA studies. Geographic access category has 36% coverage on influencing factors identified for SSA. Lifestyle category has 43% coverage on influencing factors identified for SSA.

The new factor type categorization offers a more complete representation than the existing ones. The categories are distinct from one another and cover the factors uniquely. However, certain factors from SSA studies such as the existence of a direct observation therapy center within the district,[25] false/unknown address,[26] and outpatient method[20] did not fit into any of the new categories.

Table 5 Regional comparison of predominant influencing factors

Regions	Influencing factors category
Burkina Faso[27]	Alcoholism; defaulting history; TB knowledge
Cameroon[20]	Stigmatization; wellness perceived as cured
Ethiopia[19,28,29]	Wellness perceived as cured; age group; geographic access; education level; drug adverse effect; social network (family support); TB knowledge; finance related; alternative treatment
Kenya[7]	Health care system related; social and economic factor; patient-related factor; alcoholism; therapy-related
Madagascar[26]	Transportation time; TB knowledge; sex; communication gap experienced
Namibia[2]	Distance to health care facility; wellness perceived as cured; sex; marital status; education level (literacy); social network (family support); TB knowledge; drug adverse effect; symptoms persistence; long waiting time; lack of food; substance abuse; lifestyle
Nigeria[18]	Coinfection (HIV); sex; unfavorable working condition
South Africa[30–32]	Stigmatization; wellness perceived as cured; alcoholism; tobacco usage (smoking); poverty; incentive expectation at clinic; symptoms persistence; drug adverse effect; sex; coinfection; psychological distress
Tanzania[25]	Sex; age group; distance to facility; geographic access
Zambia[33]	Wellness perceived as cured; TB knowledge; drug availability; drug adverse effect

Abbreviation: TB, tuberculosis.

Regional variation

Regional classification of the influencing factors was carried out using countries in SSA with the aim of identifying influencing factors specific to each of these regions. This classification revealed knowledge about varying predominant influencing factors for different countries (Table 5). Although, there is wide variation in the range of factors identified for different countries, the most common categories across all countries are the patient-centered, therapy, and social-related factors.

Discussion and conclusion

Using a rigorous ontology engineering methodology, we developed an ontology, the TAB-influencing factors ontology, for representing knowledge about factors that influence TAB in patients with TB. The underlying conceptual model was developed by reformulating existing categorization systems from the literature. It incorporates more dimensions than any of the current categorization systems and was successfully used to capture most of the factors that influence TB adherence behavior in SSA found in the literature.

The ontology takes an evidence-based approach by explicitly relating each factor to published clinical studies: an important consideration for health practitioners. It presents the potential for capturing details of diverse multifaceted influencing factors and their interrelationships and complexities beyond normal human abstraction, simplification, and comprehension. For instance, the diametrically opposing influencing effects that a specific factor can have under different circumstances can be effectively represented in the ontology.

The usefulness of the TAB-IF ontology was demonstrated in an open, shareable, and extendable web-based knowledge repository. The ontology formed the computational model

that underpinned the repository and provided advanced navigation, search, and filtering capabilities. The repository can be used by program officers to navigate and find potential factors affecting TB adherence emanating from clinical studies in similar communities, and to profile communities and generate risk indices that will help simplify TB patient monitoring and follow-up activities. The ontology also provided the basis for the development of a predictive model, a Bayesian decision network that may be integrated in clinical decision support tools.

The study presents a novel ontology-based approach for consolidating and structuring knowledge about TB adherence behavior. However, a number of limitations of the study should be noted. Adherence behavior is broad, complex, and difficult to assess. The current ontology does not claim to be an exhaustive representation of factors that influence TB adherence behavior. However, the ontology was designed to be extendable to reflect custom views and a changing body of knowledge around TB adherence behavior. Although the conceptual model contains more dimensions than existing categorization systems, additional dimensions can be incorporated into the ontology. The ontology was based on knowledge extracted from scientific publications, which may not exhaustively reflect all factors and categorizations experienced in practice.

Possible future research work premised on this study could be an extension of the ontology to incorporate other dimensions that are not currently included or supported by the ontology, eg, "difficulty of measurement". Further research is required to qualify the "degree of influence" in a form that is useful to further categorize and structure influencing factors. Although the ontology focused on the knowledge of TB adherence factors in SSA, the approach is potentially applicable to other diseases and regions where adherence is

a significant factor. The proposed ontology can also be used as a basis to analyze adherence in other diseases such as HIV and can be extended beyond SSA.

Disclosure

The funders had no role in study design and data collection. The authors report no other conflicts of interest in this work.

References

1. Tachfouti N, Slama K, Berraho M, et al. Determinants of tuberculosis treatment default in Morocco: results from a national cohort study. *Pan Afr Med J*. 2013;14:121.
2. Chani K. Factors Affecting Compliance to Tuberculosis Treatment in Andara Kavango Region Namibia. Pretoria: Health Studies Department, University of South Africa; 2010.
3. Johansson E, Long NH, Diwan VK, Winkvist A. Attitudes to compliance with tuberculosis treatment among women and men in Vietnam. *Int J Tuberc Lung Dis*. 1999;3(10):862–868.
4. Sabaté E, Organization WH. *Adherence to Long-term Therapies: Evidence for Action*. Geneva: World Health Organization; 2003.
5. CDC. Self-study modules on tuberculosis: patient adherence to tuberculosis treatment. *Prevention CfDCa*. Atlanta, GA: Centers for Disease Control and Prevention; 1999:1–123.
6. Munro SA, Lewin SA, Smith HJ, Engel ME, Fretheim A, Volmink J. Patient adherence to tuberculosis treatment: a systematic review of qualitative research. *PLoS Med*. 2007;4(7):1230–1245.
7. Muture B, Keraka M, Kimuu P, Kabiru E, Ombeka V, Oguya F. Factors associated with default from treatment among tuberculosis patients in Nairobi province, Kenya: a case control study. *BMC Public Health*. 2011; 11(1):696.
8. Antoniou G, van Harmelen F. Web Ontology Language: OWL. In: Steffen S, Rudi S, editors. *Handbook on Ontologies*. Berlin: Springer; 2009:91–110.
9. Gruber TR. Toward principles for the design of ontologies used for knowledge sharing. *Int J Hum Comput Stud*. 1995;43(5–6):907–928.
10. Eilbeck K, Jacobs J, McGarvey S, Vinion C, Staes C. Exploring the Use of Ontologies and Automated Reasoning to Manage Selection of Reportable Condition Lab Tests from LOINC. Proceedings of the 4th International Conference on Biomedical Ontology, ICBO 2013, Montreal, Canada, July 7–12, 2013:12–15.
11. Dieng-Kuntz R, Minier D, Ruzicka M, Corby F, Corby O, Alamarguy L. Building and using a medical ontology for knowledge management and cooperative work in a health care network. *Comput Biol Med*. 2006;36(7–8):871–892.
12. Bodenreider O. Biomedical ontologies in action: role in knowledge management, data integration and decision support. *Yearb Med Inform*. 2008;47(suppl 1):67–79.
13. Organisation IHTSD. *SNOMED CT URI Standard*. International Health Terminology Standards Development Organisation; 2014:19.
14. De Nicola A, Missikoff M, Navigli R. A software engineering approach to ontology building. *Inf Syst*. 2009;34(2):258–275.
15. Jin J, Sklar GE, Min Sen Oh V, Chuen Li S. Factors affecting therapeutic compliance: a review from the patient's perspective. *Ther Clin Risk Manag*. 2008;4(1):269–286.
16. Castelnuovo B. A review of compliance to anti tuberculosis treatment and risk factors for defaulting treatment in Sub Saharan Africa. *Afr Health Sci*. 2010;10(4):320–324.
17. Kruk ME, Schwalbe NR, Aguiar CA. Timing of default from tuberculosis treatment: a systematic review. *Trop Med Int Health*. 2008;13(5): 703–712.
18. Daniel O, Oladapo O, Alausa O. Default from tuberculosis treatment programme in Sagamu, Nigeria. *Niger J Med*. 2005;15(1):63–67.
19. Estifanos Biru S, Bernt L. Determinants of treatment adherence among smear-positive pulmonary tuberculosis patients in Southern Ethiopia. *PLoS Med*. 2007;4(2):e37.
20. Pefura Yone EW, Kengne AP, Kuaban C. Incidence, time and determinants of tuberculosis treatment default in Yaounde, Cameroon: a retrospective hospital register-based cohort study. *BMJ Open*. 2011; 1(2):e000289.
21. Ternai K, Török M, Varga K. Combining knowledge management and business process management – a solution for information extraction from business process models focusing on BPM challenges. In: Kő A, Francesconi E, editors. *Electronic Government and the Information Systems Perspective*. Vol. 8650. Springer International Publishing, Munich; 2014:104–117.
22. Chibucos MC, Mungall CJ, Balakrishnan R, et al. Standardized description of scientific evidence using the evidence ontology (ECO). *Database (Oxford)*. 2014;2014:bau075.
23. Wick M, Vatant B, Christophe B. *Geonames Ontology*. 2016. Available from: http://www.geonames.org/ontology. Accessed April 14, 2015.
24. Heckerman D, Geiger D, Chickering DM. Learning Bayesian networks: the combination of knowledge and statistical data. *Mach Learn*. 1995; 20(3):197–243.
25. Mkopi A, Range N, Lwilla F, et al. Adherence to tuberculosis therapy among patients receiving home-based directly observed treatment: evidence from the United Republic of Tanzania. *PLoS One*. 2012;7(12):e51828.
26. Comolet TM, Rakotomalala R, Rajaonarioa H. Factors determining compliance with tuberculosis treatment in an urban environment, Tamatave, Madagascar. *Int J Tuberc Lung Dis*. 1998;2(11):891–897.
27. Méda ZC, Lin Y-T, Sombié I, Maré D, Morisky DE, Chen Y-MA. Medication-adherence predictors among patients with tuberculosis or human immunodeficiency virus infection in Burkina Faso. *J Microbiol Immunol Infect*. 2014;47(3):222–232.
28. Tekle B, Mariam DH, Ali A. Defaulting from DOTS and its determinants in three districts of Arsi Zone in Ethiopia. *Int J Tuberc Lung Dis*. 2002;6(7):573–579.
29. Tadesse T, Demissie M, Berhane Y, Kebede Y, Abebe M. Long distance travelling and financial burdens discourage tuberculosis DOTs treatment initiation and compliance in Ethiopia: a qualitative study. *BMC Public Health*. 2013;13:424.
30. Cramm J, Finkenflügel H, Møller V, Nieboer A. TB treatment initiation and adherence in a South African community influenced more by perceptions than by knowledge of tuberculosis. *BMC Public Health*. 2010;10(1):72.
31. Naidoo P, Peltzer K, Louw J, Matseke G, Mchunu G, Tutshana B. Predictors of tuberculosis (TB) and antiretroviral (ARV) medication non-adherence in public primary care patients in South Africa: a cross sectional study. *BMC Public Health*. 2013;13(1):396.
32. McInerney PA, Nicholas PK, Wantland D, et al. Characteristics of anti-tuberculosis medication adherence in South Africa. *Appl Nurs Res*. 2007;20(4):164–170.
33. Kaona F, Tuba M, Siziya S, Sikaona L. An assessment of factors contributing to treatment adherence and knowledge of TB transmission among patients on TB treatment. *BMC Public Health*. 2004;4(1):68.