

ICD-11 日本語版コーディングツール作成に関する研究

研究分担者 今井 健 東京大学大学院医学系研究科

研究要旨

疾病及び関連保険問題の国際統計分類(ICD)の第 11 版改訂(ICD-11)が進められており、この国内適用へ向けた妥当性の検討が必要である。この際に適切な ICD-11 コードを検索する支援ツールの構築はこの検討作業を加速化すると共に、我が国における各種統計の基盤となる ICD コーディングの正確性担保、作業の効率化の観点から重要な課題である。本年度は ICD-11 日本語版コーディングツール作成にむけて、英語版の ICD-11 エンティティの訳出を効率化するツールの開発、並びに ICD-10,11 対応の日英索引表管理 Web プラットフォームの運用テストと課題抽出を行った。本年度成果によって WHO ICD-11 coding tool の日本語版言語拡張の実現に向けた作業効率化が図られると共に、今後取り組むべき技術的課題が明らかとなった。

A. 研究目的

疾病及び関連保健問題の国際統計分類(ICD)は、我が国では死亡統計を始め患者調査、医療保険制度、診療情報管理に広く活用されている。現在 WHO が進めている ICD-10 から 11 への改訂においては我が国の医療実態を踏まえたより適切な医療情報を将来的に確保するために、我が国での適用に向けた ICD-11 の妥当性の検討を行う必要がある。この際に、適切な ICD-11 コードを検索する支援ツールの構築は、これらの検討を加速化するのみならず、我が国における上記のような各種統計の基盤となる ICD コーディングの正確性の担保、作業の効率化の観点から極めて重要な課題である。現在、WHO は ICD-11 改訂に際し、従来の紙・冊子媒体に変わり Web ベースのアプリケーションである ICD-11 coding tool を開発しており、複雑化する ICD の構造を効率的に探索し、正確なコーディング

を支援する枠組みを整えつつある。本分担研究では、このような枠組みと整合性を担保しつつ、我が国における ICD-11 日本語版コーディングツールを作成することを目的としている。本年度は、ICD-11 日本語版コーディングツール作成にむけて、英語版の ICD-11 エンティティの訳出を効率化するツールの開発、並びにこれまで開発してきた ICD-10,11 対応の日英索引表管理 Web プラットフォームの実運用テストと課題抽出を行うことを目的とする。

B. 研究方法

ICD-11 日本語版コーディングツール作成のため、少なくとも英語版に存在する成句病名については、オーソライズされた和訳の作成が必要である。本年度は、厚生労働省政策統括官付参事官付国際分類情報管理室より、ICD-11 エンティティのカテゴリ (章) ごとに関連する各学

会へ日本語訳出の作業依頼を行った。この作業に際し、訳語の統一性・整合性の担保のため、医学用語対訳辞書を用いて日本語訳候補を導出する効率化ツールを開発した。この際には、英語版 ICD-11 エンティティの N-gram 分析を行い、頻出する表現を抽出し、これまでの我が国での慣用表現を対訳として追加した。

また、ICD-11 エンティティの日本語訳語作出のためにも、また WHO の修正勧告を適切にフォローアップしつつ ICD-10 から ICD-11 への病名单位でのスムーズな移行を管理するためにも、索引表の日英対応管理は重要な課題であるが、同室の協力を得て、これまでに開発してきた ICD-10,11 対応日英索引語管理システムの実運用テストを行い、課題点を抽出した。また、抽出された一部の課題については Web プラットフォームの改良を行った。

尚、本研究では倫理面への配慮は必要としない。

C. 研究結果

C-1) 訳語選定の効率化ツール

既に昨年度研究にて、ICD-11 日本語版コーディングツールの作成に向けては、約 97,000 語程度の ICD-11 エンティティのうち未だ仮訳が存在しない約 68,000 語(約 70%)に対する日本語用語を訳出する必要がある他、Post-Coordination 用の X Chapter、コーディングツールに特有なグローバルな言い換え可能語 (hepatic ⇔ liver) が必要であることが判明している。

近年では Google 翻訳を始めとした一般向けの Web ベースの翻訳サービスも充実化しているが、訳語の決定については関連学会の協力によるオーソライズが必要である。そこで本年度は成句としての日本語用語訳出については厚生労働省よりカテゴリ (章) ごとに関連する各学会へ作業依

頼を行った。結果は現在集約中であるが、最終的な和訳決定のためには全体的な整合性を考慮した修正が必要となる。特に ICD エンティティには「その他の明示された～」などのようにこれまで我が国に適用されてきた慣用的な表現があり、このような表現についてはカテゴリにまたがってなるべく統一されることが望ましい。そこで本年度はこのような訳語選定効率化のための支援ツールを作成した。図 1 に概要を示す。これは任意の英語フレーズに対し、形態素解析のコスト最小法と同じ方式で和訳語のトークン列に変換するものである。例えば図 1 のように

“infection by salmonella typhi” であれば、

- infection ⇔ 感染症,内寄生,汚染,伝染,感染
- by ⇔ による
- salmonella typhi ⇔ チフス菌

の各和訳語を組み合わせて表示する。辞書としては日本医学会医学用語辞典、日本医学用語シソーラス、MEID 辞書から抽出された日英対訳に、頻出する ICD 特有の慣用表現を加えた約 30 万語を用いた。

ICD翻訳チェック用辞書引きツール

元英語

- 例: 以下をコピーペーストして search を押してください
- infection by salmonella typhi
- Pyometritis
- Gastroenteritis or colitis of infectious origin
- Other specified intrahepatic cholestasis

infection by salmonella typhi

日本語

- 辞書検索

感染症<>内寄生<>汚染<>伝染<>感染 | ~による | チフス菌

- 事務局案検索

チフス菌感染症

注記

- 日本語検索は、英語のフレーズを分割し、訳語が当てられた部分を順に'|'で区切って表示しています。そのため訳語を生成する際には適宜語順を入れ替える必要があります。
- 訳語が複数ある場合は '<>' でつないでいます。

(図 1 : 訳語選定の効率化ツール)

元の英語フレーズによっては最終的に和訳語の語順を入れ替える必要があるが、確認者が目視で語順を入れ替えて訳語を決定するのは容易であることから大幅な効率化が可能である。また、既に存在している訳語がある場合は、事務局案として提示する機能もあり、これによって同一英語フレーズに対する訳語の揺れを防ぐことが可能である。現在確定した事務局案（既訳語）は約29,000語程度であるが、今後作業の進捗によってこれが増加するにつれ、一層訳語の揺れを防止することが可能となる。

C-2) ICD-10,11 対応日英索引語管理システムの運用テストによる課題抽出

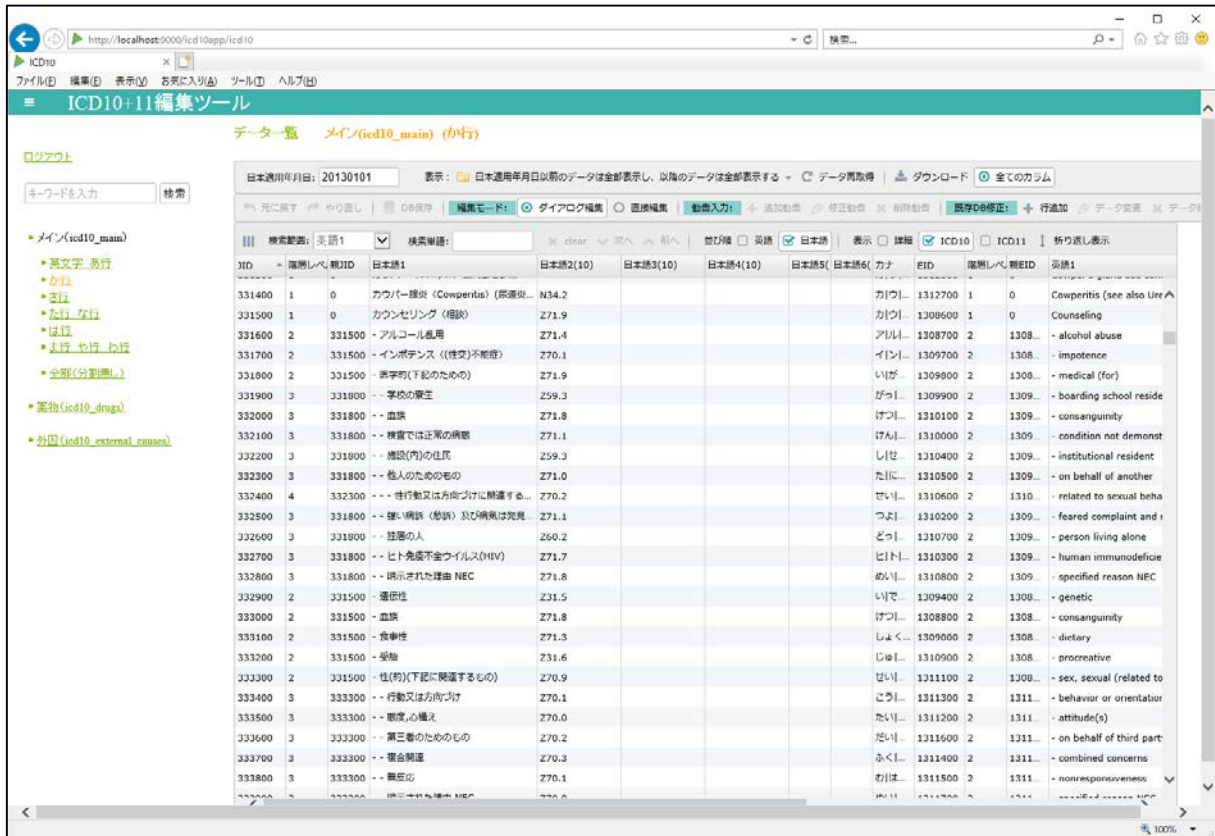
ICD-11では、内容例示表に多くの傷病名を収載する方針であるが、自由入力病名ではこれらに完全一致しないこともあり得る。このようなことから、これまでもICDでは、傷病名コーディングを支援するために、傷病名の部分文字列を元にコードをナビゲーションする「索引表」が存在している。ICD-11エンティティの日本語訳語作出のためにも、またWHOの修正勧告を適切にフォローアップしつつICD-10からICD-11への病名単位でのスムーズな移行を管理するためにも、この索引表の日英対応管理は重要である。本年度は、昨年度までに開発してきたICD-10,11対応日英索引語Web管理システムの実運用テストを行うと共に問題点の抽出、並びにユーザビリティ改善のためのプログラム修正を行った。運用テスト内容は(1) WHO ICD-10 2013年版に対する修正勧告の入力作業、(2) 国内専門委員会における修正意見内容の入力作業の2点である。

運用テストの結果、(1),(2)の実データに基づく

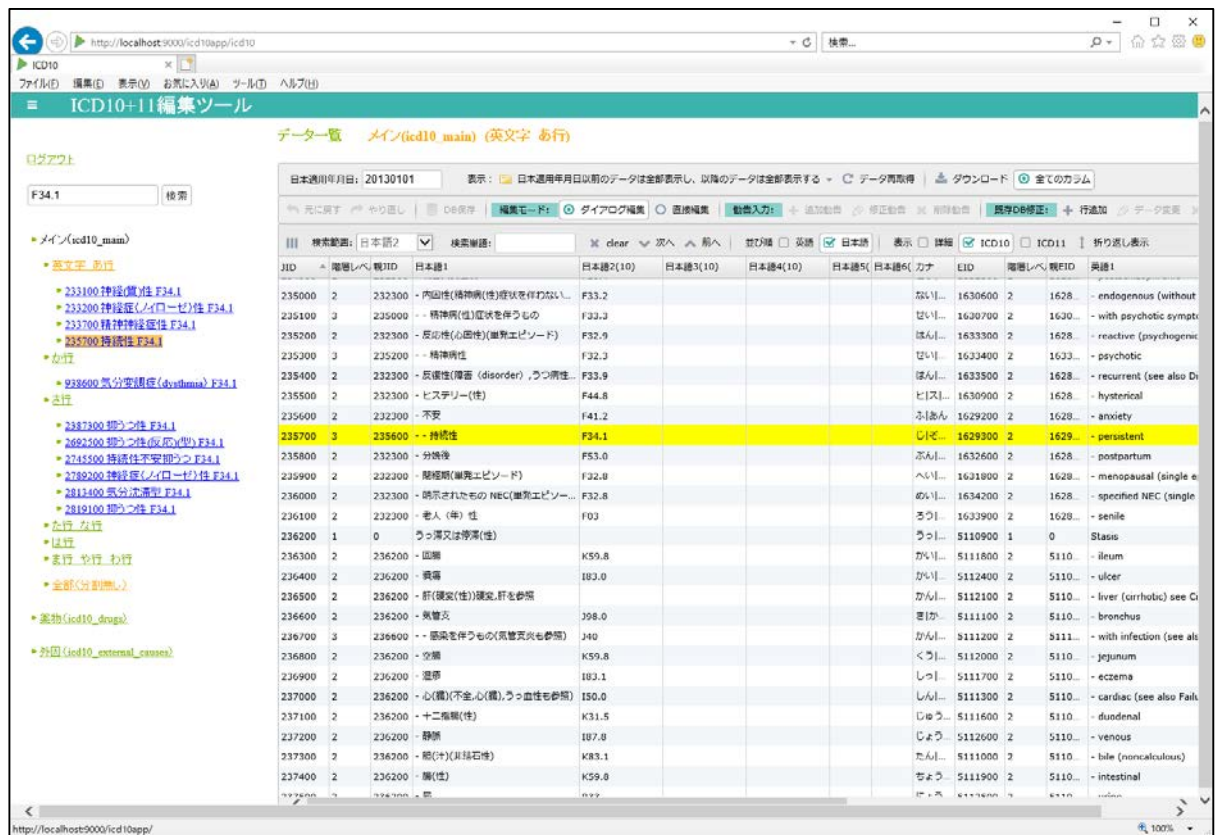
基本的な管理機能（勧告の入力・修正・削除・適用年月日を指定しての切り替え等）については問題が見られなかった。一方、今後さらなる円滑な運用のために必要と考えられた課題点は以下の通りであった。

- (a) 検索におけるユーザビリティ向上の必要性
 - 複数キーワードによる順次絞り込み、複合検索機能
 - 各索引用語の階層情報（親要素、並びに子要素）を指定した検索機能
 - 絞り込み検索時の索引用語の階層情報（親要素項目）の表示など。
- (b) 国内専門委員会における修正意見内容の入力においては、場合によって並び順が大幅に変更になることがあり、その管理が煩雑になってしまう点
 - 例えば「視力<視覚><Vision>」という親索引項目以下の一連の子要素索引語を、「視覚<Vision, visual>」と「視力<Vision, visual>」の2つに分離する場合、並び順IDの再設定が煩雑となるなど。
- (c) 実運用環境に起因する問題
 - 実運用環境（特定のブラウザ、端末性能）によってはさらなる処理速度の向上が必要である点

上記課題(a),(b)については、実運用に向けてさらなる索引語管理基盤の改良と運用手順の見直しが必要と考えられた。一方、課題(c)については、データベースを分割することで実運用環境ブラウザでの処理速度向上を図ると共に、複数データベースにまたがった検索機能を実装した。図2に分割されたデータベースの選択画面例、図3にこれらにまたがり検索を行う画面例を示す。



(図 2 : ICD-10,11 対応日英索引語 Web 管理システム ~索引語 DB の分割~)



(図 3 : 複数 DB にまたがった検索機能 ~ 例:“F34.1”の検索例~)

D. 考察

昨年度までの調査で、日本語版 ICD-10 コーディングツールの実現に向けては、運用上の観点、システム管理や整合性の観点から WHO ICD-11 Coding Tool に言語拡張の形で統合していく方式が最適であるとの感触を得ている。しかし今後これを実現するためにはいくつかの課題が存在している。

まず、約 97,000 語の ICD-11 エンティティに対する日本語訳の「統一性・整合性」の担保である。本年度、厚生労働省より各学会への日本語訳の依頼を経て、現在結果の集約中であるが、専門家の意見を尊重しながら、章ごとの訳出の揺れを極力抑える工夫と修正が必要である。英語版の ICD-11 エンティティに対する N-Gram 分析を行った結果からは、特に損傷、中毒及びその他の外因などの章 (ICD-11 22 章) を中心に頻度の多い N-Gram が抽出されている。例えば“exposure to or harmful effects of”(6-gram, 1,128 回)、“injured in collision with”(4-gram, 560 回) などである。もちろん、“Certain specified～”、“Without mention of～”など全ての章にまたがるような頻出表現も多数存在する。本年度は頻度の多い一部の頻出表現については訳語選定支援ツールのフレーズ辞書に加えたが、今後確定訳語の増加に伴い、頻出する表現に対する和訳をさらに追加することで、より一層の整合性向上とチェックの効率化を図ることが必要と考えられる。

また、これまで我が国で導入が進められてきた標準病名マスターとの整合性の確保も重要である。標準病名マスターは行政上の位置づけとしては、レセプト傷病名欄への記載として用いることが厚労省保険局医療課より通知されているとともに、2010 年には厚生労働省標準規格の 1 つとして指定されており、多くの病院で電子カルテに搭

載が進んでいる。これまでの ICD-10 は疾患カテゴリと(比較的少数の)例示病名を示したものであり、これに含まれない多くの病名については標準病名マスターがカバーしていた。しかし、ICD-11 では多くの病名を予め含んでいるため、現在進められている和訳作業でも、粒度の細かい日本語病名が数多く訳出されている。従って、診療情報とそれに基づく統計情報の継続性の観点から、今後標準病名マスターとの病名としての整合性の担保が課題となる。次年度では両者の一致と相違について分析すると共に、類似した病名の場合に標準病名を候補として推薦する機能を訳語選定支援ツールに実装する予定である。

さらに、WHO ICD-11 Coding Tool への日本語搭載にあたっては、成句としての和訳のみならず、形態素(病名の部分文字列)セットや、グローバルな言い換え可能語(hepatic ⇔ liver)など、検索エンジン特有の追加リソースも必要である。現在収集している和訳作業結果から可能な限り自動でこれらを収集することも次年度の課題である。またこれら得られたリソースを元に WHO の協力を得て次年度にて ICD-11 Coding Tool 日本語版のテスト公開を行う予定である。

一方、これらの索引語管理プラットフォームである ICD-10,11 対応日英索引語管理システムについては、実運用テストの結果、基本的な機能については運用上問題ないことが確認された。次年度は細かなユーザビリティ向上の点で、抽出された課題について改良を加えると共に、確定した ICD-11 コードを取り込んでいく予定である。

E. 結論

本年度研究では、ICD-11 日本語版コーディングツールの作成に向けて、英語版の ICD-11 エンティティの訳出を効率化するツールの開発、並びにこ

れまで開発してきた ICD-10,11 対応の日英索引表管理 Web プラットフォームの運用テストと課題抽出を行った。次年度では作成された成句病名の日本語訳と形態素リソースを元に、WHO の協力のもと、ICD-11 Coding Tool 日本語版のテスト公開を行うと共に、日英索引語管理システムの実運用を行う予定である。

F. 健康危険情報

なし

G. 研究発表

今井 健. ICD-11 の機能からみた我が国への適用について. 医療情報学 37(Suppl.):216-218, 2018.

H. 知的財産権の出願・登録状況

なし