

SNOMED Clinical Terms® UK Clinical Edition SNOMED CT Query Table

A SNOMED CT data derivative

Reduces query execution false negatives occurring due to concept inactivation



Information and technology
for better health and care

Contents

1	Introduction	5
1.1	Document Scope	5
1.2	Product Scope	5
1.3	Audience	5
1.4	Product Status	5
1.5	Product History	6
1.6	Enquiries	6
2	Background	7
2.1	Concept inactivation in SNOMED CT	7
2.2	Effect of concept inactivation on reporting	8
2.3	SNOMED CT historical relationships	8
2.3.1	UK Substitutions Table	9
2.3.2	SNOMED Query Table	9
2.4	Related SNOMED CT Content Issues	10
2.4.1	Navigational concepts	10
2.4.2	Situation with explicit context	11
3	Summary of Use Cases	14
3.1	EPR Instance Data contains inactive conceptId	14
3.2	Query contains inactive conceptId	14
3.3	Inactive conceptIds in Query and EPR instance data	14
4	Design and Content Optimisations	15
5	Limitations	16
6	File format and usage	19
6.1	Column Details	19
6.2	Recommendations for usage	21

Document Management

Revision History

Version	Date	Summary of Changes
0.1	18-Nov-2013	First draft
0.9	21-Mar-2014	Review draft
1.0	01-Apr-2014	First release
1.01	01-Oct-2014	Second release
1.02	01-Apr-2015	Third release
1.04	01-Oct-2015	Fourth release
1.05	01-Oct-2016	Sixth release – reformat to NHS Digital document template
1.06	01-Apr-2017	Seventh release – MySQL demonstrator added
1.07	01-Oct-2017	Eighth release – minor clarifications of text

Future Revisions Due

Version	Date	Summary of Changes

Reviewers

This document must be reviewed by the following people:

Reviewer name	Title / Responsibility	Date	Version
Denise Downs		21-Mar-2014	

Approved by

This document must be approved by the following people:

Name	Signature	Title	Date	Version

Glossary of Terms

Term / Abbreviation	What it stands for
HSCIC	Health and Social Care Information Centre (NHS Digital)
IHTSDO	International Health Terminology Standards Development Organisation. Organisation responsible for the management of SNOMED CT.
SNOMED CT	SNOMED CT ¹ (S ystematized N omenclature of M edicine C linical T erms)

Related Documents

Document Title
UK Concept history substitutions table doc_UKSnomedCTHistorySubstitutionTable_Current-en-GB_GB1000000_20181001
SNOMED CT: Inactive Content Technical Report http://systems.hscic.gov.uk/data/uktc/snomed/training/inactcon1202.pdf

Document Control:

The controlled copy of this document is maintained in the HSCIC corporate network. Any copies of this document held outside of that area, in whatever format (e.g. paper, email attachment), are considered to have passed out of control and should be checked for currency and validity.

¹ SNOMED® and SNOMED CT® are registered trademarks of the IHTSDO® (www.ihtsdo.org)

1 Introduction

This document describes the SNOMED CT Query Table, a computed derivative of the UK Clinical Edition of SNOMED CT. The product aims to facilitate querying over SNOMED CT data where either or both the data being queried over, or the query specifications themselves, contain SNOMED CT concept identifiers with 'inactive' status. The use cases for the product therefore overlap with those of the existing SNOMED CT History Substitution Table product, which is also one of the inputs to the computation used to produce the Query Table. The technical report '[SNOMED CT: Inactive Content](#)' provides further detail on the issue and use cases.

This SNOMED CT Query Table builds on and supercedes earlier draft designs and tooling released by NHS Digital, including the SnoGraph tool.

1.1 Document Scope

This document provides an outline of:

- the rationale for creating the new product
- the intended use cases
- the method of construction and known content issues

The document does not discuss clinical governance, safety or acceptance testing issues.

1.2 Product Scope

The Query Table product addresses 'false negative' reporting issues only in relation to concepts found in the SNOMED Clinical Terms® International and UK Clinical Editions.

It does **not** currently cover similar issues relating to concepts from SNOMED Clinical Terms® UK Drug Edition.

1.3 Audience

Suppliers and clinical users of clinical information systems already based upon, or that process SNOMED CT or that are planning future systems with these capabilities.

Secondary users of data derived from SNOMED CT systems, including large data repositories, health networks, research units, research networks, CSUs and Commissioning Groups, Public Health Observatories

NHS programmes and organisations contracting or commissioning related activity involving either suppliers or users listed above.

1.4 Product Status

From October 2018, This document plus the table it describes constitute a product at the 'Supported Product' stage as defined within the Product Lifecycle².

² <http://systems.hscic.gov.uk/data/uktc/snomed/governance/lifecycle.pdf>

'Supported Product' status means:

1. Both the release format specification of the product and the method of its content preparation shall remain fixed indefinitely *unless* a significant safety risk is identified that cannot be mitigated without changing them. Where changes are deemed necessary to improve a product then a formal consultation procedure will be undertaken which may include some or all parts of the product development process and may include an option for parallel running (i.e. support for both existing and new specification)
2. NHS Digital commits to continue to support, maintain and publish the product against that fixed specification indefinitely, subject to the considerations above or proper product termination
3. Quality assurance may be ongoing but the product is approved for deployment in live clinical systems, subject to standard safety assessment procedures associated with deployment of any product into a live environment
4. The commitment to release against a stable specification does not preclude continued parallel evolution of the specification and consequent development of improved variants which may or may not be considered as new products

1.5 Product History

Approval to prepare and release the SNOMED CT Query Table as a draft product was obtained from the Terminology Edition Committee on December 12th 2013. The first release occurred on April 1st 2014, as a Technology Preview and the committee approved the product as a Supported Product on 1st October 2018.

The algorithm used to generate the table was improved at the third (April 2015) release to include measures to similarly mitigate the risks of false reporting in respect of descendants of 363743006|Navigational concept| and Situations with the default context. The resulting table was as a consequence significantly larger than for the earlier releases.

A further improvement was made at the April 2017 release, and a sample implementation/demonstrator also added.

The additional mitigations added from the April 2015 release (in respect of navigational and default context content) are removed from the October 2017 release, pending detailed clinical safety review in respect of what are necessarily currently only partial and thus somewhat unpredictable mitigations for the risk of 'false negative' reporting.

1.6 Enquiries

Content error reports, or requests for clarification or additional functionality, should be made to information.standards@hscic.gov.uk with 'SNOMED CT Query Table' in the subject line.

2 Background

2.1 Concept inactivation in SNOMED CT

SNOMED CT, like Clinical Terms Version 3, is a dynamic terminology: not only can new content be added, but existing content may be edited or made 'inactive' where necessary. SNOMED CT partitions all its concepts into those that are '**Active**' and those that are '**Inactive**'. In RF2, this initial binary partitioning is achieved through the *active* field in the *sct2_Concept* table; RF1 releases employ a *ConceptStatus* metadata value on each concept within the *sct1_Concept* table.

'**Active**' currently further subdivides into 'current' or 'pending move', whilst '**Inactive**' has seven subdivisions: 'retired (no reason)', 'duplicate', 'erroneous', 'ambiguous', 'outdated', 'limited', and 'moved'. In RF2, this further partitioning is represented through the membership of the 900000000000489007|Concept inactivation indicator attribute value reference set (foundation metadata concept). In RF1, different *ConceptStatus* values appear in the *sct1_Concept* table; whether the concept is fundamentally 'Active' or 'Inactive' is never explicitly recorded in RF1 but rather only implied through this *ConceptStatus* value.

Refer to the SNOMED CT Technical Implementation Guide³ for detail on concept status.

In this document concepts are referred to as either *active* or *inactive*.

Although 'retired' is often synonymous in common usage with 'inactive', this document treats 'retired' as a reserved word for use only to refer to the concept status value '1' in the SNOMED CT standard (RF1).

In addition to changing their *ConceptStatus*, inactive concepts also lose all modelled attributes on inactivation. In particular, this includes losing any *IS_A* relationships that previously positioned the concept within SNOMED CT's primary taxonomic hierarchy; inactive concepts exist instead in a separate, very flat, taxonomic tree.

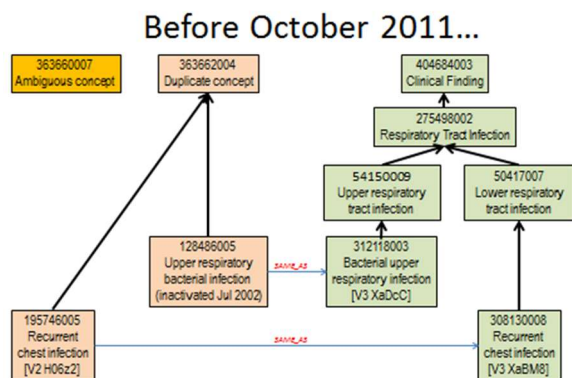


Figure 1: Position of 308130008 prior to October 2011

For example, until October 2011, [308130008|Recurrent chest infection](#) was an active concept in the main taxonomy, classified below [275498002|Respiratory tract infection](#) as shown in Figure 1. Note that [195746005|Recurrent chest infection](#) also existed, but was already inactive (being a *duplicate* of the active [308130008](#)) and, accordingly, was **not** classified as a hierarchy descendent of [275498002|Respiratory tract infection](#), although SNOMED CT content *does* state that the two are related in some other, non-hierarchical way (blue horizontal lines).

³ Publically available from the IHTSDO Document Library http://ihtsdo.org/fileadmin/user_upload/doc/

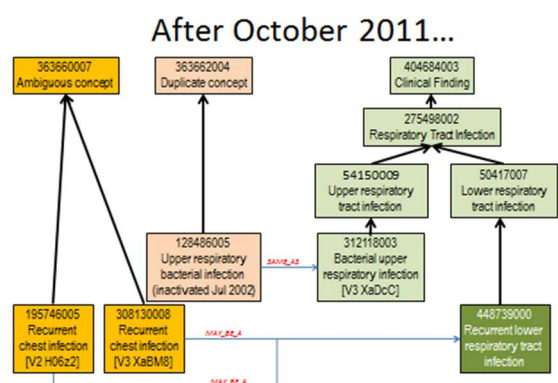


Figure 2 : Position of 308130008 after October 2011

From October 2011, [308130008|Recurrent chest infection](#) also became inactive: ‘chest’ often means the entire thorax whilst, in use, ‘chest infection’ *usually* means more specifically the lungs. Both [308130008](#) and [195746005](#) are therefore now (re-)classified as ‘ambiguous’, and [308130008](#) also moved outside the main taxonomy: neither is a subtype of [275498002|Respiratory tract infection](#). Again, however, SNOMED CT content *does* state that some other non-hierarchical type of relationships exist (blue horizontal lines).

2.2 Effect of concept inactivation on reporting

The movement of inactive concepts out of the main taxonomy means that, if SNOMED CT has been deployed into systems and regularly updated with new release data, and data has been recorded using codes like [308130008](#) that have since become inactive, then data processing will behave predictably but inappropriately in retrieval or reporting contexts, unless systems implement an appropriate mitigation.

Where query specifications have been written to exploit only SNOMED’s main taxonomy (the vertical black arrows in Figures 1 and 2) to dynamically compute the codes that should be in the result set each time a query is run then, in most cases, such inactive data will not be returned by otherwise appropriate query specifications. This will occur even though the same coded content was previously correctly returned when the same query was run against an earlier version of SNOMED CT. Clinically, concept inactivation will therefore manifest as new ‘false negatives’: patients who are no longer returned by a particular query but where the clinical expectation is that they should be.

The SNOMED CT Query Table is intended to be an appropriate and sufficient mitigation.

2.3 SNOMED CT historical relationships

The Query Table is constructed as a computed derivative of other information about inactive concepts that is represented within the SNOMED CT release data. In addition to demonstrating how inactive concepts move out of the main taxonomy, Figure 1 and Figure 2 (above) also show how *most* inactive concepts are linked to alternative active concepts, within the main taxonomy, by means of non-taxonomic ‘historical relationships’. In the figures, these links are represented by the horizontal blue lines. In Figure 1, for example, the inactive [195746005](#) has a SAME_AS relationship to [308130008](#), indicating that the two concepts are semantically equivalent. But only one of the pair sits within the traditional retrieval taxonomy.

2.3.1 UK Substitutions Table

By postprocessing all such historical relationships it is possible to construct a ‘substitutions table’ listing all currently inactive concept identifiers and, for each, all possible substitutions by an active concept (*provided* such substitutions exist – see section 5). NHS Digital therefore provides a pre-computed Substitutions Table for the UK Edition in the same Subpack as the Query Table itself.

2.3.2 SNOMED Query Table

The Substitutions Table *could* be used as part of an update process to be performed at each new release of SNOMED CT: all historically captured data, all queries and all other design artefacts involving references to SNOMED CT concepts might be examined for occurrences of newly inactivated concepts and any found are ‘annotated’⁴ with their designated active substitute concept listed in the Substitutions Table, to be used for the purposes of future retrieval. However, this search-and-replace operation can be computationally prohibitive to perform in live systems, assuming that it is even known where all design artefacts that could involve newly inactive SNOMED codes are located.

The SNOMED CT Query table offers an alternative solution: whenever a taxonomic query is executed, inactive codes are logically (but not also physically) updated with their active substitute. This logical transform is achieved by extending the standard Transitive Closure table (itself normally a required component of environments that support taxonomic querying) to also include additional rows computed as the cross-product JOIN between the Substitutions and Transitive Closure tables.

The primary intended effect of the Query Table, therefore, is that a search that should return the active concept will also return all inactive equivalents, and vice versa.

There are, however, two important caveats or limitations of using the Query Table:

- inactive concepts for which no active substitute could be identified remain as a residue of potentially ‘unreportable’ false negatives; the true clinical risk of this effect is however believed to be extremely small (see s5).
- Using the Query Table may find new false positives over and above those that would be returned if the Query Table were not used at all. This effect can be partly controlled by proper use of the ISAMBIGUOUS column (see s0)

⁴ Medicolegally, the originally entered code selection must always persist and never be changed. For taxonomy-driven reporting purposes, however, it will often be more correct to use a different code to be either computed logically on the fly (eg. by using the Query Table at run time) or stored physically as further slots in the information model (eg. by lookup on the Substitutions table).

2.4 Related SNOMED CT Content Issues

In addition to the behaviour of inactive concepts, described above, SNOMED CT contains two other areas of content that can also give rise to false negative reporting. Between April 2015 and April 2017 these risks were also but only partially mitigated by the Query Table; this functionality was withdrawn from the October 2017 release:

2.4.1 Navigational concepts

At the time of writing, 727 active SNOMED CT concepts reside in a parallel hierarchy below 363743006|[Navigational concept \(navigational concept\)](#)|. Concepts previously contained within the primary taxonomy are moved here if the set of concepts normally given as their descendants can only be determined by reference to:

- the lexical properties of their descriptions:
[273855008|Assessment scales Sw-Va \(navigational concept\)](#)|
[302302007|Enzymes A - L \(navigational concept\)](#)|
[302303002|Enzymes M - Z \(navigational concept\)](#)|
- logical properties that can not be represented in SNOMED CT's description logic:
[106220000|Genetic, molecular AND/OR cellular finding \(navigational concept\)](#)|
[362976009|Finding / observation / clinical history \(navigational concept\)](#)|
- opinion rather than objective fact
e.g. [350084001|Alternative medicines \(navigational concept\)](#)|

As with inactive concepts, the concepts in this *navigational* subhierarchy have no descendants and also lack all clinically expected ancestors.

Unlike inactive concepts, there is no modelled relationship between these concepts and some substitute in the main taxonomy, because no such substitute exists.

In most cases, the concepts found in the subhierarchy represent the projections into SNOMED-space of abstract grouper concepts found within the antecedent terminologies, principally SNOMED RT and Clinical Terms Version 3 (CTV3).

For *some* of those Navigational Concepts derived from CTV3, it is possible to determine the direct parents and children from the main SNOMED taxonomy by reference to NHS Digital's Data Migration Products.

Between April 2015 and April 2017 the SNOMED Query Table therefore also contained rows to reposition some Navigational Concepts back into the main taxonomy. These measures were, however, limited: some Navigational Concepts remained outside the Query Table scope (mainly those deriving originally from SNOMED RT) whilst those that were treated were not necessarily guaranteed to always be 'classified' entirely correctly. Although overall the number of false negatives relating to Navigational Concepts was significantly reduced, and the number of new false positives very low, these additional rows were removed from the October 2017 release pending a more detailed clinical safety assurance. No mitigation may be safer than a partial mitigation, especially if partial mitigation can be misconstrued as complete mitigation.

2.4.2 Situation with explicit context

The SNOMED CT Context Terminological Model⁵ provides a within-terminology mechanism that allows findings to be characterised by ‘context’, including: whether findings are present or absent and procedures performed or not performed; whether the coded statement relates to the subject of the record or to a relative and whether in the past, present or future. This mechanism is *mostly* intended to support contextualisation of postcoordinated expressions but it is *also* embodied in the small subhierarchy of precoordinated concepts below [243796009|Situation with explicit context|](#), wherein currently resides some 6,700 concepts (findings, family history and procedures) with a variety of contexts.

The context model further states that:

When a SNOMED CT code appears in a record without any explicitly stated context, that code is considered to have a default context. The default is "soft" in that it can be over-ridden by information carried in the structure of the record or its information model.

The default context for a clinical finding code implies that the finding has actually occurred (vs. being absent), that it applies to the subject of the record (the patient), and that it is occurring currently or occurred at a past time that is given by a date - time record linked to the code.

The default context for a procedure code implies that the procedure was completed, that it was performed on the subject of the record (the patient), and that it was done at the present time or in the past at a time that is given by a date - time record linked to the code.

In other words, for every code *without* context found in the findings or procedures hierarchy, a semantically equivalent expression implicitly also exists in which the default context has been made explicit. All such equivalent expressions will be found in the Situations subhierarchy; some— a very small minority of what is possible - have already been instantiated there:

267093002 Has a red eye (situation) 	is equal to	246676003 Injection of surface of eye (finding)
162057007 Nausea present (situation) 	is equal to	422587007 Nausea (finding)
287752006 Bursa sutured (situation) 	is equal to	88626007 Suture of bursa (procedure)
432729003 Cortisone injection given (situation) 	is equal to	432861006 Injection of cortisone (procedure)

Note that the reverse is not true; not every pre-coordinated, contextualised concept from the situations subhierarchy has a semantically equivalent *non*-contextualised one in the findings or procedures hierarchy. Many concepts below [243796009|Situation with explicit context|](#) assign a different context to the default implied by non-contextualised content:

274823004 Clonus absent (situation) 	is not equal to	36649002 Clonus (finding)
408566000 Echocardiogram declined (situation) 	is not equal to	40701008 Echocardiography (procedure)
160350005 Family history: Squint (situation) 	is not equal to	22066006 Strabismus (disorder)

The true ‘contextually equivalent’ pairs are important to identify because, although both exist within the full SNOMED CT taxonomy, a taxonomic query for one does not normally also return the other. Similarly, many credible queries can be formulated that would return one and not both.

⁵ See s6.2.3.5 in http://www.ihtsdo.org/fileadmin/user_upload/doc/en_gb/tig.html

From April 2015 to April 2017 the SNOMED Query Table therefore also contained the additional ancestor/descendent rows that were valid for truly 'context equivalent' pairs within the findings, procedures and situations hierarchies, where it was possible to identify a pair.

Thus, when using the Query Table:

162057007|Nausea present (situation)|subsumes 422587007|Nausea (finding)| and vice versa
287752006|Bursa sutured (situation)| subsumes 88626007|Suture of bursa (procedure)| and vice versa

but:

274823004|Clonus absent (situation)| still does **not** subsume 36649002|Clonus (finding)|
160350005|Family history: Squint (situation)| still does **not** subsume 22066006|Strabismus (disorder)|

How contextually equivalent pairs are identified

To explicitly state the default context for a concept in the Findings subhierarchy, the following modelling pattern is used:

```
243796009|Situation with explicit context|:  
  {246090004|Associated finding|=<finding>,  
  408729009|Finding context|=410515003|Known present|,  
  408732007|Subject relationship context|=410604004|Subject of record|,  
  408731000|Temporal context|=410512000|Current or specified|}
```

...and, as a result, all and only all such concepts *should* be autoclassified under
373573001|Clinical finding present (situation)|.

Similarly, the following modelling pattern explicitly states the default context for a procedure:

```
243796009|Situation with explicit context|:  
  {363589002|Associated procedure|=<procedure>,  
  408730004|Procedure context|=385658003|Done|,  
  408732007|Subject relationship context|=410604004|Subject of record|,  
  408731000|Temporal context|=410512000|Current or specified|}
```

...and should result in all and only all such concepts being autoclassified under
443938003|Procedure carried out on subject (situation)|

However, some procedure and finding content from the UK extensions may not be classified under these two groupers, because UK Extension content is not currently built using an autoclassifier. As a consequence, the algorithm to identify pairs of Findings or Procedures with an equivalent SNOMED CT Situation content with the default context is necessarily more complicated: all concepts with Fully Defined status under the more general 243796009|Situation with explicit context (situation)| concept are inspected. A deep structural analysis is performed on each to determine whether its modelling is an example of one or other design pattern shown above. If it is, the equivalent Finding or Procedure is determined to be the modelled value of the 363589002|Associated procedure| or 246090004|Associated finding| attribute, as appropriate.

Note, however, than many situation-with-default-context concepts exist as Primitive (ie *not* Fully Defined). For these, the non-contextualised equivalent from the Findings or Procedures hierarchies can not reliably be automatically identified, though they may exist. The Query Table therefore does not address these undetectable equivalent pairs.

Thus, for example, although the primitive situation-with-default-context concept [720008007|Prostate cancer care review done|](#) is explicitly (and correctly) modelled with a relationship pointing at its equivalent Procedure:

```
720008007|Prostate cancer care review done|:  
  {363589002|Associated procedure|=720007002|Prostate cancer care review (procedure)|,  
  408730004|Procedure context|=385658003|Done|,  
  408732007|Subject relationship context|=410604004|Subject of record|,  
  408731000|Temporal context|=410512000|Current or specified|}
```

...other situation-with-default-context primitives have no such associated_procedure relationship modelled at all:

```
819311000000107|Weight management plan completed|:  
  408730004|Procedure context|=385658003|Done|,  
  408732007|Subject relationship context|=410604004|Subject of record|,  
  408731000|Temporal context|=410512000|Current or specified|}
```

...whilst still others have a modelled associated_procedure relationship but not to the direct equivalent concept, but rather to some more abstract ancestor of it:

```
522381000000101|Ear care information leaflet given|:  
  {363589002|Associated procedure|=71388002|Procedure|,  
  408730004|Procedure context|=385658003|Done|,  
  408732007|Subject relationship context|=410604004|Subject of record|,  
  408731000|Temporal context|=410512000|Current or specified|}
```

Despite existence of [699118000|Provision of ear care information leaflet \(procedure\)|](#)

October 2017 update

Because of the limitations in the underlying modelled content outlined above, the Query Table can currently achieve only a modest and somewhat unpredictable reduction in the number of false negatives relating to 'context equivalent' pairs. Pending a more detailed clinical safety assurance, therefore, these rows are absent from the October 2017 and future releases.

3 Summary of Use Cases

The following use cases are illustrated by reference to the real historical changes to SNOMED CT content that were summarised in Figure 1 and Figure 2:

3.1 EPR Instance Data contains inactive conceptId

After October 2011, a taxonomic search for all patients with [50417007|Lower respiratory tract infection|](#) or any of its subtypes should find patients coded with *either* [308130008|Recurrent chest infection|](#) *or* [195746005|Recurrent chest infection|](#), as well as those coded with [448739000|Recurrent lower respiratory tract infection|](#).

3.2 Query contains inactive conceptId

After October 2011, a taxonomic search for all patients with [308130008|Recurrent chest infection|](#) should find patients coded with [448739000|Recurrent lower respiratory tract infection|](#).

3.3 Inactive conceptIds in Query and EPR instance data

After October 2011, a taxonomic search for all patients with [308130008|Recurrent chest infection|](#) should find patients coded with *either* [195746005|Recurrent chest infection|](#) *or* [448739000|Recurrent lower respiratory tract infection|](#) *or* [308130008|Recurrent chest infection|](#).

4 Design and Content Optimisations

A full SNOMED CT Query Table, containing all additional rows relating to all concepts ever inactivated or otherwise with an equivalent, would be a very large table indeed; in excess of 25 million rows. The UK SNOMED CT Query Table therefore makes certain assumptions that originally allowed for a much more modest table size of just under 6 million rows though both improvements to the algorithm and the ever increasing number of inactive concepts it needs to handle have caused the size to grow; the operational goal is to permit selective optimisations of the table that yield improvements in runtime compute performance but without not significantly reducing utility in most care settings.

Additional rows that would be valid in the *full* table are currently omitted if they derive from an inactive concept *and* any of the following are true:

- The inactive concept was inactivated prior to 1st January 2005, and so there should be very little or no SNOMED CT coded data that was entered prior to that time when the concept was active (but assuming that all more recent new data entry prevents users from selecting inactive concepts)
- The active substitute declared for the inactive concept is equal to any of a limited list of very high level, abstract grouper concepts in SNOMED CT, with correspondingly limited querying value even if they were included in the full Query Table:

138875005 SNOMED CT Concept (SNOMED RT+CTV3)
105590001 Substance (substance)
106237007 Linkage concept (linkage concept)
123037004 Body structure (body structure)
123038009 Specimen (specimen)
243796009 Situation with explicit context (situation)
254291000 Staging and scales (staging scale)
260787004 Physical object (physical object)
272379006 Event (event)
308916002 Environment or geographical location (environment / location)
362981000 Qualifier value (qualifier value)
363787002 Observable entity (observable entity)
370115009 Special concept (special concept)
373873005 Pharmaceutical / biologic product (product)
404684003 Clinical finding (finding)
410607006 Organism (organism)
419891008 Record artifact (record artifact)
48176007 Social context (social concept)
71388002 Procedure (procedure)
78621006 Physical force (physical force)
362955004 Inactive concept (inactive concept)
363743006 Navigational concept (navigational concept)
370136006 Namespace concept (namespace concept)
363660007 Ambiguous concept (inactive concept)

363661006 Reason not stated concept (inactive concept)
363662004 Duplicate concept (inactive concept)
363663009 Outdated concept (inactive concept)
363664003 Erroneous concept (inactive concept)
370126003 Moved elsewhere (inactive concept)
443559000 Limited status concept (inactive concept)
91722005 Entire physical anatomical entity (body structure)
301857004 Finding of body region (finding)
118234003 Finding by site (finding)
123946008 Disorder by body site (disorder)
362965005 Disorder of body system (disorder)
362958002 Procedure by site (procedure)
669771000000105 Operations by other sites or systems (procedure)
309040001 Operations by other sites or systems (procedure)
118664000 Procedure on body system (procedure)
128927009 Procedure by method (procedure)
308965007 Operations by method (procedure)
118225008 Context-dependent finding (finding)
315645005 Unspecified conditions (disorder)
24221000000103 SNOMED CT UK administrative concepts
(administrative concept)
386053000 Evaluation procedure (procedure)

5 Limitations

The Query Table is conceived as an optimised derivative of the information represented in SNOMED CT as historical relationships, asserting ‘semantic equivalence’ links between inactive concepts and their active counterparts (ie the blue lines in Figure 1 and Figure 2).

Provided such links exist, and that they are correct when they do exist, the intended effect of the Query Table is : a search that should return an active concept will also return all its inactive equivalents, and vice versa.

It is, however, known that at least some historical and taxonomic relationships represented in current SNOMED CT content are not, in fact, clinically correct. More significantly, not all valid historical relationships are actually present in current SNOMED CT content: to date (Oct 2016), 89,244 inactive concepts exist in the UK Edition without *any* active, semantically equivalent substitute; these can be easily identified within the Substitutions Table as those rows where OLDCONCEPTID is the same as NEWCONCEPTID.

For these 89,244 inactive concepts, neither the Substitutions nor Query Tables can mitigate the theoretical risk of false negative reporting.

However, the *real* clinical risk arising is likely to be very small: the huge majority relate to SNOMED codes corresponding to NHS Drug and Appliance Dictionary codes, which have no mapping to dm+d by deliberate design and which are therefore already out of scope for the Query Table.

Of the remaining 9,758, 75% were inactivated prior to 2012 and all but 156 relate to one of 18 known reasons why no active semantically equivalent concept exists:

REASON	NUM	REASON	NUM
NHS DRUG AND APPLIANCE DICTIONARY	79459	UNILATERAL/BILATERAL	34
VETERINARY MORBIDITY, ANATOMY OR PATHOGEN	4886	EXINCT CROSS-TYPE CONTENT	31
ORIGINAL CONCEPT AUTHORED IN ERROR	4034	LIST CLOSURE CATEGORY FROM ICD	31
AMBIGUOUS READ/CTV3 CODE	161	CLINICAL COURSE/ONSET	14
CTV3 METADATA	91	DISJUNCTION	14
SURGICAL APPROACHES	81	EXTERNAL ANATOMICAL FEATURES	14
ACTION WITH EXPLICIT DEVICE OR SITE	79	ANIMAL HAIR	10
VETERINARY/OBSOLETE (MICRO)ORGANISM TAXON	76	OBSOLETE NHS ORGANISATION	10
BRANDED FOODS	48	NAVIGATIONAL CONCEPT	9
		GROUPER CATEGORY FROM ICD	6

Taken together, this analysis of the timing of the inactivation and the semantic scope of the concepts involved suggests that (a) few of the codes involved could have ever plausibly been selected to capture clinical problems or interventions (or, if they were, would have constituted serious miscodes) and (b) most of the affected data is clinically no longer current.

The final residue of 156 are listed below, grouped by their SNOMED CT semantic tag; a small number (listed in bold) are already identified as clearly having directly equivalent active concepts, and content change requests have been filed to correct these errors:

10761004 Topography not applicable (body structure)	272433002 Lamella structure (body structure)
113178003 Calamus (body structure)	279187000 Arcus parieto-occipitalis (body structure)
13319006 Proximal interphalangeal joint structure of great toe (body structure)	279188005 Arcus temporo-occipitalis (body structure)
181641001 Entire flexor digitorum superficialis muscle (body structure)	280114000 Surface topography (body structure)
23708007 Superficial digital flexor tendon of hindlimb (body structure)	309163008 Complex structure derived from epithelium (body structure)
263941003 Anterior lamella (body structure)	360490005 Joint types (body structure)
264152002 Posterior lamella (body structure)	360824001 Type of ligament (body structure)

362610008|Entire psyche (body structure)
 37888007|Structure of olecranon (body structure)
 45699006|Conscious mind (body structure)
 65039001|Subconscious mind (body structure)
 89970005|Psyche structure (body structure)
 91721003|Anatomical spatial entity (body structure)
 91804004|Anatomical landmark (body structure)
 91805003|Anatomical feature (body structure)
 91813002|External anatomical feature (body structure)
 91831001|Internal anatomical feature (body structure)
 91834009|Anatomical plexus (body structure)
 91835005|Anatomical region of communication (body structure)

111035001|Hypoplastic obesity (disorder)
 111313001|Syndactyly (disorder)
 111933005|Tick fever (disorder)
 123065002|Infectious bursal disease (disorder)
 123157002|Condyloma acuminatum (disorder)
 124901003|Deficiency of vitamin B (disorder)
 14322002|Brachycephalic dwarfism (disorder)
 14960000|Secondary monoclonal gammopathy (disorder)
 197577008|Nephritis, nephrosis and nephrotic syndrome (disorder)
2300005|Ehlers-Danlos syndrome (disorder)
238848002|Cutis hyperelastica (disorder)
 240898000|Dog mite infestation (disorder)
 25495004|Tick-borne fever (disorder)
 264512007|Cellular proliferative disease (disorder)
 275509000|Infective vascular disease (disorder)
285801000119101|Left anterior hemiblock (disorder)
 34458001|Pseudoglanders (disorder)
 34671005|Villonodular synovitis (disorder)
 45064008|Primary localized osteoarthritis of thigh (disorder)
 48238008|Exercise-induced pulmonary hemorrhage (disorder)
 58673008|Lepromin reaction (disorder)
68706009|Tuberculosis of pleura (disorder)
 72743002|Sabin-Feldman syndrome (disorder)
 74507009|von Willebrand disease, type IS/III (disorder)
 75832005|Pleonotia (disorder)
 83995002|Idiot savant (disorder)

102529000|Age more than 40 years, African-American (finding)
 102530005|Age more than 40 years, African-American male (finding)
 102531009|Age more than 40 years, Native American (finding)
 102532002|Age more than 40 years, Hispanic-American (finding)
 102895009|Morose mood (finding)
 10425005|Normal presence (finding)
 106072004|Sensation related to thoracic organ (finding)
 107653005|Clinical sample finding (finding)
 168786005|Encephalography normal (finding)
 168787001|Encephalography abnormal (finding)
 184136009|Patient distance to surg ? correct (finding)
 185697008|Geriatric screening - "ghost" (finding)
 185698003|Geriatric screening - "ghost" (finding)
 275449006|Sore bottom (finding)
 275815009|Multi.- Antenatal data (finding)
 281287001|Sample left on cells (finding)
 281294003|Sample finding (finding)
 299631000000104|Seen by optometrist with special interest (finding)
 300031000000108|Developmental height centile is cause for concern (finding)
 300061000000103|Developmental weight centile is cause for concern (finding)
 300071000000105|Developmental head circumference centile is cause for concern (finding)
 366016000|Finding of appearance of cornea (finding)
 38752006|Parental legitimization (finding)
 390920000|Place cervical smear taken (finding)
 65655009|Sexual dysfunction (finding)
 81439005|Morphology not assigned in SNOMED (finding)

125286009|Acute content retention (morphologic abnormality)
 125287000|Chronic content retention (morphologic abnormality)

125288005|Recurrent content retention (morphologic abnormality)
 24416008|Multiple burns (morphologic abnormality)
 29498005|Parasarcoma (morphologic abnormality)
 314919009|General tumor morphologies (morphologic abnormality)
 43451003|Congenital deficiency (morphologic abnormality)
 58404005|Congenital impairment (morphologic abnormality)
 67294000|Partial congenital impairment (morphologic abnormality)
 77087008|Total congenital impairment (morphologic abnormality)
 80225002|Congenital ischemic atrophy (morphologic abnormality)

112217000|Refraction and accommodation, function (observable entity)
 123349008|Pregnancy delivered (observable entity)
 20339007|Estrus, function (observable entity)
 66348005|Parturition, function (observable entity)
 75395002|Hearing and speech hearing functions (observable entity)

111456007|Breach extraction without indication (procedure)
 112846001|Removal of nail (procedure)
 117803009|Echovirus 40 antibody assay (procedure)
 117808000|Eperythrozoon cuniculi antibody assay (procedure)
 117983006|Echovirus 140 antibody assay (procedure)
 169117008|Radiology-experimental animals (procedure)
 240997006|Liposuction for extravasation injury (procedure)
 241496002|Fetal echocardiography (procedure)
 252144003|Biochemical test (procedure)
 252423006|Contrast echocardiography (procedure)
 260696008|Type of colostomy (procedure)
 274403008|Pneumoventriculography (procedure)
 288146004|Fracture - compression union (procedure)
 299971000000109|Transfer of care to new provider (procedure)
 313710005|Mitochondrial antibody level (procedure)
 314134004|Reconstruction of nipple (procedure)
 31721001|Nipple reconstruction (procedure)
 31760004|Endotoxin test, animal (procedure)
 33581006|Routine patient monitoring by nurse (procedure)
 367383004|Removal of nail (procedure)
 367440009|Routine patient monitoring by nurse (procedure)
 367445004|Preventive treatment (procedure)
 36835009|Special radioisotope function studies, explain by report (procedure)
 404185003|Polymerase chain reaction test for HAV6 (procedure)
 409872001|Oraquick Rapid HIV-1 antibody test (procedure)
 431940000|Magnetic resonance imaging of heart for assessment of myocardial viability (procedure)
 54387009|Pneumoencephalography (procedure)
 72121002|Family investigation (procedure)
 83220008|Non-invasive medical procedure (procedure)
 83762000|Chemical procedure (procedure)
 86534003|Serial scanning (procedure)
 86873007|Internal cardioassist (procedure)

390908008|Lipid disorder monitoring administration (regime/therapy)

60080006|Egg (cell)

123114007|Normal presence (situation)
 160466006|Family history: Boyfriend (situation)
 160467002|Family history: Girlfriend (situation)
 168785009|Encephalography requested (situation)

120824002|Human echovirus 40 antibody (substance)
 359712008|ROOT (substance)
 97946001|Koate-HS Antihemophilic Factor (Human) (substance)

123175001|Splitting (qualifier value)
 129285002|Diagnostic surgical action (qualifier value)
 257809006|Division of object (qualifier value)
 257810001|Partial division (qualifier value)
 258002001|x (qualifier value)
 258023007|Electrical stimulation (qualifier value)
 258915008|ug/sq. L (qualifier value)

264822008|Both ligamentous (qualifier value)
28707004|Multiple (qualifier value)
30190008|Seen (qualifier value)
303430001|Introduction of substance (qualifier value)
307203009|Grade 3a (qualifier value)
307204003|Grade 3b (qualifier value)
307205002|Grade 3c (qualifier value)
307206001|Grade 4a (qualifier value)

307207005|Grade 4b (qualifier value)
307208000|Grade 5a (qualifier value)
307209008|Grade 5b (qualifier value)
360023008|Specialized action (qualifier value)
82378008|Unlisted physical agent (qualifier value)
82975001|Denied (qualifier value)

6 File format and usage

The SNOMED CT Query Table is presented as a 3-column, TAB separated file. Rows are terminated by CR/LF combination, and the first row contains the relevant column names.

The table schema is as follows:

COLUMN	LENGTH	TYPE / PATTERN	DATABASE TYPE	NOTE
SUPERTYPEID	18	SCTID	VARCHAR (18) or BIGINT(10) NOT NULL	SNOMED CT Concept Identifier. Not unique
SUBTYPEID	18	SCTID	VARCHAR (18) or BIGINT(10) NOT NULL	SNOMED CT Concept Identifier. Not unique
PROVENANCE	1	0 1 2 3	TINYINT NOT NULL	Indicates the provenance of the inference that SUBTYPEID is a descendent of SUPERTYPEID; should be used to suppress (or flag) possible false positive and negative results.

6.1 Column Details

SUPERTYPEID

The 64-bit integer identifier for a SNOMED CT concept (up to 18 digits). SNOMED Ids *may* be stored as VARCHAR(18), but indexing and lookup operations typically run faster when stored as BIGINT(10).

SUBTYPEID

The 64-bit integer identifier for a SNOMED CT concept (up to 18 digits) that, for the purposes of taxonomic querying, is to be treated *as though it were* a taxonomic descendent of the concept referenced as SUPERTYPEID.

SNOMED Ids *may* be stored as VARCHAR(18), but indexing and lookup operations typically run faster when stored as BIGINT(10).

PROVENANCE

A single digit value, encoding the provenance of the inference that SUBTYPEID is a descendent of SUPERTYPEID; currently in range 0-3. Other values may be introduced as a result of planned content improvements:

0	SUBTYPEID is related to SUPERTYPEID exclusively by means of regular taxonomic inheritance <i>plus</i> one or more of: <ol style="list-style-type: none">1. SAME_AS, REPLACED_BY or MOVED_FROM historical relationships2. Default context equivalence
---	--

	<p>3. CTV3/READ2 Navigational Subset inference</p> <p>(including possibly by chains of indefinitely many relationships exclusively of these types)</p> <p>These types of relationships are interpreted as unambiguously implying true concept identity between pairs of concepts ie <u>everything</u> true of one is also true of the other.</p>
1	<p>SUBTYPEID is related to SUPERTYPEID via regular taxonomic inheritance <i>plus</i> at least one MAY_BE relationship, but where every inactive subject of the MAY_BE relationships involved has only one candidate active SNOMED CT conceptId substitute.</p> <p>Theoretically, ambiguous inactive concepts for which only one active substitute is specified <i>should</i> imply that SNOMED CT no longer offers a means to encode any other possible interpretations of the original code and, therefore, no false positive substitutes exist to be returned.</p> <p>Practically, however, this phenomenon may also occur when SNOMED CT's set of historical MAY_BE relationships is incomplete.</p> <p>Accordingly, whilst it is more likely to be the case that the designated substitute is a true positive, and no false positives or false negatives exist, users' attention may reasonably be drawn to search results obtained using these rows.</p>
2	<p>SUBTYPEID is related to SUPERTYPEID by means of regular taxonomic inheritance <i>plus</i> at least one WAS_A relationship, so that everything that is true of SUPERTYPEID is also true of SUBTYPEID but the reverse is not known to be true. For example, the ancestors of SUPERTYPEID are also ancestors of SUBTYPEID (but SUBTYPEID may also have other ancestors). However, the descendants of SUPERTYPEID may not be also descendants of SUBTYPEID.</p>
3	<p>SUBTYPEID is related to SUPERTYPEID by means of regular taxonomic inheritance <i>plus</i> at least one overtly ambiguous MAY_BE historical relationship that is associated with an inactive concept for which more than one possible active substitute is given.</p> <p>Results obtained using these rows will therefore almost certainly include false positives, and so users should normally be alerted.</p>

6.2 Recommendations for usage

Most taxonomic query implementations should, therefore, normally legitimately *always* include lookup on rows where ISAMBIGUOUS=0. Users need not be notified of any results obtained as a result; these rows typically entirely remove the risk of either false negative or false positive reporting for affected inactive concepts.

Lookup on rows where ISAMBIGUOUS = 2 does not normally entail a risk of false positives, but users may benefit from being notified that the risk of false negatives is unlikely to have been fully mitigated.

Queries executed in an environment where false positives can be tolerated (e.g. case finding where a second validation by human inspection step will occur) can also normally include lookup of rows where ISAMBIGUOUS=1 or 3, but that portion of the result set obtained by this route should be highlighted to the user tasked with weeding out the false positives.

In safety critical situations where false positives can not be tolerated, and no review to weed them out can be implemented, the ISAMBIGUOUS=1 or 3 rows should usually NOT be consulted. For example, when running reports to automatically enrol patients for interventions.

6.3 Sample implementation/test harness

A small Query Table test harness implementation was made available from April 2017. It takes the form of a simple table of simulated coded EPR data, and a MySQL script file to load that data and then to query it both with and without the Query Table.

The SQL script file requires the prior existence of a 'snomed' schema containing at least the following three tables, as named below:

1. rf2_concept_sp - a recent snapshot of the rf2_concept table from the RF2 distribution of the UK Edition of SNOMED CT
2. TRANSITIVECLOSURE - a precomputed transitive closure table for the same edition of SNOMED CT
3. QUERYTABLE - the Query Table as distributed

Further outline instructions on how to set up and then use the simple test harness are provided at the top of the same SQL script file.