



# A Concise Ontology to Support Research on Complex, Multimodal Clinical Reasoning

Sabbir M. Rashid<sup>(✉)</sup> , Jamie McCusker , Daniel Gruen ,  
Oshani Seneviratne , and Deborah L. McGuinness 

Rensselaer Polytechnic Institute, Troy, NY, USA  
{rashis2,mccusj2,gruend2,senevo}@rpi.edu, dlm@cs.rpi.edu

**Abstract.** When clinicians perform tasks involving clinical reasoning, such as the diagnosis or treatment of diabetes, multiple forms of reasoning, including deduction and abduction, are often employed. Ontologies designed to provide a foundation for clinical decision support systems have been encoded based on Clinical Practice Guidelines. Nevertheless, existing approaches solely allow deductive rules for clinical reasoning, with ontologies too large or complex to support tractable abductive reasoning. We follow existing guidelines and standards to design the Diabetes Pharmacology Ontology, a concise ontology – an ontology engineered by adhering to the Minimum Information to Reference an External Ontology Term principle and following an agile design approach. We claim that use cases that incorporate multiple forms of reasoning, such as those aimed at supporting both deduction and abduction, are better supported by concise, rather than complete and comprehensive, ontologies. We demonstrate how Personal Health Knowledge Graphs have been implemented using our ontology and evaluate the abductive capability of modules included with our ontology. We openly publish the resources that have resulted from this work, as listed below. This work demonstrates how multimodal semantic reasoning – deduction and abduction – can be used to emulate tasks involving clinical reasoning and thus has the potential to support practitioners with clinical decision-making.

- **Ontology:** <https://purl.org/twc/dpo/ont/diabetes-pharmacology-ontology.ttl>
- **Bioportal:** <https://bioportal.bioontology.org/ontologies/DPCO>
- **URL:** <https://tetherless-world.github.io/diabetes-pharmacology-ontology>
- **Documentation:** <https://bit.ly/dpo-documentation>
- **GitHub:** <https://github.com/tetherless-world/diabetes-pharmacology-ontology>
- **License:** <https://www.apache.org/licenses/LICENSE-2.0>
- **DOI:** <https://doi.org/10.5281/zenodo.7454721>

**Keywords:** Clinical Practice Guidelines · Standards · Case Studies · Diabetes · Ontology · Pharmacotherapy · Personal Health Knowledge Graph · Reasoning · Deductive Reasoning · Abductive Reasoning

Supported by Rensselaer Polytechnic Institute.

## 1 Introduction

The use of clinical reasoning, a complex process involving cognition, meta-cognition, and discipline-specific knowledge ?, is required for clinicians to complete complex medical tasks involving information comprehension and decision-making. Although the intricacies of how humans perform clinical reasoning is an ongoing topic encompassing much debate ?, it is clear that multiple distinct types of inference are used ?. While the dialectics of logical philosophy is beyond the scope of this article, we find that the Select and Test Model (ST-Model), in which an expert chooses a plausible hypothesis that is subsequently confirmed or falsified through testing, is a practical epistemological framework for our research on medical reasoning ??.

Since the ST-Model presents the cyclic and back-and-forth nature of medical problem-solving tasks in a manner resembling hierarchical decomposition, a divide-and-conquer approach employed by humans for solving large problems ?, we believe that for the purpose of our research, the ST-Model is representative enough of the type of reasoning a clinician may employ. To create a clinical decision support system capable of employing ST-Model-like reasoning, we initially focus on the facilitation of deduction and abduction.

We commence this research by incorporating domain-specific and semantic standards to design a knowledge representation that supports both types of reasoning. We present a guideline and standards-driven approach for crafting a clinical ontology. Given a specific use case for the diagnosis and treatment of type 2 diabetes mellitus, we design the Diabetes Pharmacology Ontology (DPO).

### 1.1 Motivation

While there are multiple factors that come into play when attempting to predict reasoning performance, empirically it has been shown that reasoning over large and complex ontologies can be very time-consuming ?. Despite such concerns, existing diabetes ontologies focus on completeness and encoding as much knowledge as possible ?. These ontologies solely support deduction to accomplish clinical reasoning tasks such as differential diagnosis or therapy planning.

When a clinician reasons over patient information to perform clinical decision-making, multiple forms of reasoning are used, including abstraction, deduction, abduction, and induction ?. Non-monotonic reasoning problems involving abduction are notoriously more challenging to solve than their monotonic counterparts ? and have greater considerations in terms of tractability ????. Therefore, this work is motivated by the need of supporting multiple forms of reasoning to better emulate how clinicians perform clinical reasoning tasks.

### 1.2 Contribution and Claims

The main contribution of this article is that we develop an approach for semantic knowledge representation that supports multimodal reasoning and demonstrate

how the approach can be used for clinical decision-making by publishing a concise, FAIR-compliant ontology, the Diabetes Pharmacology Ontology (DPO), which is used in our approaches for differential diagnosis and therapy planning.

Due to performance and tractability concerns, it becomes important to consider the size and complexity of an ontology that will be used in any setting, which is often more important to consider when it is to be used for multimodal reasoning. We are able to support our use case involving clinical reasoning that incorporates multiple forms of reasoning by creating a concise, rather than complete and comprehensive, ontology. This is achieved by following the Minimum Information to Reference an External Ontology Term (MIREOT) <sup>1</sup> principle when linking to external vocabularies and using an agile <sup>2</sup> design strategy to create both an ontology and a Personal Health Knowledge Graph (PHKG) <sup>3</sup>. An agile approach allows us to meet the needs of our use case while limiting the number of concepts included, thus limiting the size of our ontology. Therefore, we define a concise ontology as one that adheres to both the MIREOT principle as well as an agile design methodology, where the minimum required amount of information for external classes is included and concepts in the ontology are limited to those required by a given use case.

One way we keep our ontology concise is by only including the pharmacotherapy factors and antihyperglycemic treatments found in Table 9.2<sup>1</sup> of Chap. 9 <sup>4</sup> of the ADA Guidelines, omitting other therapies from our ontology. In Sect. ??, we discuss the importance of conciseness as determined from the evaluation of abductive capability presented in Sect. ?? by answering a set of competency questions using the AAA Abox abduction solver <sup>5</sup>. Our ontology and PHKG support both deductive and abductive reasoning and can demonstrate how semantic reasoning can be used to emulate clinical decision-making tasks, including differential diagnosis and therapy planning.

Our ontology is based on existing standards and best practices. We leverage content from the American Diabetes Association (ADA) Clinical Practice Guidelines <sup>6</sup> to inform specific branches of the ontology. Additionally, we consider Ontology Design Patterns (ODPs) <sup>7</sup>, the HCLS dataset specification <sup>8</sup>, and the Data on the Web Best Practices <sup>9</sup> when designing, scoping, and annotating our ontology. The Fast Healthcare Interoperability Resources (FHIR) <sup>10</sup> specification is a standard for representing clinical information that employs a composition and constraint-based modeling approach. In Sect. ??, we demonstrate how a PHKG is created using either FHIR or an upper ontology.

The resources resulting from this research are FAIR <sup>11</sup>, based on guiding principles for the discovery, use, and reuse of data, which we demonstrate by evaluating our ontology on principles for developing computational biomedical knowledge (CBK) infrastructure <sup>12</sup>. The ontology and PHKGs are published and readily available, use standard vocabulary, and are adequately documented.

---

<sup>1</sup> For easy reference, we include this table as supplementary material: [tetherless-world.github.io/diabetes-pharmacology-ontology/#supplementary-material](https://tetherless-world.github.io/diabetes-pharmacology-ontology/#supplementary-material).

## 2 Related Work

### 2.1 Existing Diabetes Ontologies

To support Clinical Decision Support Systems (CDSS) with the diagnosis and treatment of diabetes, an ontological approach for incorporating recommendations from computerized Clinical Practice Guidelines is proposed ?. While a relatively simple flat ontology (as opposed to hierarchical) is created including classes for diagnoses, lab tests, patient information, risk factors, and symptoms, diagnostic and therapy planning rules are written in SWRL ?, allowing for the categorization of patients in terms of both type 1 and type 2 prediabetes and diabetes. The concepts included in this ontology are similar to the top-level classes that we include in our ontology, but we expand upon each concept hierarchically, resulting in a richer representation.

Another ontology including SWRL rules for diabetes diagnosis is the Diabetes Diagnosis Ontology (DDO). This ontology was designed with the goal of systematically developing complete ontologies for the diagnosis of diabetes mellitus ?, resulting in the creation of comprehensive OWL2 diabetes knowledge representation interoperable with OBO Foundry ? ontologies. A major limitation of DDO is the absence of support for therapy planning, addressed by the Diabetes Mellitus Treatment Ontology (DMTO) ?, an extension of DDO.

DMTO is an OWL 2 ontology based on the SHOIQ(D) description logic in which type 2 diabetes treatment plans are modeled based on CPGs which includes SWRL rules for diagnosis and treatment ?. Unfortunately, when trying to run the rules ourselves, we ran into several errors. Furthermore, while DMTO is indeed a comprehensive ontology, resulting in great domain coverage, an ontology of this size is not compatible with timely abductive reasoning. Despite the inclusiveness of DMTO, it does not meet all the requisites to abductively plan treatments for patients, such as direct links from treatments to associated pharmacotherapy factors and appropriately represented concept restrictions to trigger the abductive reasoner used in our approach.

### 2.2 Clinical Practice Guidelines

Clinical Practice Guidelines (CPGs) are collections of statements intended to optimize patient care by assisting the understanding of factors involved in complex medical decision-making, including potential benefits, harms, or alternatives of a specific medical decision, as well as demographic and socioeconomic considerations ?. For a set of guidelines to be trustworthy, they should be developed by experts from various associated disciplines through a systematic review of existing literature, should provide evidence, clear explanations, and quality ratings for recommendations, and should be revised appropriately in light of new findings. For this work, we leverage an existing CPG, the Standards of Medical Care in Diabetes ? by the American Diabetes Association (ADA).

### 3 Ontology Design Approach

We now describe our design methodology for creating DPO, design-related scoping requirements, and our approach to linking external vocabularies.

#### 3.1 Use Case

DPO arose from the need for a structured vocabulary to aid in clinical decision-making. We consider a use case with the goal of supporting clinicians with clinical reasoning tasks, such as differential diagnosis, therapy planning, or plan critiquing. We ground this in the setting where patients have elevated blood glucose levels.

A primary requirement of this use case is that the support provided should emulate the type of rationality a clinician is likely to apply. In particular, a proof of concept is required, where it must be demonstrated that the approach can support multimodal reasoning. We find that by adhering to the ST-Model, we can approximate human clinical decision-making involving multiple forms of reasoning. Inversely, the proof of concept requirement nullifies the requirement that complete coverage is realized. It is not necessary to demonstrate that any scenario in a given domain is addressable, but rather support for such content is possible with the proper extensions.

While complete domain coverage is not necessary, the use case does require that provided recommendations for the portion covered should align with domain-specific standards. As an elevated blood glucose level is an indicator of prediabetes or diabetes, a scoping requirement is that the technology created should center on these medical conditions. In particular, we focus on type 2 diabetes mellitus and leverage the well-used ADA Standards of Care Diabetes CPG. Since clinical decision-making employs multiple forms of reasoning and earlier work has been conducted involving the encoding of CPGs in ontologies, we found that the use of Semantic Web technologies was a natural choice for meeting these initial requirements.

Due to privacy considerations and to avoid HIPAA violations, the use of actual patient data is not a requirement for this use case. Nevertheless, it is required that the data used should resemble actual patient data. We find that a NetCE course includes relevant case studies on diabetes and is a good reference point for constructing hypothetical patients.

#### 3.2 Design Methodology

While an initial use case may be in mind when designing an ontology, it is arguably impossible to know all the future uses the ontology will have. We provide an initial set of concepts in our extensible ontology while remaining open to including additional concepts that may be needed as a consequence of growing use and reuse. Two ontology design approaches that allow for modular updates include the Agile & eXtreme Design (XD) methodologies.

Approaches involving modular updates are often referred to as modular ontology design ?.

When using an agile design approach, simplicity is encouraged, especially when representing rudimentary ideas. Essential features should be implemented foremost, while additional features can be included in the future. Inspired by Agile, the XD methodology asserts that ontologies should only contain concepts and properties that are essential for the particular task for which the ontology is being designed. Extensions of an XD ontology typically transpire iteratively, where further improvements are incorporated by considering the needs of the end-user and involving the customer in the design process.

### 3.3 Design Requirement-Related Scoping

A design requirement is that our ontology must leverage a CPG so that intelligent systems that use the ontology can follow guideline recommendations. We scope our study to clinical cases relevant to type 2 diabetes mellitus. Therefore, the Standards of Medical Care in Diabetes by the American Diabetes Association (ADA) became an apparent choice of a CPG to use. Chapter 9 of the ADA guidelines ?, titled Pharmacologic Approaches to Glycemic Treatment, contains treatment information and factors that can be leveraged for diabetes therapy planning. Therefore, to create a concise rather than complete ontology, we scope the therapies and pharmacotherapy factors included in our ontology to those mentioned in this article, omitting additional diabetes-associated factors and therapies mentioned in the literature. For example, Licogliflozin and Sotagliflozin are SGLT2 inhibitors that we do not include in our ontology. Diagnostic factors and clinical measurements included in the ontology are scoped based on the diabetes-related NetCE case studies. To test the representation capability of the ontology, three of these case studies (patients K, H, and B in ?) are used to create example patients.

### 3.4 Linking External Vocabularies

To make an ontology interoperable with other ontologies, it is necessary to link to external vocabularies. Unfortunately, importing entire ontologies using *owl:import* statements often results in unnecessarily large overhead, especially when the ontologies being imported have a substantial number of concepts, often including many concepts that are not directly applicable. It is not uncommon for an ontology to import other ontologies which are not even used.

The inclusion of *owl:import* statements introduces the risk of vastly increasing the size of an ontology, saturating it with unnecessary information, especially when multiple high-level ontologies are imported or an ontology is imported for the use of a single class or property. Too many classes can have a negative impact on reasoning complexity and computation time, especially with non-monotonic forms of reasoning, such as abduction.

Therefore, for our use case involving multimodal reasoning, it is necessary to minimize the number of external concepts included. The MIREOT guidelines provide specifications for the minimum amount of information required when including concepts from external ontologies. Rather than importing entire ontologies, only necessary concepts should be included. The MIREOT guidelines state that, for an external concept to be consistently referenced, the only required information includes the ontology namespace, the URI for the specific term being imported, and the URI for the superclass of the term being imported ?.

We adopt this approach in terms of directly including external concepts in our ontology as well as the minimal set of information required for reference. When including a superclass of a linked concept, in the cases that (i) the superclass is not directly linked to a term in the ontology, (ii) is not a subclass of another linked concept, or (iii) is a top-level class, we assign the superclass to be a subclass of the concept `dpo:ExternalClass`. By doing so, we reduce the hierarchical clutter by putting all external classes under a single branch.

In addition to the minimal set, we also include the labels of external concepts as also recommended in the MIREOT guidelines, simply to aid in readability. The external concepts are directly linked to internal concepts using *owl:equivalentClass*. For an external concept that includes a definition, we assign the definition to the linked internal concept. For all definitions acquired from external resources, we append a definition source attribution to the end of the definition string that references the external resource, concept, or URL that the definition is adopted from.

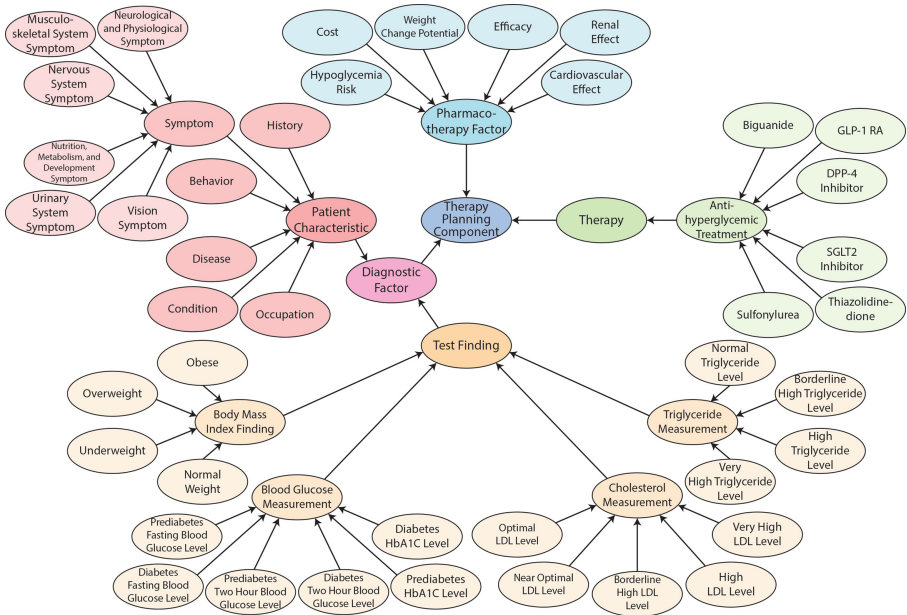
## 4 Diabetes Pharmacology Ontology

To address the needs of our use case, it is necessary to include in our ontology diagnostic factors, such as patient characteristics and test findings, therapies, and pharmacotherapy factors. Shown in Fig. ?? is a simplified representation of DPO, depicting key concepts that extend from the root concept, `dpo:TherapyPlanningComponent`, and some of their subclasses.

While additional classes extend from some of the concepts shown, due to space considerations we omit several branches of the ontology and limit this diagram to extend up to three concepts from the root. The three top-level branches are `dpo:PharmacotherapyFactor`, `dpo:DiagnosticFactor`, and `dpo:Therapy`. All other branches of the ontology extend from these top-level concepts. In addition to the existing concepts included in our ontology, DPO is extensible by allowing the incorporation of new classes as a subclass of an existing concept.

### 4.1 Pharmacotherapy Factor

Organized in the ADA CPG Table 9.2 are type 2 diabetes treatments based on drug-specific factors. Such factors, including efficacy, weight change association, cost, and risk correlations for several medical conditions or diseases, correspond to the columns of Table 9.2. The `dpo:PharmacotherapyFactor` branch is based



**Fig. 1.** DPO is rooted at `dpo:TherapyPlanningComponent`, which has three main branches. These top-level concepts include `dpo:PharmacotherapyFactor`, `dpo:DiagnosticFactor`, and `dpo:Therapy`. Diagnostic factors are split into `dpo:PatientCharacteristic`, shown in light red, and `dpo:Measurement`, shown in light orange. Pharmacotherapy factors are shown in light blue and therapies in light green. (Color figure online)

off of this table, where the above-mentioned columns are used to form the immediate subclasses of `dpo:PharmacotherapyFactor`. In turn, potential values for each entry of a column informed the terms included under each of these concepts.

In addition to the factors mentioned above, Table 9.2 also incorporates a column titled “Additional Considerations”. This column covers additional therapy considerations, such as FDA Black Box warnings ? as well as additional known risks and reactions. To constrain the scope of the initial implementation of DPO, we have not yet included in our ontology the considerations from this column. This extension may be considered as part of future work. Pharmacotherapy factors are mapped to external classes from National Cancer Institute Thesaurus (NCIT) ?. In particular, NCIT is used due to its broad coverage and since most of the concepts included in `dpo:PharmacotherapyFactor` is not included in the other external vocabularies we are linking.

## 4.2 Therapy

The rows of the ADA CPG Table 9.2 comprise of categorizations of therapies commonly used to treat type 2 diabetes. These categorizations are leveraged to inform the top-level therapies we included in our ontology. While there are many



drugs associated with these therapy categorizations, it is not our intention to encode an exhaustive list of therapies. Instead, we limit our scope of therapies to include only those mentioned in Table 9.2. Therapy concepts are mapped to external classes from NCIT, Chemical Entities of Biological Interest (ChEBI) <sup>?</sup>, and Logical Observation Identifiers Names and Codes (LOINC) <sup>?</sup>.

### 4.3 Diagnostic Factor

A NetCE course <sup>?</sup> includes several diabetes-related case studies. To test the application of our work on hypothetical diabetes patients, three case studies included in this course are semantically encoded to create a PHKG for each patient, as described in Sect. ???. These case studies detail the attributes of each patient, including demographic information, treatment history, existing conditions, and symptoms. The case studies also include lab measurements of the patients taken at specific visits. The `dpo:DiagnosticFactor` branch is created based on the NetCE case studies, and is separated into two main branches, `dpo:PatientCharacteristic` and `dpo:TestFinding`.

Concepts in the `dpo:PatientCharacteristic` branch, corresponding to patient attributes, are mapped to external classes from NCIT, HP <sup>?</sup>, and the Symptom ontology<sup>2</sup>. The concepts in the `dpo:TestFinding` branch are based on the NetCE lab measurement, mostly corresponding to glucose and cholesterol-related readings. Test finding concepts are mapped to external classes from NCIT, the Symptom ontology, LOINC, and EFO <sup>?</sup>.

## 5 Personal Health Knowledge Graph

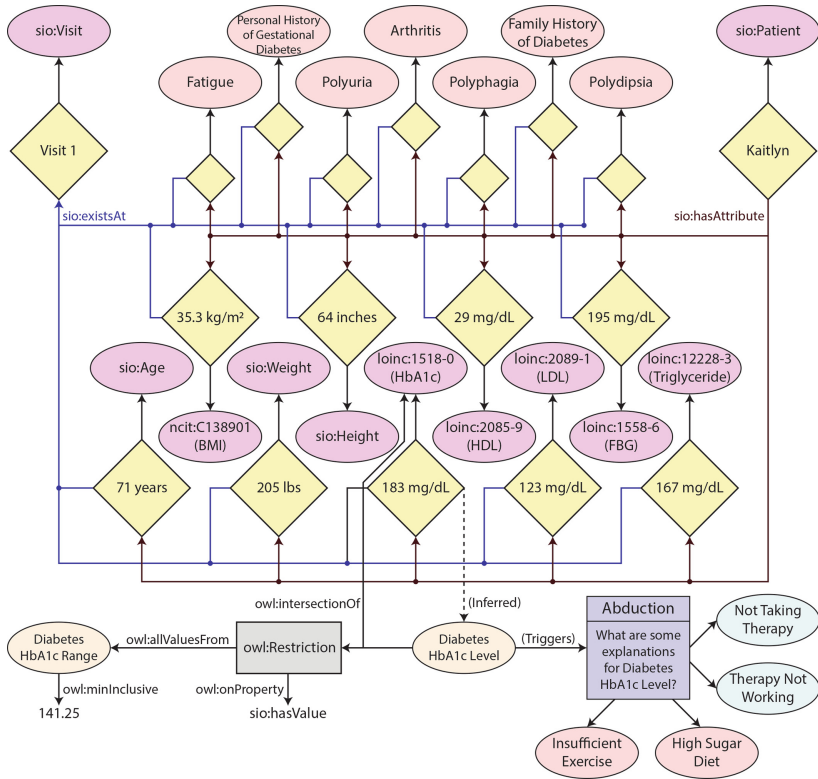
A Personal Health Knowledge Graph (PHKG) is a knowledge resource linking a patient's relevant medical information and personal data, typically for use in personalized health applications <sup>?</sup>. To test the applicability of our ontology, we have created PHKGs for several hypothetical patients based on diabetes-related NetCE case studies. A portion of one such PHKG is depicted in Fig. ??.

In addition to the use of concepts from DPO, our PHKGs leverage an upper ontology, such as the Semanticscience Integrated Ontology (SIO) <sup>?</sup>. We have also created a PHKG based on FHIR <sup>?</sup> rather than SIO for demonstration purposes and to provide support for a healthcare-related standard specification. Nevertheless, we find that the use of SIO results in a much more straightforward and concise representation. Both representation formats are included in our GitHub repository<sup>3</sup>. The file titled `maria.ttl` is the FHIR representation while the other files leverage SIO.

When using SIO, the patient is encoded as an instance of `sio:Patient` that has instances of attributes linked using `sio:hasAttribute`. Each attribute exists at an instance of a timepoint, linked using `sio:existsAt`, corresponding to the

<sup>2</sup> <https://bioportal.bioontology.org/ontologies/SYMP/>.

<sup>3</sup> <https://github.com/tetherless-world/diabetes-pharmacology-ontology/tree/main/kb>.



**Fig. 2.** A Personal Health Knowledge Graph includes instances of patients. A top-level ontology like SIO along with concepts in DPO is used for representing patient attributes. Deductive rules from the ontology are used to infer attribute categorizations. Ontology classes are shown as ellipses. External classes are pink while internal classes are color-coded as described in Fig. ?? . Instances are shown as yellow diamonds. (Color figure online)

visit at which an attribute was recorded. Patient attributes include symptoms, conditions, demographic information, and measurement values. Deductive rules in the ontology can be used to make inferences regarding the recorded attributes.

One such rule is depicted towards the bottom of Fig. ?? , where a hemoglobin A1C (HbA1c) measurement is categorized to be in the diabetes range. Such categorizations can be used to trigger abductive hypotheses. One may ask the question, “What are some explanations for why the patient’s HbA1c is in the diabetic range?” The encoding of this rule is discussed in the following section.

## 6 Application: Supporting Abductive Reasoning

The abductive reasoning problem arises when a knowledge base  $\mathcal{K}$  does not entail an observation  $\mathcal{O}$ , that is  $\mathcal{K} \not\models \mathcal{O}$ , resulting in the search for an explanation  $\mathcal{E}$  that when included with  $\mathcal{K}$  would result in the entailment of the observation,

$\mathcal{K} \cup \mathcal{E} \models \mathcal{O}$ ?. For a sufficient solution, several constraints need to be considered, including consistency ( $\mathcal{K} \cup \mathcal{E} \not\models \perp$ ), minimality ( $\mathcal{E}$  is a ‘*minimal explanation*’ for  $\mathcal{O}$ ), relevance ( $\mathcal{E} \not\models \mathcal{O}$ ), and explanatoriness ( $\mathcal{K} \not\models \mathcal{O}, \mathcal{E} \not\models \mathcal{O}$ )?.

An existing approach demonstrates how abduction can be utilized to address a use case related to the diagnosis of diabetes mellitus-related conditions?. While this work presents how rules based on symptoms can be formed to abductively arrive at diagnoses, the described ontology is not implemented. Nevertheless, this work provides examples of abductive rule representation that have influenced our approach. In fact, we leverage the AAA ABox abduction solver by the same authors? as the abductive inference engine we employ.

Simple abductive explanations arise through subsumption. To entail the observation that an instance is of type a class, the instance could be one of the subclasses of that class. Therefore, an approach for representing possible abductive explanations is by assigning possible hypotheses as a union of classes. For example, we can define `dpo:DiabetesHbA1cLevel` to be equivalent to the union of classes that serve as potential explanations, as listed below.

```
dpo:DiabetesHbA1cLevel rdf:type owl:Class ;
  owl:equivalentClass
    [ rdf:type owl:Class ;
      owl:unionOf (
        dpo:TherapyNotEffective
        dpo:NotComplyingWithTherapy
        dpo:InsufficientExercise
        dpo:HighSugarDiet ) ] ;
  rdfs:subClassOf dpo:Measurement .
```

As part of our resources, we include in our GitHub repository<sup>4</sup> modules written in RDF on which to apply abductive queries. Due to the tractability limitations associated with abductive reasoning?, we recommend the use of specific modules for targeted abductive queries rather than running the reasoner over the entire ontology. Nevertheless, if time and computational capability are not a deterrent, the modules can be combined with the ontology and/or additional knowledge to result in a more comprehensive set of explanations.

Included is a module for an example scenario involving unexpected weight gain, as well as modules related to the use case discussed in this article, such as a diagnostic module, HbA1c module, and therapy module. The therapy module represents specific antihyperglycemic medications prescribed based on pharmacotherapy factor considerations. We further include abductive queries for these modules in the form of commands<sup>5</sup> that can be sent to the abduction solver.

<sup>4</sup> <https://github.com/tetherless-world/diabetes-pharmacology-ontology/tree/main/ont/modules>.

<sup>5</sup> [https://github.com/tetherless-world/diabetes-pharmacology-ontology/blob/main/abduction\\_commands](https://github.com/tetherless-world/diabetes-pharmacology-ontology/blob/main/abduction_commands).

## 7 Evaluation

A recent article [?](#) promotes three guiding principles for developing computational biomedical knowledge (CBK) and its associated infrastructure; CBK should be FAIR, trustworthy, and open. We use these principles to evaluate the DPO, which we include in the Supplementary Materials section of our website.<sup>6</sup>

In addition to evaluating how our ontology met the principles for developing CBK, we evaluate the abductive reasoning support of our approach. To do so, we compile a set of competency questions that we test whether we can answer using abduction. These competency questions are listed as follows.

- $C_1$  What are some causes for an HbA1c level in the diabetes range?
- $C_2$  How can we explain the patient having insufficient exercise?
- $C_3$  Why might the patient not be taking a medication?
- $C_4$  What therapies have high efficacy?
- $C_5$  What therapies have potential for weight loss?
- $C_6$  What therapies have potential atherosclerotic cardiovascular disease benefit?

The first 3 questions are used to test diagnostic capability while the latter 3 relate to therapy planning. To test these competency questions, a Dell laptop with a Ubuntu 20.04.5 LTS operating system is used. The computer has an Intel Core i7-8550U CPU running at 1.8 GHz with 4 cores and 8 threads. The laptop contains 16 GB of RAM. Version 0.11 of the AAA Abox abduction solver<sup>7</sup> is used, with the negations and avoid loops parameters set to true, the abduction approach set to reduction, and the timeout set to 120s.

**Table 1.** Modules used to evaluate competency questions. The columns correspond to ontology details returned by the AAA solver. “Concepts” refers to the number of classes in the module. “Roles” refers to the number of properties. “Individuals” refers to the number of instances. “Tbox” refers to the number of terminological axioms. “Abox” refers to the number of individual-associated assertions.

Module	Concepts	Roles	Individuals	TBox	ABox
hba1c	13	1	7	18	7
exercise	12	1	2	21	2
compliance	18	1	2	28	2
therapy	117	1	81	230	115

We use 4 modules with a varying number of concepts and individuals, as shown in Table ???. A unique module is used for each diagnostic question, while a single therapy module is used for the therapy-related questions. Abducibles

<sup>6</sup> <https://tetherless-world.github.io/diabetes-pharmacology-ontology/#supplementary-material>.

<sup>7</sup> Available here: <https://dai.fmph.uniba.sk/~pukancova/aaa/>.

are particular concepts, roles, or individuals that are provided to the reasoner to limit the reasoning space. Given the complete set of relevant abducibles, an abductive query can be run without providing any abducibles, by providing all of the relevant abducibles, or by segmenting the abducibles and running multiple queries using each of the segments. When creating segments, we keep necessary abducibles in all of the segments, such as those appearing in the observation, and split the rest. For this evaluation, we create at most 3 segments per query.

The depth of an explanation corresponds to the number of assertions that can appear in an explanation. We test each query with depths of 1, 2, and 3. For a depth  $n$ , we record the time  $t_n$  in seconds that it takes to find the set of explanations at that depth. If the set is not found within the time limit, which we have set to 2 min, the query times out, denoted using T/O. We also record the number of explanations found. The results of this evaluation are shown in Table ???. The full set of evaluation output logs are available on our GitHub.<sup>8</sup>

## 8 Discussion

We briefly discuss the results of the evaluation, the impact of this work, some limitations, and future research directions.

### 8.1 Results

We have evaluated the capability of competency questions to be answered abductively using modules included with our ontology. While the therapy-related queries did time out when specifying a depth of 3, upon examining the explanations provided, we found that for all of the modules, the set of expected explanations could be obtained using a depth of 2. Depending on the representation of the module, the expected explanations could even be reached using a depth of 1. Therefore, we have demonstrated that our approach allows us to abductively answer competency questions related to diagnosis and therapy planning.

These results demonstrate the importance of conciseness when performing abductive reasoning. Modules with fewer concepts allowed for abductive answers to be obtained faster and to a greater depth. We also found that greater depth calculations could be obtained by dividing the set of abducibles into segments. However, this requires knowledge of the set of relevant abducibles. Nevertheless, for simpler modules, such as the diagnostic modules, we can obtain the expected results without specifying abducibles, again showing the value of conciseness.

### 8.2 Scientific Impact

We have published an original ontology as well as PHKGs that exemplify how the ontology can be used with an upper ontology or as a FHIR adherent representation. We have introduced our resources in this article and have included further

<sup>8</sup> [https://github.com/tetherless-world/diabetes-pharmacology-ontology/tree/main/evaluation\\_results](https://github.com/tetherless-world/diabetes-pharmacology-ontology/tree/main/evaluation_results).

**Table 2.** Abductive competency query results. For multiple segments, the number of abducibles per segment is shown as an average, and the computation times and number of unique explanations are shown as sums. \*The second segment with 1 less abducible actually did finish computing after 109.48s, but the first segment timed out.

CQ	#Segments	#Abducibles/Segment	Depth	$t_1$ (s)	$t_2$ (s)	$t_3$ (s)	#Exp.	Module		
$C_1$	0	$\infty$	1	1.01	N/A	N/A	4	hba1c		
			2	1.01	14.2	N/A	9			
			3	1.14	14.65	T/O	9			
	1	18	1	<b>0.91</b>	N/A	N/A	4			
			2	1.04	11.52	N/A	9			
			3	0.92	11.67	T/O	9			
	2	12	3	1.44	<b>6.04</b>	<b>49.04</b>	9			
	$C_2$	0	$\infty$	1	0.39	N/A	N/A		3	exercise
				2	0.38	0.66	N/A		3	
3				0.39	0.66	1.33	3			
1		9	1	0.38	N/A	N/A	3			
			2	<b>0.36</b>	<b>0.45</b>	N/A	3			
			3	0.39	0.49	<b>0.49</b>	3			
$C_3$		0	$\infty$	1	0.49	N/A	N/A	6	compliance	
				2	0.46	0.93	N/A	6		
				3	0.58	0.94	2.23	6		
	1	11	1	0.48	N/A	N/A	6			
			2	<b>0.38</b>	<b>0.48</b>	N/A	6			
			3	0.57	0.57	<b>0.57</b>	6			
	$C_4$	0	$\infty$	1	T/O	N/A	N/A	0		therapy
		1	26	1	8.26	N/A	N/A	14		
				2	8.47	T/O	N/A	14		
2		13.5	2	7.64	T/O*	N/A	14			
3		9.33	2	7.95	49.52	N/A	14			
			3	<b>7.4</b>	<b>48.75</b>	T/O	14			
$C_5$		0	$\infty$	1	T/O	N/A	N/A	1	therapy	
	1	26	1	7.76	N/A	N/A	11			
			2	7.25	T/O	N/A	12			
	2	13.5	2	6.86	T/O	N/A	12			
	3	9.33	2	<b>6.59</b>	<b>48.42</b>	N/A	11			
			3	7.07	48.95	T/O	12			
	$C_6$	0	$\infty$	1	T/O	N/A	N/A	0		therapy
1		26	1	6.96	N/A	N/A	7			
			2	6.75	T/O	N/A	8			
2		13.5	2	<b>5.68</b>	T/O	N/A	9			
3		9.33	2	6.1	51.53	N/A	7			
			3	6.07	<b>50.51</b>	T/O	12			

descriptions on a dedicated website. While other diabetes ontologies based on CPGs do exist, our work is innovative and advances the state-of-the-art in that we consider both deductive and abductive capabilities in its design. Unlike other diabetes ontologies, we focus on the creation of a concise rather than comprehensive ontology to better support tractable abductive reasoning. Our contributions include not just the shared resources, but also the approach that may be leveraged in other work. While we have not yet validated our technique using actual patient data, we do apply our approach to hypothetical diabetes case studies.

We apply our approach to the diagnosis and treatment of diabetes, a prevalent health problem affecting 37.3 million people in the United States (11.3% of the U.S. population), with an additional 96 million people aged 18 years or older (38.0% of the adult U.S. population) affected with prediabetes ?. Our approach to designing the ontology and representing rules that can trigger abductive queries can also be generalized to other medical conditions. Therefore, this work and the resulting resources can potentially aid clinicians with a wide range of clinical decision-making tasks. Our resources and approach are of interest to the Semantic Web community since they illustrate how multimodal reasoning can be implemented for a practical application.

### 8.3 Limitations and Future Directions

Due to scoping considerations, only a subset of diabetes information is encoded in our ontology. This limitation is justified by the need for our approach to allow for abductive reasoning and the requirement of our use case for only a proof of concept rather than complete coverage. Nevertheless, a future extension of this work includes incorporating more knowledge. Another limitation is that for many queries to run quickly, such as those related to therapy planning, abducibles need to be explicitly specified to constrain the Abox abduction solver to only return explanations involving the expected concepts. This constraint reduces computation time but requires preexisting knowledge of the expected results.

We plan to encode additional components from the ADA guidelines. Moreover, we wish to review relevant literature to find further pharmacotherapy factor associations and allow support for more antiglycemic therapies. Since we test our approach using data based on diabetes-related case studies, one limitation of our PHKGs is that they are based on hypothetical rather than real patient data. Therefore, potential future research involves the validation of our approach using actual patient data.

## 9 Conclusion

We have introduced DPO, which we have used as the vocabulary and knowledge representation resource for our approach to supporting multimodal clinical reasoning for the diagnosis and treatment of diabetes. Unlike earlier ontologies that focus on comprehensiveness, we instead design a concise ontology able to support multimodal reasoning. We have presented and evaluated our approach, and have discussed the impact of this research, its limitations, and future work.

**Acknowledgements.** This work is supported by IBM Research AI through the AI Horizons Network. We would like to acknowledge Amar K. Das for his guidance. We also thank the members of the Tetherless World Constellation, especially Sola Shirai, Shruthi Chari, and Danielle Villa for their feedback.

## References

- Alharbi, R.F., Berri, J., El-Masri, S.: Ontology based clinical decision support system for diabetes diagnostic. In: 2015 Science and Information Conference (SAI), pp. 597–602. IEEE (2015)
- Arocha, J.F., Wang, D., Patel, V.L.: Identifying reasoning strategies in medical decision making: a methodological guide. *J. Biomed. Inform.* **38**(2), 154–171 (2005)
- Association, A.D.: Introduction: standards of medical care in diabetes-2022 (2022)
- Bender, D., Sartipi, K.: HL7 FHIR: an agile and restful approach to healthcare information exchange. In: Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems, pp. 326–331. IEEE (2013)
- Blomqvist, E., Hammar, K., Presutti, V.: Engineering ontologies with patterns-the extreme design methodology (2016)
- Centers for Disease Control and Prevention: National diabetes statistics report (2022). <https://www.cdc.gov/diabetes/data/statistics-report/index.html>
- Committee, A.D.A.P.P., Committee: A.D.A.P.P.: 9. Pharmacologic approaches to glyemic treatment: Standards of medical care in diabetes-2022. *Diab. Care* **45**(Supplement\_1), S125–S143 (2022)
- Courtot, M., et al.: MIREOT: the minimum information to reference an external ontology term. Aug-2009 (2009)
- Degtyarenko, K., et al.: ChEBI: a database and ontology for chemical entities of biological interest. *Nucleic Acids Res.* **36**(suppl\_1), D344–D350 (2007)
- Delong, C., Preuss, C.V.: Black box warning (2022)
- of Medicine (US). Committee on Standards for Developing Trustworthy Clinical Practice Guidelines, I., Graham, R., Mancher, M., Miller Wolman, D.: Clinical practice guidelines we can trust. National Academies Press Washington, DC (2011)
- Du, J., Qi, G., Shen, Y.D., Pan, J.Z.: Towards practical ABox abduction in large OWL DL ontologies. In: Twenty-Fifth AAAI Conference on Artificial Intelligence (2011)
- Du, J., Wang, K., Shen, Y.D.: A tractable approach to ABox abduction over description logic ontologies. In: Twenty-Eighth AAAI Conference on Artificial Intelligence (2014)
- Dumontier, M., et al.: The semanticscience integrated ontology (SIO) for biomedical research and knowledge discovery. *J. Biomed. Semant.* **5**(1), 1–11 (2014)
- El-Sappagh, S., Ali, F.: DDO: a diabetes mellitus diagnosis ontology. *Appl. Inform.* **3**, 1–28 (2016)
- El-Sappagh, S., Kwak, D., Ali, F., Kwak, K.S.: DMTO: a realistic ontology for standard diabetes mellitus treatment. *J. Biomed. Semant.* **9**(1), 1–30 (2018)
- Elsenbroich, C., Kutz, O., Sattler, U.: A case for abductive reasoning over ontologies. In: OWLED, vol. 216 (2006)
- Gangemi, A.: Ontology design patterns for semantic web content. In: Gil, Y., Motta, E., Benjamins, V.R., Musen, M.A. (eds.) ISWC 2005. LNCS, vol. 3729, pp. 262–276. Springer, Heidelberg (2005). [https://doi.org/10.1007/11574620\\_21](https://doi.org/10.1007/11574620_21)
- Gray, A.J., Baran, J., Marshall, M.S., Dumontier, M.: Dataset descriptions: HCLS community profile. Interest group note, W3C (2015). <http://www.w3.org/TR/hcls-dataset>



- Gyrard, A., Gaur, M., Shekarpour, S., Thirunarayan, K., Sheth, A.: Personalized health knowledge graph. In: *CEUR Workshop Proceedings*, vol. 2317. NIH Public Access (2018)
- Hammar, K., Hitzler, P., Krisnadhi, A.: *Advances in Ontology Design and Patterns*, vol. 32. IOS Press (2017)
- Horrocks, I., et al.: SWRL: a semantic web rule language combining OWL and RuleML. *W3C Member Submission* **21**(79), 1–31 (2004)
- Hubauer, T., Lamparter, S., Pirker, M.: Automata-based abduction for tractable diagnosis. In: *International Workshop on Description Logics*, pp. 360–371 (2010)
- Hunt, A., Thomas, D.: The trip-packing dilemma [agile software development]. *IEEE Softw.* **20**(3), 106–107 (2003)
- Islam, R., Weir, C.R., Jones, M., Del Fiol, G., Samore, M.H.: Understanding complex clinical reasoning in infectious diseases for improving clinical decision support design. *BMC Med. Inform. Decis. Mak.* **15**(1), 1–12 (2015)
- Kang, Y.B., Pan, J.Z., Krishnaswamy, S., Sawangphol, W., Li, Y.F.: How long will it take? Accurate prediction of ontology reasoning performance. In: *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 28 (2014)
- Kindler, H., Fischer, B., Densow, D., Fliedner, T.M.: An architecture for medical knowledge-based assistance systems. In: *Proceedings IEEE Symposium and Workshop on Engineering of Computer-Based Systems*, pp. 442–449. IEEE (1996)
- Kumar, A., Smith, B.: Oncology ontology in the NCI thesaurus. In: Miksch, S., Hunter, J., Keravnou, E.T. (eds.) *AIME 2005. LNCS (LNAI)*, vol. 3581, pp. 213–220. Springer, Heidelberg (2005). [https://doi.org/10.1007/11527770\\_30](https://doi.org/10.1007/11527770_30)
- Lóscio, B.F., Burle, C., Calegario, N.: Data on the web best practices. *W3C recommendation* (2017). <https://www.w3.org/TR/dwbp/>
- Loughlin, M., Copeland, S.M.: Humans, machines and decisions: clinical reasoning in the age of artificial intelligence, evidence-based medicine and COVID-19. *J. Eval. Clin. Pract.* **27**(3), 475 (2021)
- Malone, J., et al.: Modeling sample variables with an experimental factor ontology. *Bioinformatics* **26**(8), 1112–1118 (2010)
- McCusker, J., et al.: Guiding principles for technical infrastructure to support computable biomedical knowledge. *Learn. Health Syst.* **n/a**(n/a), e10352 (2022). <https://doi.org/10.1002/lrh2.10352>, <https://onlinelibrary.wiley.com/doi/abs/10.1002/lrh2.10352>
- McDonald, C.J., et al.: LOINC, a universal standard for identifying laboratory observations: a 5-year update. *Clin. Chem.* **49**(4), 624–633 (2003)
- NetCE: Diabetes pharmacology (2022). <https://www.netce.com/coursecontent.php?courseid=2488>
- Presutti, V., Daga, E., Gangemi, A., Blomqvist, E.: eXtreme design with content ontology design patterns. In: *Proceedings of Workshop on Ontology Patterns* (2009)
- Pukancová, J., Homola, M.: Abductive reasoning with description logics: Use case in medical diagnosis. In: *Description Logics* (2015)
- Pukancová, J., Homola, M.: Tableau-based ABox abduction for the ALCHO description logic. In: *Description Logics* (2017)
- Pukancová, J., Homola, M.: The AAA ABox abduction solver. *KI-Künstliche Intell.* **34**(4), 517–522 (2020)
- Ramoni, M., Stefanelli, M., Magnani, L., Barosi, G.: An epistemological framework for medical knowledge-based systems. *IEEE Trans. Syst. Man Cybern.* **22**(6), 1361–1375 (1992)
- Reyes-Cabello, A.L., Aliseda-Llera, A., Nepomuceno-Fernández, Á.: Towards abductive reasoning in first-order logic. *Log. J. IGPL* **14**(2), 287–304 (2006)

- Robinson, P.N., Mundlos, S.: The human phenotype ontology. *Clin. Genet.* **77**(6), 525–534 (2010)
- Schumaker, E.: What is a HIPAA violation? ABC News (2021)
- Smith, B., et al.: The obo foundry: coordinated evolution of ontologies to support biomedical data integration. *Nat. Biotechnol.* **25**(11), 1251–1255 (2007)
- Stefanelli, M., Ramoni, M.: Epistemological constraints on medical knowledge-based systems. In: Evans, D.A., Patel, V.L. (eds.) *Advanced Models of Cognition for Medical Training and Practice*. NATO ASI Series, vol. 97, pp. 3–20. Springer, Heidelberg (1992). [https://doi.org/10.1007/978-3-662-02833-9\\_1](https://doi.org/10.1007/978-3-662-02833-9_1)
- Wilkinson, M.D., et al.: The fair guiding principles for scientific data management and stewardship. *Sci. Data* **3**, 1–9 (2016)