



OPEN ACCESS

Semantic analysis of SNOMED CT for a post-coordinated database of histopathology findings

Walter S Campbell,¹ James R Campbell,² William W West,¹ James C McClay,³ Steven H Hinrichs¹

¹Department of Pathology and Microbiology, University of Nebraska Medical Center, Omaha, Nebraska, USA

²Department of Internal Medicine, University of Nebraska Medical Center, Omaha, Nebraska, USA

³Department of Emergency Medicine, University of Nebraska Medical Center, Omaha, Nebraska, USA

Correspondence to

Dr Walter S Campbell, University of Nebraska Medical Center, Department of Pathology and Microbiology, 985900 Nebraska Medical Center, DRC 2, Room 8064, Omaha, NE 68198-5900, USA; wcampbel@unmc.edu

Received 1 November 2013

Revised 21 April 2014

Accepted 26 April 2014

Published Online First

15 May 2014

ABSTRACT

Objective This research investigated the use of SNOMED CT to represent diagnostic tissue morphologies and notable tissue architectures typically found within a pathologist's microscopic examination report to identify gaps in expressivity of SNOMED CT for use in anatomic pathology.

Methods 24 breast biopsy cases were reviewed by two board certified surgical pathologists who independently described the diagnostically important tissue architectures and diagnostic morphologies observed by microscopic examination. In addition, diagnostic comments and details were extracted from the original diagnostic pathology report. 95 unique clinical statements were extracted from 13 malignant and 11 benign breast needle biopsy cases.

Results 75% of the inventoried diagnostic terms and statements could be represented by valid SNOMED CT expressions. The expressions included one pre-coordinated expression and 73 post-coordinated expressions. No valid SNOMED CT expressions could be identified or developed to unambiguously assert the meaning of 21 statements (ie, 25% of inventoried clinical statements). Evaluation of the findings indicated that SNOMED CT lacked sufficient definitional expressions or the SNOMED CT concept model prohibited use of certain defined concepts needed to describe the numerous, diagnostically important tissue architectures and morphologic changes found within a surgical pathology microscopic examination.

Conclusions Because information gathered during microscopic histopathology examination provides the basis of pathology diagnoses, additional concept definitions for tissue morphometries and modifications to the SNOMED CT concept model are needed and suggested to represent detailed histopathologic findings in computable fashion for purposes of patient information exchange and research.

Trial registration number UNMC Institutional Review Board ID# 342-11-EP.

BACKGROUND AND SIGNIFICANCE

SNOMED CT, originally developed by the College of American Pathologists (CAP) and now the product of the International Health Terminology Standards Development Organization (IHTSDO), is the international *lingua franca* for encoding clinical findings within the electronic health record (EHR) and has been adopted for use in the USA, Canada, UK, Australia, and many other nations. To satisfy Meaningful Use requirements, the Office

of the National Coordinator (ONC) requires SNOMED CT to be used to encode problem lists and selected findings within the EHR.¹⁻³ This requirement includes encoding of clinical findings for communication of patient health summaries between healthcare entities. The CAP's Cancer Checklists^{4 5} also incorporate pre-coordinated SNOMED CT⁶ for many required reporting elements to document standardized cancer reporting at a summative level. The continued development of SNOMED CT expressions that accurately represent the clinician's intended meaning is important for patient care, transitions of care, and patient outcome research.

Surgical pathology practice entails the examination of histologically prepared glass slides by a pathologist, and development of a report that is the summary of the pathologist's findings. The contents of the diagnostic report consist of the pathologist's interpretation of the images and the clinical information provided beforehand.^{7 8} Specific diagnostic features noted in the microscopic exam may be referred to in the final report, but explicit details of tissue architectures and diagnostic morphologic changes are not exhaustively represented in the diagnostic report. SNOMED CT encoding of detailed microscopic findings is a process that is untested as most surgical pathology databases today are created by natural language processing of final dictated reports. The explicit and detailed findings of the diagnosing pathologist may not be easily converted into computable terms for reporting purposes, patient information communication, or research.

Microscopic findings and tissue morphometries recorded in computable form can be used to enhance the surgical pathology report and could be incorporated into clinical decision support systems for histopathology-based diagnoses.^{9 10} In conjunction with digital images, encoded microscopic findings can enrich medical training programs with detailed, annotated pathology image repositories which could also serve as a resource to facilitate computer aided diagnostic devices. Ultimately, an accurately indexed library of microscopic findings is the foundation for achieving the promise of 'Big Data' wherein genomic associations with abnormal tissue morphologies can be identified.

SNOMED CT expressions for clinical findings may sometimes be represented by a pre-coordinated definition. Pre-coordinated SNOMED CT definitions are published by the IHTSDO and consist of a unique SNOMED CT concept identifier linked to



Open Access
Scan to access more
free content



CrossMark

To cite: Campbell WS, Campbell JR, West WW, et al. *J Am Med Inform Assoc* 2014;**21**:885–892.

a set of relationships that bind the term to the concepts within the SNOMED CT concept hierarchy that capture the unambiguous semantic meaning of the finding. Expressions which cannot be fully defined within the SNOMED CT model are identified as semantically incomplete and flagged as 'primitive'. When pre-coordinated content is not fit for purpose, IHTSDO provides guidance to properly construct post-coordinated SNOMED CT expressions wherein syntactically normalized SNOMED CT expressions are developed at the time of recording of the finding. SNOMED CT concepts and relationships may be combined in such a way to unambiguously represent a clinical finding following the SNOMED CT concept model as explained in the SNOMED CT User Guide,¹¹ and more recently the SNOMED CT Starter Guide,¹² to expand the expressivity of SNOMED CT.

The objective of this research was to investigate SNOMED CT as an expressive terminology to describe detailed histopathologic findings in order to capture surgical pathology findings in a discrete, explicit, and interoperable fashion. *It was hypothesized that microscopic histopathologic findings could be accurately represented using SNOMED CT.* The research further sought to elucidate attributes, values, and syntax of SNOMED CT that might be required to enhance the expressivity of SNOMED CT for histopathologic findings in surgical pathology.

MATERIALS AND METHODS

Twenty-four breast biopsy cases (13 malignant and 11 non-malignant diagnoses) were selected for review from cases previously signed out as part of the University of Nebraska Medical Center Department of Pathology and Microbiology breast pathology service. Efforts were made to include cases that demonstrated a variety of diagnostic features as noted in the final diagnoses reported in the laboratory information system (LIS).

The surgical pathologists reviewed digital whole slide images (WSI) of histologically prepared glass slides of the selected cases to identify tissue architectural features of diagnostic importance and tissue morphologies contributing to the final diagnosis. Histopathologic features supporting the final diagnoses were marked up using WSI viewing software tools. Each marked up feature was annotated by the individual pathologist in their own words, thereby creating a series of stated assessments. The diagnostic comments and statements contained in the image annotations and the final, sign-out report as recorded in the LIS were categorized and reduced to 95 lexically distinct statements (table 1).

After each case was reviewed, marked up, and annotated, the authors (WSC, JRC) analyzed the meaning of the clinical statements based on the underlying semantics intended by the pathologists and identified pre-coordinated SNOMED CT expressions or developed post-coordinated SNOMED CT expressions to accurately and comprehensively represent each microscopic finding. The SNOMED CT concept model as defined in the 2012 SNOMED CT Editorial Guide¹¹ and the 2012 SNOMED CT Technical Guide¹³ was strictly observed. SNOMED CT July 2012 international release was the reference terminology release. The Clinclue Xplore SNOMED CT browser utility (The Clinical Information Consultancy Ltd, UK, 2011) was used to perform word searches to identify possible concepts to include in the definitional expression of each histological finding.

On development of post-coordinated SNOMED CT expressions for each histological feature, the expressions were reviewed by the pathologists who made the statements to ensure the integrity of the expressions between the stated definition and the intended meanings. Changes to the post-coordinated

expression were made as necessary to ensure that the SNOMED CT expression accurately defined the intent of the pathologist's stated assessment and remained consistent with the SNOMED CT concept model as specified in the Technical Guide.¹³ In particular, the SNOMED CT concept model consists of a limited number of top level concept hierarchies, including |clinical finding|. Each concept within a hierarchy is defined by a series of attribute-value pairs. Attributes represent definitional aspects of the concept (eg, a clinical finding is defined by attributes such as |finding site|, |abnormal morphology|, and |finding method|). Each attribute is paired with a concept value that the attribute asserts (eg, to assert that the finding is in the breast, then |finding site|=|structure of breast|). The model specifies the definitional requirements and constraints that must be followed to properly construct a concept within the top level hierarchy (ie, allowable attribute domains and the range of allowable concept values). A senior SNOMED CT terminologist (JRC) also reviewed each expression for semantics and adherence to the SNOMED CT concept model. Cases where no SNOMED CT expression could be developed to accurately represent the intended meaning with adherence to SNOMED CT editorial rules were inventoried and classified by reason of encoding failure.

RESULTS

Of the 24 breast biopsy cases, 95 diagnostic or pre-diagnostic features were marked up and annotated by two pathologists (WWW, SHH) or were found to be explicitly stated in the final diagnostic summary. Sixty-nine statements represented conclusive or probabilistic diagnostic assertions, and 26 pathologist statements were descriptive in nature. Seventy-three unique post-coordinated SNOMED CT expressions were constructed from these stated definitions.

The meaning of complex statements which represented conjunctions such as 'fibrocystic changes including stromal fibrosis, apocrine metaplasia, cyst formation, and hyperplastic changes' were captured by employing the SNOMED CT expression syntax for complex expressions (box 1). Only one finding of 74 was represented by a pre-coordinated concept, |calcification of breast (finding)|.

Representation of numbers

Eleven of the 13 malignant cases involved linear measurement of a tumor extent. Attribute-value pairs for |observable entity (observable entity)| were used for recording dimensions of tumors or ductal carcinoma in situ (DCIS) involvement within the biopsy using values of |tumor size, invasive component, greatest linear dimension (observable entity)| or |linear extent of involvement of carcinoma in specimen obtained by needle biopsy (observable entity)|. A SNOMED CT standard for numerical representation is currently under ballot by IHTSDO,¹³ but to date has not been approved. Therefore, these post-coordinated expressions including dimensions could not accurately be rendered with the current concept model. As a statement of clinical finding, the SNOMED model extension under ballot expects that as an attribute-value pair of | has interpretation (attribute)|={numerical observation+units of measure}. Since this is a known limitation of the SNOMED expression syntax undergoing ballot review, the condition was only counted once in our inventory of SNOMED CT limitations.

This left a total of 24 statements of the total 95 unique clinical statements (25%) that could not be adequately represented

Table 1 Lexically unique physician statements captured in annotations and diagnostic reports

Extent of DCIS in biopsy	Apocrine hyperplasia	Lymphocytic aggregation	Abnormal epithelial cells	Cyst formation
Focal hyperplasia without atypia	Abnormal epithelial cells infiltrating stroma	Multifocal—papillomatosis	Area of pathology with cystic change	Patchy lymphocytic mastitis
Focal tubular formation of epithelial cells	Extensive periductal sclerosis	DCIS with possible invasion	Invasive ductal carcinoma; grade 1/3	DCIS suspicious for microinvasion
Microcalcification in DCIS	Mild, sclerosing adenosis	Intralobular fibrosis	No malignancy	No atypia
DCIS with necrosis, solid growth pattern, high nuclear grade	Possible ductal hyperplasia requiring further study	Papillary proliferation of epithelium and stromal tissue	Dense fibrous replacement of normal architecture	Invasive adenocarcinoma with tubulolobular features, Nottingham grade 2/3
Invasive lobular carcinoma	Breast tissue with sparse ducts and lobules	Focal area of apocrine metaplasia (blebs)	Necrosis with possible microcalcifications	Proliferative fibrocystic changes
Breast tissue effaced by probable neoplastic process	Ductal carcinoma in situ, solid growth pattern, high nuclear grade	Dense hyalinized connective tissue with lymphocytic infiltration	Fragmented portions of fibroadenomatoid nodules	Biphasic pattern, benign epithelial component and benign stromal components
Dilated duct with inspissated proteinaceous material and hyperplastic changes	p63—focally positive but not definitive for presence of myoepithelial cells	Simple epithelium overlying dense fibrous connective tissue forming large cystic structure—ectasia	Associated DCIS, low nuclear grade, cribriform growth pattern, non-necrotic	Ductal carcinoma in situ, micropapillary and cribriform growth, intermediate grade with necrosis
Normal lymphocytes	Normal epithelial cells	Ductal ectasia	Fragment of papilloma	Nodular adenosis
Mild cystic changes	Fibroadenoma	Papilloma	Calcification	Microcalcifications
Fibroadenomatoid change	Extensive periductal inflammation	Sclerosing adenosis	Invasive lobular adenocarcinoma	Portion of cyst wall—benign
Florid usual hyperplasia	Fibroadenomatoid changes	Adenosis	No associated DCIS	Nodular focus
Invasive ductal adenocarcinoma, high grade	Ductal hyperplasia, non-atypical (usual)	Invasive ductal adenocarcinoma, Nottingham grade 3/3	Nests and irregular chords of pleomorphic epithelial cells	Nodular adenosis with associated columnar cell changes
Apocrine change	Lobular carcinoma in situ	Perineural invasion	Fibrosis	Cystic dilated ductule
Questionable area of adenosis and possible neoplasia requiring further exam	Proliferation of epithelial cells with abundant eosinophilic cytoplasm and distinct apical blebs. Minimal nuclear pleomorphism without mitoses or invasion	Intraductal papillary proliferation of epithelial and stromal elements without atypia or pleomorphism—papillomatosis	AE1/AE3—keratin stain highlights connections between cell groups and only a few individual cells	Ductal carcinoma in situ, solid growth pattern, intermediate nuclear grade with necrosis
Invasive ductal adenocarcinoma, Nottingham grade 2/3	Greatest contiguous linear extent of invasive carcinoma	Possible microcalcification within ducts	Invasive lobular carcinoma, Nottingham grade 2/3	Linear pattern/formation suggestive of lobular pattern
Mildly dilated duct with inspissated proteinaceous material	DCIS—cribriform growth, micropapillary overgrowth, intermediate nuclear grade	DCIS with solid and cribriform growth pattern, intermediate nuclear grade with necrosis	Abnormal epithelial cells with infiltrating pattern and lobular formation	Invasive chords of neoplastic ductal cells surrounded by reactive fibrous connective tissue
Possible nodular adenosis	Benign breast tissue	LCIS (e-cadherin used)	Radial scar	Columnar cell changes
Fibrocystic changes	Apocrine metaplasia	Elastosis	Cyst	Stromal fibrosis

DCIS, ductal carcinoma in situ; LCIS, lobular carcinoma in situ.

Box 1 Stated finding of fibrocystic changes refined by specified abnormal morphologies

Fibrocystic changes including stromal fibrosis, apocrine metaplasia, cyst formation, and hyperplastic changes [IS A]

- ▶ 404684003|clinical finding|
- ▶ 363698007|finding site|=279009002|glandular structure of breast|,
- ▶ 116676008|associated morphology|=367647000|fibrocystic changes|,
- ▶ 116676008|associated morphology|=367643001|cyst|,
- ▶ 116676008|associated morphology|=81274009|apocrine metaplasia|,
- ▶ 116676008|associated morphology|=1112674009|fibrosis|,
- ▶ 116676008|associated morphology|=31390008|epithelial hyperplasia|,
- ▶ 418775008|finding method|={104210008|hematoxylin and eosin stain method|+}
- ▶ 252416005|histopathology test|+104157003|light microscopy|}

Box 2 Formalism for invasive ductal carcinoma with associated DCIS

Invasive ductal carcinoma with associated DCIS [IS A]

- ▶ 404684003|clinical finding|
- ▶ 363698007|finding site|=279009002|glandular structure of breast|,
- ▶ 116676008|associated morphology|=82711006|infiltrating duct carcinoma|,
- ▶ 47429007|associated with|=404684003|clinical finding|
- ▶ (363698007|finding site|=64633006|lactiferous duct structure|,
- ▶ 116676008|associated morphology|=86616005|intraductal carcinoma, non-
- ▶ infiltrating, no ICD-0 subtype|),
- ▶ 418775008|finding method|={104210008|hematoxylin and eosin stain method|+}
- ▶ 252416005|histopathology test|+104157003|light microscopy|}

by SNOMED CT expressions. In all, valid SNOMED CT expressions were constructed for 75% of the assessment statements.

Related findings

Concepts subsumed by the clinical finding hierarchy are defined by a set of attributes and a range of concept values that include |finding site| with allowable values of |anatomical or acquired body structure| and its subtypes or the attribute |associated morphology| with allowable values of |morphologically abnormal structure| and its subtypes. Each attribute is paired with a defined SNOMED CT concept value (eg, 31737007|structure of small lactiferous ductules| or 31390008|epithelial hyperplasia|), thus creating a list of attribute–value pairs. The SNOMED CT concept model dictates which attributes and values may be used to define |clinical finding (finding)|.¹¹ A |clinical finding (finding)| is considered ‘fully-defined’ when all necessary and sufficient permissible attribute–value pairings have been enumerated to unambiguously define the |clinical finding| concept in question.

The defining attributes required for each post-coordinated expression implicit to the histologic methods employed and the specimens examined as part of this research project consisted of |finding site|, |associated morphology| and |finding method|. The |severity| qualifier was utilized when necessary to assert degrees of the |morphologically abnormal structure| attribute–value pair (eg, severe epithelial hyperplasia or mild, hyalinized fibrosis).

Breast cancer diagnostic statements often asserted the presence of cancer (eg, DCIS) and a histologic grade or Nottingham score. In the current SNOMED CT concept model, histologic grade and Nottingham scores of carcinomas are defined as pre-coordinated, primitive |clinical finding (finding)| concepts and are not included in the domain of defining attributes of |clinical finding (finding)|. However, the defining attribute, 47429007|associated with (attribute)| is in the allowed domain and can be paired with a defined clinical finding concept to assert the presence of a related clinical finding. Therefore, the defining attribute, |associated with (attribute)|, was paired with a clinical finding concept value of the appropriate histologic grade or Nottingham score such as, |nottingham combined grade I: 3–5

points|, to assert a cancer diagnosis with a histologic grade or Nottingham score. To assert concomitant conditions that must be enumerated in synoptic reports,^{14 15} such as DCIS found in the presence of invasive ductal carcinoma, the abnormal morphology concepts were listed individually as shown in **box 2**. The microscopic |anatomical or acquired body structure (body structure)| values in this study were limited to six specific SNOMED CT |anatomical or acquired body structure (body structure)| concept codes pertaining to the glandular structure of the breast, three codes pertaining to non-glandular connective tissue, and one code generalizing breast structure.

Missing SNOMED CT concepts

A defined SNOMED CT concept for the lumen of the breast duct or ductule did not exist in the July 2012 release but was pre-coordinated in July 2013 (64633006|structure of lumen of lactiferous duct (body structure)|. The lumen of the breast duct was used in one finding in the 24 cases analyzed as part of this study.

To assert the proper concept from the procedure hierarchy for the findings noted by light microscopy, histopathology, and hematoxylin and eosin stain, three procedure concept values were combined. Namely, concept values for |light microscopy|, |histopathology test| and |hematoxylin and eosin stain method| were combined to assert that the clinical finding was made by the light microscopy examination of a histology specimen prepared with hematoxylin and eosin stain. For findings noted on slides prepared with immunohistochemistry (IHC) stains, the procedure value for hematoxylin and eosin stain method| was replaced with the value |IHC procedure|. However, no SNOMED CT concept codes for the specific IHC |procedures required for differential diagnoses were defined (ie, the p63 stain method, AE1/AE3 (pan keratin) or e-cadherin). As such, the findings made by these procedures could not be defined completely, but rather remained generalized to an IHC procedure. This resulted in three findings where SNOMED CT expressions could not be sufficiently defined within the constraints of the 2012 international release.

All clinical findings in the 24 breast biopsy cases analyzed were defined by 44 abnormal morphology values that were

Box 3 Two associated morphology concepts joined (bolded) to signify a singular, abnormal morphology

Epithelial hyperplasia with atypia |IS A|

- ▶ 404684003|clinical finding|
- ▶ 363698007|finding site|=31737007|structure of small lactiferous ducts|,
- ▶ 116676008|associated morphology|=
- ▶ **(31390008|epithelial hyperplasia| + 44085002|atypia suspicious for malignancy|)**,
- ▶ 418775008|finding method|={104210008|hematoxylin and eosin stain method|+}
- ▶ 252416005|histopathology test|+104157003|light microscopy|}

paired with the defining attribute, |morphologically abnormal structure|. In six stated definitions, two or more concepts were joined to assert the co-occurrence of two or more abnormal morphologies observed in the same |anatomical or acquired body structure (body structure)| and whose co-occurrence signified a unique, singular finding and not a simple co-occurrence of unrelated, distinct abnormal morphologies. For example, the clinical statement ‘epithelial hyperplasia with atypia’ required the binding of the concepts |epithelial hyperplasia| and |atypia suspicious for malignancy| to create a new concept that asserts that epithelial hyperplasia with atypia |IS A| epithelial hyperplasia and |IS A| atypia suspicious for malignancy (box 3).

Uncertainty and significant negatives

Within the SNOMED CT concept model, clinical statements that assert conditions of probability or the specific absence of a finding require the use of |situation with explicit context|. Box 4 demonstrates the use of |situation with explicit context| to express the verbal statement of ‘no malignancy of breast’. The conjugation of |situation with explicit context (situation)| was required when conditions of probability and/or absence of a finding occurred in combination with the positive presence of another finding. For example, the stated finding of ‘focal hyperplasia without atypia’ entailed the creation of a conjunction of 243796009|Situation with explicit context (situation)| using the post-coordinated value of epithelial hyperplasia with the 404684003|Clinical finding (clinical finding)| attribute along with the |finding context| of |known present|. This |situation with explicit context (situation)| was grouped with the |situation with explicit context (situation)| consisting of the |clinical finding (clinical finding)| attribute with the post-coordinated value for epithelial cell atypia and the finding context of |known absent| (box 5). Other examples of exclusionary findings included hyperplasia without atypia and cystically dilated ductules without atypia. Therefore, if the scope of implementation of a surgical pathology database is to include statements of probability or clinical absence, explicit context must be modeled for all findings.

Absence of cellular architecture

SNOMED CT concepts describing morphologic features as described by the pathologists and included in the stated assessment of the observation which included descriptions for cellular formations were not present in allowable concept hierarchies for clinical findings and/or no concept definition was present

Box 4 Formalism for no malignancy of breast using |situation with explicit context|

No malignancy of breast |IS A|

- ▶ 243796009|situation with explicit context|:
- ▶ {408729009|finding context|=410516002|known absent|,
- ▶ 246090004|associated finding|=404684003|clinical finding|=
- ▶ (363698007|finding site|=76752008|breast structure|,
- ▶ 116676008|associated morphology|=86049000|malignant neoplasm, primary|
- ▶ 418775008|finding method|={104210008|hematoxylin and eosin stain
- ▶ method|+}
- ▶ 252416005|histopathology test|+104157003|light microscopy|}),
- ▶ 410510008|temporal context value|=410585006|current – unspecified|,
- ▶ 408732007|subject relationship context|=410604004|subject of record|}

Box 5 Expression joining two situations with explicit context to represent focal epithelial hyperplasia without atypia

Focal hyperplasia without atypia |IS A|

- ▶ 243796009|situation with explicit context|:
- ▶ {408729009|finding context|=410515003|known present|,
- ▶ 246090004|associated finding|=404684003|clinical finding|=
- ▶ (363698007|finding site|=76752008|breast structure|,
- ▶ 116676008|associated morphology|=36949004|focal epithelial hyperplasia|
- ▶ 418775008|finding method|={104210008|hematoxylin and eosin stain
- ▶ method|+}
- ▶ 252416005|histopathology test|+104157003|light microscopy|}),
- ▶ 410510008|temporal context value|=410585006|current – unspecified|,
- ▶ 408732007|subject relationship context|=410604004|subject of record|}
- ▶ +
- ▶ 243796009|situation with explicit context|:
- ▶ {408729009|finding context|=410516002|known absent|,
- ▶ 246090004|associated finding|=404684003|clinical finding|=
- ▶ (363698007|finding site|=4212006|epithelial cell|,
- ▶ 116676008|associated morphology|=44085002|atypia suspicious for
- ▶ malignancy|,
- ▶ 418775008|finding method|={104210008|hematoxylin and eosin stain
- ▶ method|+}
- ▶ 252416005|histopathology test|+104157003|light microscopy|}),
- ▶ 410510008|temporal context value|=410585006|current – unspecified|,
- ▶ 408732007|subject relationship context|=410604004|subject of record|}

Table 2 Descriptive assessments which could not be fully defined within the SNOMED CT concept model and 2012 international release content

Linear pattern/formation suggestive of lobular pattern
Focal tubular formation of epithelial cells
Nodular focus
Nests and irregular chords of pleomorphic epithelial cells
Invasive chords of neoplastic ductal cells surrounded by reactive fibrous connective tissue
Abnormal epithelial cells with infiltrating pattern and lobular formation
Dense fibrous replacement of normal architecture
Breast tissue with sparse ducts and lobules
Intraductal papillary proliferation of epithelial and stromal elements without atypia or pleomorphism
Papillary proliferation of epithelium and stromal tissue
Area of pathology with cystic change
Proliferation of epithelial cells with abundant eosinophilic cytoplasm and distinct apical blebs. Minimal nuclear pleomorphism without mitoses or invasion
Simple epithelium overlying dense fibrous connective tissue forming large cystic structure
Biphasic pattern, benign epithelial component and benign stromal components
Portion of cyst wall
Nodular adenosis
Fragment of papilloma
Fragmented portions of fibroadenomatoid nodules
Cyst formation
Patchy lymphocytic mastitis

within any concept hierarchy to describe the observed tissue morphometry. Therefore, valid post-coordinated SNOMED CT expressions could not be created for stated assessments such as ‘nests and irregular chords of pleomorphic epithelial cells’ or ‘dense hyalinized connective tissue’. This condition prevented the creation of post-coordinated SNOMED CT expressions for 20 findings (or 21% of the stated expression) in this dataset (see [table 2](#)). To express the pathologists’ statements of degree of morphology observed within the histologically prepared slide, the 272141005|Severities (qualifier value)| attribute was used.

DISCUSSION

The utility of SNOMED CT was evaluated as a means to represent the microscopic findings as stated by surgical pathologists in 24 breast biopsies. Sixty-nine of the 95 (75%) listed clinical assessments could be accurately and comprehensively encoded with existing SNOMED CT content and the current concept model. The remaining 26 (25%) clinical statements could not be adequately represented using the July 2012 international release of SNOMED CT. The areas in which SNOMED CT lacked adequate expressivity could be categorized into two groups. The first group was represented by assessments for which no defined SNOMED CT concepts existed in the July 2012 SNOMED CT release. The second group could not be represented with SNOMED CT expressions because of constraints in the current SNOMED CT concept model.

The absence of defined SNOMED CT concepts encountered in this research was primarily limited to specific IHC stains. The concept definition 117617002| IHC procedure (procedure)| exists but does not provide the specificity to describe 404684003|Clinical finding (clinical finding)| by the unique IHC procedure used by the pathologist. Enumeration of SNOMED CT concept definitions for specific IHC stains would

be required to achieve the expressivity required by anatomic pathologists in their current daily practice.

The single limitation based on |Anatomical or acquired body structure (body structure)| encountered in this study was the absence of a defined concept for the lumen of the breast duct. This deficiency has been corrected in the 2013 SNOMED CT release and no longer presents a definitional issue for purposes of the findings of this research project.

The SNOMED CT expression of clinical statements that used descriptive language was difficult, and often not possible to achieve in this study. Twenty clinical statements could not be encoded using post-coordinated SNOMED CT expressions because no concept codes existed in the July 2012 international release that asserted the proper meaning of the stated clinical definition. For example, the diagnostic expression ‘nests and cords of pleomorphic epithelial cells’ could not be formed into a SNOMED CT expression because pattern or shape concepts were not defined for ‘nests’ or ‘cords’. Concepts for some cellular or tissue formations do exist within the SNOMED CT concept model, but they are found in the qualifier value/formations/descriptors hierarchy. The SNOMED CT concept model will have to support additional concepts for tissue morphometrics and architectural features within the associated morphology hierarchy if this meaning is to be properly recorded.

Definitive or conclusive abnormal morphology statements could be represented by defined SNOMED CT concepts. However, descriptive statements consistent with the conclusive abnormal morphology statement could not be represented using SNOMED terminology. For example, duct ectasia is a defined SNOMED CT concept, 110420004|duct ectasia|. However, the architectural features that describe duct ectasia, that is, ‘simple epithelium overlying dense fibrous connective tissue forming large cystic structures’ cannot be represented using SNOMED CT. SNOMED CT permits synonym descriptions for defined concepts which can accommodate descriptive utterances, but development of concept definitions for the basic tissue architectural features may be a better approach for use in histopathology.

The practice of surgical pathology is largely that of pattern recognition by the pathologist of tissue specimens viewed by light microscopy with a given clinical context. The use of descriptive language concerning architectural features, shapes, and patterns of tissue formations is an important part of reaching differential diagnoses and in training pathology residents. Descriptive statements of tissue architecture within the SNOMED CT concept model present a challenge for the use of SNOMED CT findings expressions at the microscopic level. Restricting SNOMED CT expressions to definitive, conclusive abnormal morphology concepts without providing a descriptive layer of permissible concepts places artificial limitations on characterizing observed tissue morphometrics. The development of a concept hierarchy of architectural concepts to be used within the CLINICAL FINDING hierarchy should be investigated. It should be determined which architectural concepts are definitional and used to distinguish between diagnostic conditions/disorders and those architectural concepts that serve as qualifiers of definitional concepts. Application of both uses of architectural concepts can be found in the anatomic pathology diagnostic practice. This differentiation of architectural concepts is important and will affect their representation within the overall SNOMED CT concept model.

The current release of SNOMED CT, the Technical Users Guide¹³ and the Editorial Guide¹¹ do not adequately address or define how to properly express certain clinical statements

important to defining surgical pathology microscopic findings. In 11 of the 12 cancer cases reviewed, diagnostic statements concerning the greatest linear extent of invasive carcinoma and the linear extent of involvement of carcinoma in needle biopsy were explicitly stated. A valid SNOMED CT expression could be constructed to assert the clinical statement with the exception of the numerical value of the measurement. Therefore, the linear measurement of the extent of carcinomas could not be recorded. This issue has been noted by IHTSDO and is currently under ballot for incorporation into the SNOMED CT concept model.¹⁴

Clinical expressions containing the positive presence of a morphologic abnormality and a pertinent absence of another morphologic abnormality required the conjugation of two situations with explicit context. As previously discussed, the statement ‘usual hyperplasia’ was represented in SNOMED CT (box 5) using a conjunction of situations with explicit context. One situation states the absence of atypia suspicious for malignancy, and the other situation explicitly lists the default situational context of a clinical finding, that is, the attribute–value pairings of 410510008|temporal context value|=410585006|current – unspecified| and 408732007|subject relationship context|=410604004|subject of record|.

It would seem to be an unnecessary burden to require a SNOMED CT expression database for surgical pathology to be maintained as |situation with explicit context (situation)| for all instances of clinical findings. However, in correspondence with the IHTSDO head terminologist, description logic constraints dictate that a clinical finding cannot be conjoined with a situation with explicit context, nor can the description logic classifier compute equivalence between a situation with explicit context and clinical finding. Situations with explicit context expressions can only be conjoined with other situations with explicit context expressions.¹⁵ The underlying logic constraint prohibits the conjoining of two concepts from different top-level hierarchies. The soft context (ie, default context) of the clinical finding hierarchy is that the finding is present, the subject is the patient, and the temporal context is current. However, |situation with explicit context| is a top-level hierarchy. Therefore, a clinical finding and a situation cannot be combined to create a single concept expression. A clinical finding, however, can be expressed in the situation hierarchy by explicitly stating the soft context, in which case it can be conjoined with another situation. This guideline is not described in the published SNOMED CT documentation and is likely a little known fact to SNOMED CT users. Post-coordinated databases which employ any uncertainty or statements of clinical absence must therefore include the additional attribute–value data for ALL clinical findings if they are to be supported by description logic query engines.

Using the terminology to assert presence, absence, negation, or temporal context invites robust debate concerning the proper role of a terminology model and that of an information model. On one extreme, terminologies specialize in the definition of concepts and the relationships between them. At the other end of the modeling spectrum, information models specialize in the management of definitions which often includes temporal and existential information. Between the two extremes, either modeling method can be employed. The level of success realized by either model is dependent on the particular use case and the binding of the terminology model and information model for that use case. SNOMED CT seeks to represent clinical concepts. Clinical concepts entail asserting the existence (or level of absence) of and the temporal context of clinical findings. IHTSDO includes a mechanism to represent this type of

information within the terminology. The scope of this study was to review the ability of SNOMED CT as a terminology, in its current state, to represent histopathology findings and not evaluate the merits of alternative approaches.

The construction of SNOMED CT expressions describing the histologic grade of an identified tumor was technically possible according to grammatical guidelines. This was done by pairing the associated morphology defining attribute with the abnormal morphology value of |intraductal carcinoma, noninfiltrating, no ICD-0 subtype| and using the defining attribute, |associated with|, and the clinical finding concept of |DCIS nuclear pleomorphism, grade 1: monotonous nuclei, 1.5–2.0 red blood cells diameters, with finely dispersed chromatin and only occasional nucleoli (finding)|. Using each attribute–value pair in a single, post-coordinated expression is allowed by the current clinical finding guidelines, but it is unclear if expressing nuclear grade as a finding concept is appropriate. Nuclear grade may better be represented as an observable entity concept. Histologic grade and Nottingham score represent measurement concepts, albeit with an amount of subjectivity, and therefore, would be better expressed as other measurement concepts. This is a definitional problem within the SNOMED CT release and has been communicated with IHTSDO for resolution.

Alternatively, each clinical finding could be expressed independently, that is, DCIS as one finding and histologic grade 1 as another, separate finding. This approach is syntactically straightforward but subject to ambiguous interpretation. DCIS is a clinical finding whose meaning is well understood logically and by the clinician. Histologic grade, however, represents a clinical finding that is meaningless without an associated abnormal morphology subject to grading (eg, DCIS). Therefore, the representation of each clinical finding independently of the other is not useful.

Employing the defining attributes for 404684003|Clinical finding (clinical finding)| consistent with the SNOMED CT model definitions is demonstrated in the case of invasive carcinoma. To assert that the invasive ductal carcinoma has associated DCIS (as reported in CAP’s cancer checklist), the post-coordinated expression for DCIS can be paired with the 47429007|associated with| attribute in a complex expression to assert the meaning of ‘invasive carcinoma with associated DCIS’ (box 2). Similar expression construction can be employed to express other Cancer Checklist data elements such as associated necrosis, microcalcifications, and lobular carcinoma in situ.

CONCLUSIONS

Despite the limitations discussed with using SNOMED CT expressions as a vehicle to describe diagnostic features noted by microscopic examination and whole slide imaging, the SNOMED CT international release in July 2012 was adequate to express pathologist interpretations of architectural findings for 75% of the stated statements listed in this study. Defined SNOMED CT concepts existed for each definitively described abnormal morphology in this set of breast biopsies. Furthermore, defined 442083009|Anatomical or acquired body structure (body structure)| attribute–value pairs could be identified or constructed using the current SNOMED CT conceptual content.

SNOMED CT in the July 2012 release did not allow for sufficient and specific expressions for descriptive statements of histologic findings described by the examining pathologists. This deficiency presents an issue to knowledge capture and representation for microscopic, anatomic pathology assessments. Descriptive, tissue architecture information recorded by the

pathologist represents the thought process of the diagnostician, and is therefore important to represent. The diagnostic thought process in combination with the resultant, definitive diagnoses represents the expression of knowledge of the physician.

Conclusive information by itself is valuable for categorization of findings (assuming the conclusions are correct), and descriptive findings alone provide data that may or may not have meaning. Both elements of information are necessary to represent knowledge. Continued development of the SNOMED CT concept model and conceptual content that pertains to microscopic examination of histologically prepared tissue specimens is required in order for the terminology to be effective in surgical pathology knowledge capture and knowledge management. A possible approach is to create a specialization of the concept 399984000[abnormal shape] within the [morphologically abnormal structure] hierarchy. Tissue and cellular formation, pattern, and other architectural descriptors can be subsumed within this new concept hierarchy. This approach can be developed within a local extension of SNOMED CT with a subsequent submission to IHTSDO for balloting consideration.

Detailed tagging of microscopic pathologic findings with a controlled terminology such as SNOMED CT is necessary to link tissue morphometrics with diagnostic conclusions. Encoded histopathology descriptions and findings support the aggregation and reuse of image findings that can be used for training residents, creating automated diagnostic systems, and conducting translational research using histologic imagery. However, expansion of the SNOMED concept space and current concept model to accommodate descriptive language is necessary for broad adoption of the terminology in the histopathology reporting process.

Acknowledgements The authors wish to express their thanks to David Markwell (IHTSDO) and Daniel Karlsson (Linköping University) for their assistance with SNOMED CT expression formation and formalisms.

Contributors SC conceived the study. SC, JC and SH designed the study and wrote the manuscript. WW and SH marked up and annotated histologic findings and validated SNOMED CT expression meaning. SC and JC identified and developed all SNOMED CT expressions. JM provided data regarding SNOMED CT concepts and alternative approaches for terminology expression, as well as, editorial input. All authors contributed to the final editorial process of the manuscript.

Competing interests None.

Provenance and peer review Not commissioned; externally peer reviewed.

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 3.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/3.0/>

REFERENCES

- 1 About Snomed. <http://www.ihtsdo.org/snomed-ct/snomed-ct0/>. Updated 2012.
- 2 Snomed CT. <http://www.connectingforhealth.nhs.uk/systemsandservices/data/ukkt/snomed>. Updated 2012.
- 3 Office of the National Coordinator for Health Information Technology, Department of Health and Human Services. Health information technology: initial set of standards, implementation specifications, and certification criteria for electronic health record technology. interim final rule. *Fed Regist* 2010;75:2013–47.
- 4 Lester SC, Bose S, Chen YY, et al. Protocol for the examination of specimens from patients with ductal carcinoma in situ of the breast. *Arch Pathol Lab Med* 2009;133:15–25.
- 5 Lester SC, Bose S, Chen YY, et al. Protocol for the examination of specimens from patients with invasive carcinoma of the breast. *Arch Pathol Lab Med* 2009;133:1515–38.
- 6 An overview of the College of American Pathologists Cancer Checklist. http://www.cap.org/apps/docs/committees/cancer/cancer_protocols/Overview_CAP_Cancer_Checklists_090115.pdf. Updated 2009.
- 7 Fletcher CDM. Symposium on errors, error reduction, and critical values in anatomic pathology. *Arch Pathol Lab Med* 2006;130:602–3.
- 8 Hamilton PW, van Diest PJ, Williams R, et al. Do we see what we think we see? The complexities of morphological assessment. *J Pathol* 2009;218:285–91.
- 9 Garcia Rojo M, Punys V, Slodkowska J, et al. Digital pathology in Europe: coordinating patient care and research efforts. *Stud Health Technol Inform* 2009;150:997–1001.
- 10 Daniel C, Booker D, Beckwith B, et al. Standards and specifications in pathology: image management, report management and terminology. *Stud Health Technol Inform* 2012;179:105–22.
- 11 Spackmen K, ed. *SNOMED CT editorial guide*. 3.20th edn. Copenhagen, Denmark: International Health Terminology Standards Development Organization, 2012.
- 12 International Health Terminology Standards Development Organization. *SNOMED-CT starter guide*. 2-22-2014 (US English) ed. Copenhagen: International Health Terminology Standards Development Organization, 2014.
- 13 International Health Terminology Standards Development Organization. *SNOMED-CT technical implementation guide*. July 2012 International Release (US English) ed. Copenhagen: International Health Terminology Standards Development Organization, 2012.
- 14 International Health Terminology Standards Development Organization. *RF2 specification change request: addition of concrete domains*. Change Request Date 2011 06 24 version 0.6 ed. Copenhagen: International Health Terminology Standards Development Organization, 2011.
- 15 Spackmen K. E-mail regarding post-coordination. 1 October 2012.