Author’s response to reviews

Title: Classifying the Precancers: A Meta Data Approach

Authors:
Dr Jules J Berman (bermanj@mail.nih.gov)
Donald E Henson (hensond@comcast.net)

Version: 2 Date: 15 May 2003

I thank both reviewers for reviewing the manuscript.

The first reviewer, Hans-Ulrich Prokosch, accepted the paper without revision.

The second reviewer, Catherine A. Smith had both compulsory and discretionary revisions. These revisions correct errors in the text, clarify text and improve the manuscript.

Under compulsory revisions, are:

1. The reviewer asked that the paper clarify that the abstract, to be published in Archives of Pathology, was the first public presentation of the precancer classification. This clarification is included in Acknowledgments:

This work was previously presented as an abstract to APIII, Advancing Pathology Informatics, Imaging and the Internet, October 2-4, 2002, in Pittsburgh, PA. The presentation abstract will appear in the Archives of Pathology and Laboratory Medicine in the fall, 2003.

2 Misspelling (clasification to classification) is corrected.

3. Colon is added on page 7.

4 (see also 5 and 6). The reviewer asks for clarification of the differences between UMLS and the metathesaurus. This is added to the revised version, in Methods, as shown:

The National Library of Medicine's UMLS (Unified Medical Language System) is a set of tools that facilitate the use of medical terminologies and the semantic relationships between terms and vocabularies. The UMLS Metathesaurus is one of three knowledge sources within the UMLS and contains concepts and terms from about 100 different medical vocabularies. The primary UMLS metathesaurus file used in the construction of the precancer terminology is MRCON. The 2003 MRCON file is over 150 Mbytes in length and contains over two million different terms belonging to nearly a million different concepts. MRCON and the entire UMLS metathesaurus are available at no cost from the National Library of Medicine at:

http://www.nlm.nih.gov/research/umls/

An example of some records from the MRCON file is shown:

C0004763|ENG|P|L0004763|VO|S1397347|Barretts Esophagus|0|
C0004763|ENG|P|L0004763|VO|S1397348|Esophagus, Barretts'|0|
C0004763|ENG|P|L0004763|VO|S1459012|barrets esophagus|2|
C0004763|ENG|P|L0004763|VO|S1940341|Barretts esophagus|0|
Notice that numerous variant terms for Barrett's esophagus all map to the same number in the first column, C0004763. This is the UMLS CUI (Concept Unique Identifier) for Barrett's esophagus.

The authors collected precancer terms from the UMLS. After review of the terms, the authors added supplemental terms from their own knowledge. Every additional term added by the authors matched a pre-existing UMLS concept. About 10% of the precancer synonyms were contributed by the authors.

After the terminology was assembled, the authors created a classification system and assigned each precancer term to one of the precancer classes. The entire classification was prepared as a metadata document using XML (eXtensible Markup Language) annotation.

5. The reviewer needs some clarification of what is new in the proposed terminology and what is re-used from existing terminologies. She recommends that we specify the source of terms.

Actually, the innovation in the article is the introduction of a classification of the [pre-existing] terms. About 90% of the terms were already in UMLS. The authors added several hundred terms that were, essentially, variant representations of concepts present in UMLS. The precancer terminology effort required expertise in pathology but was essentially an aggregation effort. Although it was a lot of work to gather the precancer terminology, it did not represent the innovative part of the paper. The creation of a a classification (the categories for the precancers) and the assignment of each precancer term to one of the biological categories was the innovative and intellectually challenging work of the paper. In other words, the terminology pre-existed in UMLS. The classification was entirely the work of the authors.

We attempted to clarify these points in Methods, as shown in the text shown in point 4, above.

6. The reviewer recommends that we clarify whether the new terms are matches to pre-existing metathesaurus concepts. They are. We added terms to the UMLS concepts and didn't need to add any new concepts. The revised version includes this point as shown in text for point 4, above.

Reviewer 2, Catherine Smith also advised discretionary revisions.

1. The reviewer notes that the use of the [two-word] term meta data is not common parlance. We agree that it is awkward. It seems self-evident that it should be one word (as in metaphysical, metamorphosis, metatarsal). Before submitting the original manuscript, I spent quite a while looking to see how different authors handled the term (meta data, metadata, meta-data). Surprisingly, most authorities used "meta data." But I just did a Google search on medatada, and many well-known sites use the one-word term. So, since the reviewer likes metadata, and we have always preferred metadata, I've changed the manuscript to use the one-word form consistently.

2. The reviewer asked that we explain the meaning of the flanking tags where they are first used in the manuscript. We do so in the revised manuscript, as follows:

In the example, five metadata tags are employed: <concept>, <cui>, <precancer_class>, <term>, <synonym>, along with and their corresponding closure tags (marked by a slash character). Because XML is case-sensitive, lower-case letters were consistently employed to simplify implementation. The <concept> tag indicates that a new concept will follow. Since all of the precancer concepts derive from or correspond to existing UMLS concepts, it was convenient to assign each precancer concept with the UMLS Concept Unique Identifier (CUI) and mark these with
a <cui> tag. Each precancer concept is assigned one of the precancer classes. In this case, the term "PTLD, post-transplant lymphoproliferative disorder" is assigned to the precancer class of "Acquired diffuse hyperplasias/metaplasias." The term is flanked by <term> tags and the class designation is flanked by <precancer_class> tags. Because the term is a synonymous variant, it is nested in <synonym> tags. Term and synonym tags are used for each of the term variants of the single concept.

I would like to add one comment. Both reviewers designated the manuscript as "A paper whose findings are important to those with closely related research interests." It is a conceit of every author that his/her paper has great significance to mankind. However, I think it's worth stressing that precancers are very important lesions. If we had a way of eliminating all precancers, we'd solve the whole cancer problem. On May 5, I attended an internal NCI (National Cancer Institute) organizational meeting on precancers. The issue of the classification came up in the context of 1) knowing which lesions are precancers, 2) correctly diagnosing precancers so that they don't report their findings on lesions that they thought was one lesion while in fact it was another, and 3) acquiring a classification that would serve as a scaffold for molecular biologic annotation and 4) being able to find relationships between heterogenous biomedical datasets that could be related to terms in the classification. This classification paper was mentioned throughout the meeting, and there was great interest in having a publication to serve as a baseline for additional work, ASAP. So, in my humble opinion, this paper has importance to anyone at-risk for cancer (i.e. everyone).

These changes satisfy every recommended revision. We hope that the re-submitted manuscript can be accepted for publication.