

医学的研究およびICD改正に対応した 索引データの整備に関する検討

フジワラ ケン ジ
藤原 研司*

目的 ICD（通称：国際疾病分類）は、現在ICD-10までの改訂が行われており、さらにICD-11に向けて国際的に検討が進められている。WHOのICD改正・改訂作業に、より迅速に対応するため、英語と日本語を関連づけした電子化索引データを整備する。

方法 索引データの整備は、①WHOの英語版（ICD-10第2版）と日本語版（ICD-10（2003年版準拠））の索引の統合、②英語版にのみ記載された項目、日本語版にのみ記載された項目の特定、③階層構造をもつデータの同時検索システムの構築、④WHOのICD-10第2版以後の一部改正を作成したデータに追加の内容で行ない、英訳、和訳が統一的に行われているか等を検証した。

結果 索引データ（約63,000行）について、和訳、英訳が統一的に記載されたものとするとともに、階層構造のデータを一度に検索可能なデータとした。これにより、全体の約1割弱については、日本で独自に索引に追加したものであること、同一の英語であっても1次索引項目に続く項目によっては和訳が異なること等が判明した。

結論 今回整備されたデータは、ICDの分類としての基本構造を検討する際の基本的データとして利用可能であり、今後、ICD-11の分類構造について検討を行う際の問題点抽出等に大きな役割を果たすことができると考えられる。

キーワード ICD改正、電子化索引データ、索引データの整備

I はじめに

ICD（疾病及び関連保健問題の国際統計分類：通称国際疾病分類）は、現在、第10回改訂版のICD-10までの改訂（一部改正を含む）が行われており、さらにICD-11に向けて国際的に検討が進められている。ICDはわが国においても広範に活用され、国民の健康の保持、改善にとって医療の現場や政策の策定において極めて有用なデータ作成の基となっている。

ICD-10の出版物は、第1巻「総論」、第2巻「内容例示」、第3巻「索引」で構成され、現在、索引は、WHOから提示された英語版の

索引とこれに対応した日本語版の索引（製本および電子データ）がある。しかし、WHOから提示された英語版の索引と日本語版の索引との関連を同時にみることでできるデータは整備されていない。

日本語版の索引は、英語版の索引とは違った独自の修正を行ってきた歴史がある。例えば、日本独自の病名等もわずかながら「内容例示」（コード別に具体的な傷病名を記載）に「※印」を付して掲載されているが、索引には反映されていない。そのため、※印で明示されている独自の病名が索引では検索できない。また、NOS（“not otherwise specified”の略で、他に何らかの説明や記載がないものの意味）を伴う項目名は、英語版にはないが、日本語版索引に

* 横浜労災病院院長

は多数掲載されている。

日本語独自の索引の部分は、英語版の改正作業には現れないため、見落とされることが多く、作業を行う前にはこれらの問題点を解決するため、日本語と英語を同時にみることのできる索引を開発する必要があった。

しかし、WHOが勧告する一部改正を円滑に日本語版に反映し、確認するためには、索引データを英語に翻訳し、アルファベット順から五十音順に並び替える作業が必要となるが、日本語の単語と英語の単語が関連づけられた電子データが整備されていないため、担当者が紙ベースで確認作業をした上で整備をすることとなり、作業が大変複雑になり、情報提供に時間がかかっている。

現在使用している索引は、日本人としてはなじみにくい英語の階層構造をそのまま踏襲した階層構造で、病名を検索する際は、まず一番重要となる病態、次に病名を修飾する形容詞を探す必要がある。そのため、ある程度医学知識が必要となり、医学知識のないものが初めて索引を利用するのは難しく、索引を利用するための技術（こつ）を学ぶ必要があり、それを学ぶための教育機関が存在している状況である。

本研究では、英語と日本語に関連づけした電子化索引データを整備し、電子化データがより有効に活用できるように、保存データの形式も検討することで、現在行っているWHOのICD改正・改訂作業により正確で迅速に対応でき、さらにICDの有用性や普及の向上に寄与することを目的とする。

また、英語版を踏襲した日本人にはなじみにくい階層構造の索引から、日本人になじみのある辞書形式（例えば病名を一列表記にし、先頭の文字の五十音順にするなど）にする可能性についても検討を行う。

Ⅱ 方 法

本研究においては、電子的索引管理ツールの開発に関する研究を実施し、そのツールでは、①WHOの英語版（ICD-10第2版）と日本語

版（ICD-10（2003年版準拠））の索引の統合、②英語版にのみ記載された項目、日本語版にのみ記載された項目の特定、③階層構造をもつデータの同時検索システムの構築、④WHOのICD-10第2版以後の一部改正を作成したデータに追加を行うこととした。

また、このツールを使って、用語の英語訳と日本語訳をまとめて、英訳、和訳が統一に行われているか検証する。同時に、今後、WHOから出版されると予想されるICD-10第3版の日本語版作成に備えるため、このツールが実際に利用可能なものであるかを検証した。

さらに、ICDの各項目に該当する病名をマッピングする（1つのコードに包含されるべき病名を一覧にする）。この作業で、今まで、たびたび医療現場から指摘のあった「構造上項目は存在するが、実際には該当する病名が存在しないもの」を特定する。これは、ICDの構造の不備を実証することとなり、ICDの改善を行うための重要な基礎資料となる。

加えて、ツール開発に関連して日本人にとってなじみやすい表記（例えば辞書のような一列表記）に再構築することが可能か検討する。そのような索引表記が実現できれば、特別に研修を受けていなくても、いわゆる一般の辞書を引く作業と同様の作業で、病名を検索し、ICDコードを見つけることが可能となる。そこで、診療情報管理士の資格をもたない、病院職員に管理ツールを使って作業を行ってもらい、利用可能かどうかを検証した。

Ⅲ 結 果

（1）WHOの英語版（ICD-10第2版）と日本語版（ICD-10（2003年版準拠））の索引の統合

用語の英語訳と日本語訳をまとめて、英訳、和訳が統一的に記載される索引データを整備した（図1）。

このデータ整備では、日本語索引データ（第Ⅰ編本体、第Ⅱ編外因、第Ⅲ編薬剤の項目合わせて約63,000）について、今まで第1次索引項

（日本語索引のゴシック体部分）のおよそ1割程度しか英語訳を付していなかったものに、第2次、第3次項等含め全体の約9割強に英語訳を付すことができた。具体的には、日本語のエクセルシートと、英語のエクセルシートを並列させて、該当コードを一致させ、候補となる項目を絞り、その中から該当する項目を割り当てる作業を行った。これについては、医師で診療情報管理士の資格をもつ、ICDに詳しい者の協力を得て、英、和の対応が正しいか検証を行った。

全体の約1割弱については、日本で独自に索引に追加したものであることが判明した。これらには、日本語に対応するWHO原本の英語が存在しなかった。しかし、今後は、次回ICD改正時に漏れのないように対応するため、可能な限り参照となるような英語を付すことを、検討すべきと考えられる。

また、日本独自に、「内容例示」に追加した病名の中で、索引のコードが英語版とは相違するものがあることが明らかとなった。これらはWHOに対して、今後修正の提案を行うことを

検討すべき事項であると考えられる。なお、日本独自の病名であり、かつ、日本語版索引に記載されていない病名もあったが、これについても新たに電子的索引データに追加し、このツールのさらなる充実を図った。

（2） 英語版にのみ記載された項目、日本語版にのみ記載された項目の特定

英語版には、ラテン語やフランス語など英語以外の言語が引用されていたり、表現の違う英語が存在して、それを和訳すると英語は違っても同じ日本語になる項目があることも明らかになった。

（例）

○第1次索引項目に対する英訳が複数になる場合
遺棄（新生児） T74.0 Abandonment T74.0
Desertion (newborn) T74.0
胃痛 Gastralgia R10.1
Gastrodynia R10.1
エコノモ（von Economo-Cruchet）病 A85.8
Von Economo-Cruchet
disease A85.8
Economo's disease A85.8

○第2次項目等で英訳が複数になる場合
萎縮（atrophy）、萎縮性

－胃（性） K29.4 – gastric K29.4
– stomach K29.4
－眼球 H44.5 – eyeball H44.5
– globe H44.5
－副腎（被膜）（腺） E27.4 – adrenal
(capsule) (gland) E27.4
– suprarenal
(capsule) (gland) E27.4

また、英語の第1次索引項に続く項目を日本語訳した時にすべて同じ項目に導かれないもの

図1 英訳、和訳が統一的に記載される索引データの整備

現行の日本語索引データ

胃腸炎（急性）（流行性）（推定される感染（性）） （敗血症性）（腸炎および分類項目A09の注も参照）	A09
－アレルギー性	K52.2
－ウイルス（性）	A08.4
－ノロウィーク様	A08.1
－ロタウィルス	A08.0
－急性伝染性	A08.3
－重度、乳児の	A08.3
－小児（急性）	A08.3



英語訳を追加した索引データ

胃腸炎（急性）（流行性）（推定される感染（性）） （敗血症性）（腸炎および分類項目A09の注も参照）	A09	Gastroenteritis (acute) (epidemic) (presumed infectious) (septic) (see also Enteritis and note at category) A09
－アレルギー性	K52.2	－allergic
－ウイルス（性）	A08.4	－viral
－ノロウィーク様	A08.1	－Norwalk agent
－ロタウィルス	A08.0	－rotaviral
－急性伝染性	A08.3	－acute infectious
－重度、乳児の	A08.3	－severe, of infants
－小児（急性）	A08.3	－infantile (acute)
－明示された型 NEC	A08.3	－specified type NEC
－流行性	A08.1	－epidemic
－ノロウィーク型	A08.1	－type Norwalk

も存在した。例えば，“abuse”という語は、それに続く語が薬物、軟化剤、溶剤等の場合には、「乱用」と訳すが、小児、心理的、性的等が続く場合は、「虐待」と訳される。

このように、項目の特定ができたことは、今後ICDの改正・改訂の提案に対し、留意すべき用語が明らかになり、一括変換では対応できない部分の検証作業に、大変有効であると考えられた。

(3) 階層構造をもつデータの同時検索システムの構築

これまで階層構造をもつ索引データの検索は、ある病名を検索して、探したい用語はヒットしてもその属性（親、子）の関係が見つけられず、使用に耐えなかった。今回、既存のテキストデータ約64,000項目のすべてにヨミフリを付した。これは、本来ソフトがもっている自動のヨミフリ機能では、病名のような特殊な読みに対応できないことから階層構造のデータを一度に検索するためには、親、子、孫等の関係が明らかになるようにヨミフリを付しておく必要があるためである。結果、病名を第一次索引項目、第二次索引項目、第三次索引項目等、統合して検索することが可能となった。これにより、索引に掲載されているすべての病名を網羅して検索することができるようになった。

また、ツール開発に関連して日本人にとってなじみやすい表記（例えば一列表記）に再構築することが可能か検討した。英語と日本語では修飾語の語順が変わるため階層構造となった索引を第1次索引項、第2次索引項、第3次索引項とそのまゝ一列に並べるだけでは病名として成り立たないが、ある程度法則性をもって結合させることが可能となれば、医学的知識のない者であっても辞書を引く要領で、コーディングが可能と考えられた。しかし、英語を一列に並べて結合させた病名のうち医学用語辞典の英文とマッチングできたものは全体の約0.1%程度に過ぎなかったため、正しく結合させるための作業が膨大になることが予想されることから、本研究においては断念し、今後の検討課題とし

た。

(4) WHOのICD-10第2版以後の一部改正分を作成した電子的索引データに追加

ここでは、WHOのICD-10第2版以後の一部改正分を集約し、今回作成した電子的索引データへの追加を試みた。WHOから提示されたICD-10第2版以後の一部改正分は約10,000項目あるが、それらについて作成した電子的索引データに追加するためのファイル整理を行った。

- ① 一部改正の適用年を明示した。
- ② 項目、コードとも変遷が明確になるようにした。
- ③ 日本オリジナル項目を明記し、その際に参照可能な英文も出来るだけ付記した。
- ④ 英語索引と日本語索引を一対一対応させる際に、第1次索引項目を分解せざるを得ない項目（例：abuse）を明記した。

以上により、研究成果としての電子的索引データを作成した。

Ⅳ おわりに

WHOから発信されている情報によると、ICD-11以降の構造は、オントロジーを駆使した多軸となる可能性があるが、基本構造はICD-10を継承するとされている。本研究においては、分類構造の問題点を分析し、改正・改訂の提案に対応するために、索引データを電子化し、英語と日本語の対応関係を明確にすることができた。

また、日本語と英語の対訳作業により、日本語訳の不統一な部分が明確となり、今後は、統一的な翻訳作業を行うことが可能となった。ただし、管理上電子データと書籍版は、同一データである必要があることと、今回修正したもののすべてを書籍に反映させるとした場合、膨大な作業となり、現場に大変な混乱を引き起こすこととなると考えられた。改訂の作業が開始された現状においては、修正は、明らかな誤りであるものに留めておくべきと考えられた。検証作

業から得られた情報は蓄積し、改訂時に情報として、今後、ICD-11に反映させることができる重要な知見である。

さらに、ICD-11改訂に向けて、医療現場で使用する病名の標準化推進のために、ICDで例示する病名は、医学用語辞典に掲載されている用語にするようにとの要望については、ICDの索引に引用されている病名が、医学用語辞典の病名とどの程度の齟齬があるのかを検証する方法を模索していた。その結果、今回の研究成果により英語を仲介することで修正すべき病名の特定が可能となり、この作業を大きく前進させることができると考えられる。

ICD-11は医療情報の電子化された状況での活用を想定したものとなると発表されており、

分類項目にはそれを定義付けるための情報（用語）が付けられることになっている。これは、あらゆる医療用語から関連する分類項目を電子的に検索できる環境になると考えられ、紙ベースの索引の利用は今後大きく変わると考えられる。ただし、ICD-11以降も索引データの改正・改訂作業は続いていく可能性が高い。

最後に、研究成果が来るべきICD-11の改訂に具体的に活用されることを希望するものである。

謝辞

本研究に際してご協力いただいた、日本診療情報管理学会の方々に心から感謝いたします。