

<原著>

疾病および関連保健問題の国際統計分類第11回改訂版 (ICD-11) フィールドトライアルにおけるコーディング結果の一致性の評価

西大明美^{1,2)}, 木村映善^{2,3)}, 瀬戸僚馬¹⁾, 佐藤洋子^{2,4)}, 星佳芳²⁾, 緒方裕光⁵⁾, 水島洋²⁾

- ¹⁾ 東京医療保健大学医療保健学部医療情報学科
²⁾ 国立保健医療科学院研究情報支援研究センター
³⁾ 愛媛大学医学部医療情報学講座
⁴⁾ 静岡県立総合病院リサーチサポートセンター統計解析室
⁵⁾ 女子栄養大学疫学・生物統計学研究室

Assessment of consistency in the 11th revision of the International Statistical Classification of Diseases and related health problems (ICD-11) field trial coding results

NISHIO Akemi^{1,2)}, KIMURA Eizen^{2,3)}, SETO Ryoma¹⁾, SATO Yoko^{2,4)},
HOSHI Keika²⁾, OGATA Hiromitsu⁵⁾, MIZUSHIMA Hiroshi²⁾

- ¹⁾ Division of Healthcare Informatics, Faculty of Healthcare, Tokyo Healthcare University
²⁾ Center for Public Health Informatics, National Institute of Public Health
³⁾ Department of Medical Informatics, Medical School of Ehime University
⁴⁾ Division of Statistical Analysis, Research Support Center, Shizuoka General Hospital,
⁵⁾ Laboratory of Epidemiology and Biostatistics, Graduate School of Kagawa Nutrition University

抄録

目的 : ICD-10からICD-11への改訂において分類の構成が大きく変化し、エクステンションコード（重症度や解剖学的部位などの補助情報）が新設されるなど、コーディング内容に変化が生じており、コーディングの質に影響を与える可能性がある。本研究では、我が国で2017年に実施されたICD-11フィールドトライアルのコーディングのデータを用いて、WHOからゴールドスタンダード（GS）コードが提供された診断用語のコーディング結果の一致性について分析することを目的とした。具体的には、評価者のICD-10のコーディング結果をWHOのICD-10のGSと比較して正解率を算出し、同様に評価者のICD-11のコーディング結果をWHOのICD-11のGSと比較して正解率を算出した。その後、「ICD-10の正解/不正解」と「ICD-11の正解/不正解」の一致性を分析するために今回の研究を行った。

方法 : 我が国のICD-11フィールドトライアルは、2017年8月1日から9月10日を評価期間として、診療情報管理士298名が参加して実施された。その結果を基に診断用語コーディングのICD-10とICD-11のそれぞれの正解率およびICD-11の主病名の正解率を算出した。それらの結果と、コードの特徴との関連を検討した。コードの特徴としてGSコード数、主病名のコードの桁数、その他のコード（Yコード）や詳細不明コード（Zコード）の有無、エクステンションコードの有無をあげた。更に有効回答評価者の評価結果のICD-10/ICD-11のコーディングの一致性の評価としてGwet's AC1を用いて検討した。

連絡先：西大明美
〒154-8568 東京都世田谷区世田谷3-11-3
3-11-3 Setagaya, Setagaya, Tokyo 154-8568, Japan.
Tel: 03-5799-3711
Fax: 03-5799-3713
E-mail: a-nishio@thcu.ac.jp
[令和3年6月28日受理]

結果：ICD-11において正解率が高い診断用語は、GSコード数が少なく、その他のコード（Yコード）や詳細不明コード（Zコード）、エクステンションコードを必要としないものが多かった。ICD-11において正解率が5%未満と低い診断用語は、GSコード数が複数であり、主病名のコードの桁数が4～6桁と多く、その他のコード（Yコード）や詳細不明コード（Zコード）、エクステンションコードを必要とするものが多かった。サイトメガロウイルス性大腸炎の正解率はICD-10における36.55%からICD-11における89.85%と、19の診断用語の中で最も向上していた。これはICD-11においては、疾病ごとにコードが割り当てられたため、詳細な分類が可能となったことが影響している。Gwet's AC1値が0未満と低かった診断用語も、エクステンションコードを必要とするものが多かった。

結論：詳細な分類が可能となった診断用語は、正解率が高く、ICD-11への改訂が、コーディング結果への向上に繋がっていた。一方で、GSコード数が複数であり、エクステンションコードを必要とする診断用語は、正解率が低かった。今後、複数のコードやエクステンションコードの使い方について、十分な教育コンテンツを準備することが必要である。今回のフィールドトライアルは一部の資料を除き英語環境下での実施であったため、国内適用の前には、十分な研修と完全な日本語環境によるフィールドトライアルの実施が求められる。

キーワード：ICD-11, WHO, 診断用語コーディング, エクステンションコード, フィールドトライアル

Abstract

Objective: Significant changes were introduced to the International Classification of Diseases in the 11th revision (ICD-11). Moreover, extension codes (supplementary information such as the severity of conditions and anatomical sites) have been introduced, which may change the content of coding and affect its quality. This study by using the coding data from a field trial conducted in Japan in 2017, we aimed to analyze the consistency in the results of the coding of diagnostic terms for which the gold standard (GS) codes were provided by the WHO. Specifically, we calculated the percentage of correct answers by comparing the raters' coding results based on ICD-10 with the WHO GS for ICD-10 codes. Similarly, we calculated the percentage of correct answers by comparing the raters' coding results based on ICD-11 with the WHO gold standard for ICD-11 codes. Subsequently, we analyzed the consistency between "correct/incorrect answers based on ICD-10" and "correct/incorrect answers based on ICD-11" to determine the necessary measures for the implementation of accurate coding.

Methods: A field trial involving 298 health information managers was conducted in Japan from August 1 to September 10, 2017. We calculated the percentage of correct answers in the coding of diagnostic terms based on ICD-10 and ICD-11. Moreover, the percentage of correct answers in the coding of the main conditions based on the ICD-11 was also calculated. We examined the relationship between these results and the characteristics of the codes. The characteristics of the codes included the number of GS codes, the number of digits in the codes for main conditions, the presence/absence of other specified codes (Y codes)/ unspecified codes (Z codes), and the presence/absence of extension codes. In addition, Gwet's AC1 was used to evaluate the consistency between coding based on ICD-10 and that based on ICD-11 in the evaluation results of valid response raters.

Results: The percentage of correct answers for cytomegalovirus colitis improved from 36.55% (coded using ICD-10) to 89.85% (coded using ICD-11), the highest among the 19 diagnostic terms. This was due to the assignment of a code for each disease in ICD-11, which enabled detailed classification. There were many diagnostic terms with a high percentage of correct answers based on ICD-11, that had few GS codes, few digits in the code for the main condition, and did not require extension codes, other specified codes (Y codes), or unspecified codes (Z codes). Low percentages (<5%) of correct answers in the diagnosis coding based on ICD-11 were noted for diagnostic terms that required multiple codes or extension codes. Diagnostic terms with a low Gwet's AC1 value of less than 0 also required extension codes.

Conclusion: The percentage of correct answers was high for diagnostic terms for which a detailed classification was introduced. The 11th revision of the ICD led to improved coding results. In contrast, the percentage of correct answers was low for diagnostic terms that required multiple codes or extension codes. The results of this study indicate that it is necessary to prepare sufficient educational content on the use

of multiple codes and extensions. Since this field trial was conducted in an English environment (with the exception of some materials), sufficient training and a field trial conducted in a complete Japanese environment are required before the application of ICD-11 in Japan.

keywords: ICD-11, WHO, Morbidity Line Coding, Extension Code, Field Trial

(accepted for publication, June 28, 2021)

I. はじめに

疾病及び関連保健問題の国際統計分類第11回改訂版 (International Statistical Classification of Diseases and Related Health Problems ICD-11) は, WHOが2018年6月に公表し, 世界保健総会が2019年5月に採択した[1].

第10回改訂 (ICD-10) は当時の疾病, 障害および死因の分類のために, 詳細な疾病を特定するのではなく統計上のカテゴリー化をしていくという設計思想のもと開発されており, 1つのコードに対して様々な疾病が分類されるケースがある. ICD-10の「.9」は, 詳細不明の病態に使用される. また, 「.8」は, 多くの病態を含むその他の病態に使用される[2]. 例えば, Cytomegalovirus Colitis (1206: サイトメガロウイルス性大腸炎) のICD-10のゴールドスタンダード (Gold Standard:GS) として設定されていたB25.8はサイトメガロウイルス病で大腸炎だけでなく, 脊髄炎, 脳炎, 脳脊髄炎, 網膜炎, 網脈絡炎などの様々なサイトメガロウイルスによる疾病を表すその他のコードである.

ICD-11は基本的に疾病とコードが1対1の関係になるよう詳細な分類[3]となっている. 例えば, サイトメガロウイルス性大腸炎のICD-11のGSとして設定されていた1A24は, 大腸炎のみを表すコードである. また, コードの数はICD-10の約14,000からICD-11は約32,000と分類項目数が増えた[3]. 更に, 索引用語の大幅な増加, 章構成や扱う概念の増加, 分類軸や分類の構成, コード体系の変更など多岐にわたる相違点がある[1]. 新たに追加される章 (第4, 7, 17, 26, V, X章) のうち, X章:

エクステンションコードのコード数は, 約14,000である. エクステンションコードは重症度, 時間軸 (急性・慢性の別など), 病因 (感染病原菌), 解剖学的詳細 (部位), 組織病理 (新生物におけるより詳細な組織の分類) など詳細な補足情報を追加できるものであり, ICD-10にはなかった領域をカバーする分類の構成で, かなり大きな存在になる[3].

ICD-11のコード体系は, 1桁目 (章を表す英数字), 2桁目 (英字), 3桁目 (数字), 4桁目 (英数字), 5桁目 (英数字), 6桁目 (英数字), 7桁目 (英数) となっている[3].

Chronic obstructive pulmonary disease with pneumonia and HIV disease (1199: 肺炎及びHIV感染症に伴う慢性閉塞性肺疾患) のICD-11のGSは, CA12.2/CA20.Z/IC4Zである (図1, 表1). /で区切られているコードはシステムコードといて, 単独で使用できるコードであり, GSコード数が複数である. 主病名CA12.2のコードの桁数は5桁である. 桁数の数え方は厚生労働省の「ICD-

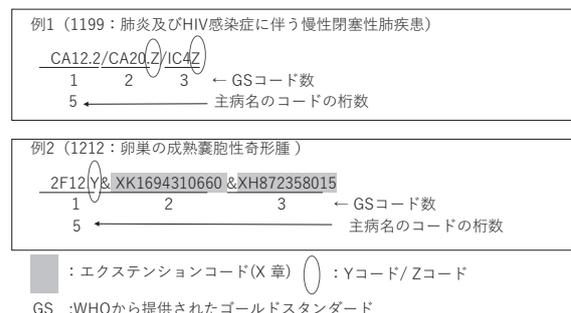


図1 ICD-11 コード

表1 GS (ICD-10およびICD-11) 診断用語

評価項目 NO	章	分野/ カテゴリー	診断用語		GS(ゴールドスタンダード)	
			英語	日本語	ICD-10	ICD-11
1192	I	Sepsis	Puerperal sepsis	産褥敗血症	O85	JA70.1
1193			Sepsis due to Escherichia coli (E. coli)	大腸菌による敗血症	A41.5	1B40
1194			Staphylococcus sepsis	ブドウ球菌敗血症	A41.2	1B46
1195			Methicillin-Resistant Staphylococcus Aureus (MRSA) Septicemia	メチシリン耐性ブドウ球菌 (MRSA) 敗血症	A41.0&U82.1	1B46&MJ81.11
1196			Sepsis due to urinary tract infection	尿路感染による敗血症	A41.9 N39.0	GC58.Z/MJ75.1
1197		Central line associated Escherichia coli (E. coli) sepsis.	大腸菌による中心静脈カテーテル関連敗血症	T82.7 A41.5 Y84.8	1B40/PG81.6/PH11.26	
1199	I	HIV	Chronic obstructive pulmonary disease with pneumonia and HIV disease	肺炎及び HIV 感染症に伴う慢性閉塞性肺疾患	B24 J44.0	CA12.2/CA20.Z/IC4Z
1200			Kaposi's sarcoma of the soft palate with AIDS	軟口蓋の AIDS 関連カポジ肉腫	B21.0 B24 C46.2	1C4Z/2B86.Y&XA1012292065
1201			AIDS-related dementia	AIDS 関連認知症	B22.0 B24 F02.4*	6C35.4/1C4Z/6C30.4
1202			Pneumocystis pneumonia (PCP) with AIDS	AIDS を伴うニューモシスチス肺炎 (PCP)	B20.6 B24 B59 J17.3*	CA20.31/1C4Z
1203		Wasting syndrome due to HIV	HIV 消耗性症候群	B22.2	1C4Z	
1204	II	Diarrhea	Clostridium difficile diarrhea	クロストリジウム・ディフィシル下痢	A04.7	1A04
1205			Acute gastroenteritis and dehydration	急性胃腸炎及び脱水症状	A09.9 E86	1A71&XT786106375/5C20.1
1206			Cytomegalovirus Colitis	サイトメガロウイルス性大腸炎	B25.8	1A24
1207	II	Neoplasm	Right breast angioliopoma	右乳房の血管脂肪腫	D17.1	2E60.1Z&XK876572005
1208			Moderately differentiated invasive adenocarcinoma of the duodenum	十二指腸の中分化型浸潤性腺癌	C17.0	2C05.11
1209			Pleomorphic adenoma of right parotid gland	右側耳下腺の多形腺腫	D11.0	2E71.1&XK876572005&XH1856218901
1210			Neurofibromatosis	神経線維腫症	Q85.0	1CID.1
1212			Mature teratoma of ovary	卵巣の成熟嚢胞性奇形腫	D39.1	2F12.Y&XK1694310660 &XH872358015

11のコード体系」の資料に基づいた[3]。CA20.ZとIC4Zの「Z」はICD-11では詳細不明のコードとなる。(図1, 表1)。

Mature teratoma of ovary (1212: 卵巣の成熟嚢胞性奇形腫)のICD-11のGSは2F12.Y&XK1694310660 &XH872358015である。&で続くコードXK1694310660やXH872358015はICD-11で新設されたエクステンションコードである。2F12.Yの「Y」はその他のコードである(図1, 表1)。

WHOは2017年に世界各国, 地域でフィールドトライアルを実施した[4]。我が国は世界保健機関国際分類ファミリー(World Health Organization Family of International Classification WHO-FIC)協力センターである国立保健医療科学院がフィールドトライアルセンター(field trial Center FTC)となり, 平成29年度厚生労働科学研究班「社会構造の変化を反映し医療・介護分野の施策立案に効果的に活用し得る国際統計分類の開発に関する研究(研究代表者 今村知明)」において2017年8月1日から9月10日を評価期間として, 実施された。調査結果は2017年10月のWHO年次総会で報告された[4,5]。

ICD-11の改訂におけるコーディングへの影響を調査するためのフィールドトライアルでは, ICD-11の医療現場における疾病分類への適用性, コーディング内容の信頼性, 統計上の有用性などを検討したうえで, 抽出された課題について解決・検討することが求められている[6]。フィールドトライアルは実際の医療現場でICDコーディングを行う実務者によって行われるコーディングを評価する。フィールドトライアルの結果は, 改訂によって生じるコーディング内容の変化が統計データに与える影響を検討する際に有用な情報となるだけでなく, 改訂された疾病分類の合理性を高めるためにも必要不可欠である[7-10]。これらはICD-10からICD-11への改訂において分類の構成が大きく変化するのに合わせてコーディング内容に変化が生じており, コーディングの質に影響を与える可能性がある[3]。ICD-11の改訂において, 詳細な分類となったことで明確になったこと, 複雑になったことによる影響, 更に, 分類の構成の変化, 新設されたエクステンションコードによる影響の評価が必要である。

国際的に共通のフィールドトライアルを実施するためにWHOからガイドライン(ICD-11リファレンスガイド)[11]やウェブベースのプラットフォーム(ICD Field Implementation Tool: ICD-FIT)[12]および診断用語セット(Diagnostic Term Set (morbidity) for Line Coding Testing 2017)とケースシナリオ(ICD-11 Morbidity Coding Field Test Case Scenarios)が提供される。

我が国のICD-11フィールドトライアルについては, 我々の報告で, ICD-11コードが, GSと一致することの影響として, GSコード数が少ないこと, ICD-10コードの一致率が高いこと, 個別評価で, 「難しくなかった」「詳細度がちょうどいい」と答えたことが同定された[9]。また, 複数のコードやエクステンションコードが必要となる診断用語について, 更に詳細な評価を行い, コーディ

ングルールを整備する必要がある[9]。GSが提供された診断用語について, ICD-10とICD-11の一致率は複数のコードやエクステンションコードが必要となる診断用語はいずれも5%未満であり, より詳細な分析が必要である[13], と述べている。

佐藤らの研究では, ICD-10とICD-11の「コーディングの難しさ」「コードの詳細さ」「コードの曖昧さ」が章ごとに異なることが明らかになり, 更に詳細な適用性の評価が必要[14]。各国および我が国のICD-11フィールドトライアルの実施概要が報告されたが, 参加国総数および内訳はWHOから公表されておらず, 明らかになっていない[4]。診断用語コーディングにおける分野ごとの「コーディングの難しさ」「コードの詳細さ」「コードの曖昧さ」をICD-10とICD-11で比較したところ, 総じてICD-10の評価が高かったが, ICD-11で新しく追加された章についても評価は良好であった[15], と報告されている。

上述したように, ICD-10からICD-11への改訂による影響が評価されているが, コーディングの質の向上のための具体的な方策を検討するためには, 個別の診断用語やコーディングにおける影響を詳細に評価する必要がある。そこで本研究では, 我が国で2017年に実施されたICD-11フィールドトライアルのコーディングのデータを用いて, WHO からGSコードが提供された診断用語のコーディング結果の一致性を分析することを目的とした。具体的には, 評価者のICD-10のコーディング結果をWHOのICD-10のGSと比較して正解率を算出し, 同様に評価者のICD-11のコーディング結果をWHOのICD-11のGSと比較して正解率を算出した。その後, 「ICD-10の正解/不正解」と「ICD-11の正解/不正解」の一致性を分析するために今回の研究を行った。

II. 方法

本研究では, 我が国のICD-11フィールドトライアルの結果を基に診断用語コーディングを分析した。フィールドトライアルは, 2017年8月1日から9月10日を評価期間として実施された。日本診療情報管理学会と日本病学会に所属している医療機関への広報を通して, 評価者となる診療情報管理士をオンラインでリクルートした。評価者のキャリアや経験を問わず, 診療情報管理士の資格取得者の中から300名以上を目標に参加者を募った。378名の評価者が応募し, 全員が参加を承諾した。ICD-FITを使用する際に, 事前登録が必要となり, メールアドレスを登録し, 評価者NO, パスワードを通知した。評価者は評価期間中に自分の都合のつく時間帯にフィールドトライアルの評価用システムICD-FITのWebサイトにログインして回答を入力した。

なお, ICD-11のコードは2017年8月時点のものであり現在のコードと一致しないものもある。

ICD-FIT上のコーディング作業は英語環境下で行われ

た。我が国は非英語圏であるため、平成29年度厚生労働科学研究班「社会構造の変化を反映し医療・介護分野の施策立案に効果的に活用し得る国際統計分類の開発に関する研究(研究代表者 今村知明)」[5]において、ローカライズを実施した。ICD-11リファレンスガイドを翻訳し、フィールドトライアルの評価者に日本語訳のPDFファイルを配布した。ICD-FITの個別評価項目と診断用語の英語と日本語の対照表を作成して提供した。

フィールドトライアルではWHOより、298の診断用語が提示され、評価者は、ICD-10とICD-11のそれぞれにおいてコーディングした。その結果の中からWHOよりGSが提供された19の診断用語のデータのみ抽出した。19の診断用語の分野内訳は感染症分野(第1章)が14件、新生物分野(第2章)が5件であった。感染症分野の内訳は敗血症分野が6件、ヒト免疫不全ウイルス(Human Immunodeficiency Virus, HIV)が5件、下痢分野が3件である(表1)。

なお、WHOからGSが提供された診断用語は19の診断用語のみであった。

1. ICD-10/ICD-11のコーディング正解率の分析および、コードの特徴との関連の検討

19の診断用語ごとの有効回答評価者数および、ICD-10とICD-11のそれぞれの正解率および、ICD-11における主病名の正解率を算出した。それらの結果と、コードの特徴との関連を検討した。コードの特徴として、GSコード数、主病名のコードの桁数、その他のコード(Yコード)や詳細不明コード(Zコード)の有無、エクステンションコードの有無をあげた。

2. ICD-10/ICD-11間の一致性の評価

有効回答評価者の評価結果のICD-10/ICD-11間のコーディングの一致性評価としてGwet's AC1[16-18]を用いて検討した。Gwet's AC1の値については-1から1の範囲の値を取り完全不一致から完全一致までを示す[19]。

3. 統計解析ソフト

統計解析はR v.3.5.1を用いた。Gwet's AC1値[16-18]の算出にはsirtパッケージ[20]を利用した。

本研究は、国立保健医療科学院倫理委員会の承認を得た(承認番号NIPH-IBRA#12308)

III. 結果

1. ICD-10/ICD-11のコーディング正解率の分析および、コードの特徴との関連の検討

診療情報管理士378名の評価者のうち未回答は80名であった。未回答を除き298名分を分析した。WHOよりGSが提供された19の診断用語に対する総回答数は3,959件であった。診断用語により回答数にばらつきがあった(表2)。

ICD-11において正解率が80%以上と高い診断用語は、Staphylococcus sepsis(1194:ブドウ球菌敗血症)等であった(表2)。ICD-11において正解率が5%未満と低い診断用語は、Pleomorphic adenoma of right parotid gland(1209:右側耳下腺の多形腺腫)等であった(表2)。ブドウ球菌敗血症等は、GSコード数が少なく、その他のコード(Yコード)や詳細不明コード(Zコード)、エクステンションコードを必要としないものが多かった(表2)。一方、右側耳下腺の多形腺腫等は、GSコード数が複数であり、

表2 ICD-10とICD-11のコーディング結果の一致性と正解率

評価項目NO	診断用語	n	ICD-10		ICD-11					ICD-10/ICD-11間の一致性 Gwet'sAC1
			正解率(%)	正解率(%)	主病名の正解率(%)	GSコード数	主病名のコードの桁数	YZコードの有無	エクステンションコードの有無	
1192	Puerperal sepsis	291	93.81	95.88	95.88	1	5			0.901
1193	Sepsis due to Escherichia coli (E. coli)	278	51.08	87.77	91.01	1	4			0.275
1194	Staphylococcus sepsis	275	52.36	80.36	87.27	1	4			0.126
1195	Methicillin-Resistant Staphylococcus Aureus (MRSA) Septicemia	253	11.86	31.23	56.52	2	4			0.636
1196	Sepsis due to urinary tract infection	238	18.07	1.68	65.97	2	5	✓		0.770
1197	Central line associated Escherichia coli (E. coli) sepsis	216	0.00	0.00	62.50	3	4			1
1199	Chronic obstructive pulmonary disease with pneumonia and HIV disease	194	0.52	0.00	2.58	3	5	✓		0.995
1200	Kaposi's sarcoma of the soft palate with AIDS	197	0.00	2.03	41.62	3	4	✓	✓	0.979
1201	AIDS-related dementia	195	0.00	0.51	69.74	3	5	✓		0.995
1202	Pneumocystis pneumonia (PCP) with AIDS	180	0.00	11.11	15.00	2	6	✓		0.876
1203	Wasting syndrome due to HIV	178	86.52	88.76	91.01	1	4	✓		0.756
1204	Clostridium difficile diarrhea	198	92.42	78.79	80.30	1	4			0.698
1205	Acute gastroenteritis and dehydration	185	18.92	1.08	65.95	3	4		✓	0.756
1206	Cytomegalovirus Colitis	197	36.55	89.85	92.89	1	4			-0.205
1207	Right breast angioliopoma	186	66.67	0.00	41.40	2	6	✓	✓	-0.200
1208	Moderately differentiated invasive adenocarcinoma of the duodenum	173	81.50	67.63	69.94	1	6			0.516
1209	Pleomorphic adenoma of right parotid gland	168	75.60	3.57	55.95	3	5		✓	-0.426
1210	Neurofibromatosis	182	66.48	82.42	86.81	1	5			0.441
1212	Mature teratoma of ovary	175	4.00	0.00	11.43	3	5	✓	✓	0.958

n:回答者数

Y:その他のコード

Z:詳細不明コード

主病名のコードの桁数が4~6桁と多く、その他のコード(Yコード)や詳細不明コード(Zコード), エクステンションコードを必要とするものが多かった(表2)。

2. ICD-10/ICD-11間の一致性の評価

ICD-10とICD-11の正解率に差がある診断用語は、サイトメガロウイルス性大腸炎、右側耳下腺の多形腺腫等であった(表2)。そのうち、ICD-11において正解率が向上した診断用語は、サイトメガロウイルス性大腸炎であった(表2)。一方、ICD-11において正解率が低くなっていた診断用語は右側耳下腺の多形腺腫等であった(表2)。サイトメガロウイルス性大腸炎については、ICD-10において、正解率が36.55%であったものが、ICD-11において89.85%と、19の診断用語の中で最も向上していた。ICD-11において、1A24として大腸炎のみを表すコードが割り当てられているという特徴があった(表2)。右側耳下腺の多形腺腫の正解率はICD-10においては、75.60%であったものが、ICD-11において3.57%と顕著に低くなっていた。ICD-11において、GSコード数が複数であり、主病名のコードの桁数も多く、その他のコード(Yコード)や詳細不明コード(Zコード), エクステンションコードを必要とするものが多かった。(表2)。

Gwet's AC1値が、0.901以上と高かった診断用語は、Puerperal sepsis (1192:産褥敗血症)、卵巣の成熟嚢胞性奇形腫等であった(表2)。そのうち、ICD-10とICD-11の正解率がどちらも90%以上と高かった診断用語は、産褥敗血症であった(表2)。一方、ICD-10とICD-11の正解率がどちらも5%未満と低かった診断用語は、卵巣の成熟嚢胞性奇形腫等であった(表2)。産褥敗血症については、ICD-11のGSコード数が少なく、その他のコード(Yコード)や詳細不明コード(Zコード), エクステンションコードを必要としないというコードの特徴があった。卵巣の成熟嚢胞性奇形腫等については、ICD-11のGSコード数が複数であり、その他のコード(Yコード)や詳細不明コード(Zコード), エクステンションコードを必要とするものが多いというコードの特徴があった。

Gwet's AC1値が、0未満と低かった診断用語は、サイトメガロウイルス性大腸炎、右側耳下腺の多形腺腫であった(表2)。そのうち、サイトメガロウイルス性大腸炎については、ICD-10において、正解率が36.55%であったものが、ICD-11において89.85%と、19の診断用語の中で最も向上していた。一方、右側耳下腺の多形腺腫の正解率はICD-10においては、75.60%であったものが、ICD-11において3.57%と顕著に低くなっていた。この理由は、前述した通りであると考えられる。

IV. 考察

この研究はICD-11のフィールドトライアルにおいて正確なコーディングが行えるかを検討した我が国で初めての論文である。評価者は298名と規模が大きく、19の診

断用語のコーディング結果の一致性について詳細な検討を行った。

フランスの研究は2人の評価者間のコーディング結果の一致率の分析[21]であった。つまり、今回の研究と全く同じ目的で実施されたフィールドトライアルは、このフランスの研究のみであった。現在公開されている研究のなかでは、正確なコーディングが行えるかを検討した世界でも最大規模のものである。WHO-FICネットワーク年次会議2016年~2020年における、ICD-11フィールドトライアルに関する研究発表(Poster Booklet WHO-FIC Network Annual Meeting)の報告国は、カナダ[22]、オーストラリアとニュージーランド[23]、韓国[24]であった。評価者はそれぞれ8名、49名、12名であった。各国と比較しても評価者数が多かったことで、Gwet's AC1値を算出することができ、より多くの知見を得ることができたと考える。

評価者のICD-10とICD-11のコーディング結果をWHOのICD-10とICD-11のGSと比較して正解率を算出したところ、ICD-10における正解率が低い診断用語は、その他のコードが付与されたものであった。しかし、それらの診断用語はICD-11においては、疾病ごとにコードが割り当てられており、正解率が高くなっていた。これは、ICD-11はコードの数が増えて、より詳細な疾病を特異的にコーディングできるようになったことに起因していると考えられる。ただし、ICD-11において正解率が高い診断用語は、GSコード数が少なく、その他のコード(Yコード)や詳細不明コード(Zコード), エクステンションコードを必要としないものが多かった。

カナダのICD-11フィールドトライアルでもICD-11の主病名のコードのみを比較した場合の正解率が79%であったのに対し、複数のコードを有する診断用語は、正解率が9%まで低下した[22]、と報告されている。韓国のICD-11フィールドトライアルのパイロットテストでも、全体評価において、その他のコード(Yコード)が十分に検討されていない[25]、と報告されている。

Gwet's AC1値が低かった診断用語は、ICD-10とICD-11の正解率に差があった。そのうちICD-10の正解率に比べICD-11において正解率が向上した診断用語は、疾病ごとにコードが割り当てられていた。一方、ICD-10の正解率に比べICD-11において正解率が顕著に低くなっているものは、GSコード数が複数であり、主病名のコードの桁数も多く、その他のコード(Yコード)や詳細不明コード(Zコード), ICD-11において新設されたエクステンションコードを必要とするものが多かった。

以上の結果より、ICD-11はより特異的なコーディングが可能になった反面、詳細なコードを必要とする正確なコーディングを行うためには、ICD-11リファレンスガイドの整備、教育プログラムの工夫が必要である。カナダのICD-11フィールドトライアルにおいても、ICD-11の一貫性や信頼性を高めるためにはポストコーディングの教育を充実させることが必要だと述べてい

る[22]. 韓国のICD-11フィールドトライアルにおいてもICD-10と大きく異なるため, 詳細なガイドラインが必要であると述べており[24], 日本でも検討が必要である. 複数のコードやエクステンションコードを必要とすることで正解率が非常に低い右側耳下腺の多形腺腫等, 詳細な分類が可能となったことで正解率が高いサイトメガロウイルス性大腸炎も同様にICD-11の分類の構成について研修を行えばより高い正解率が期待できる. 経験年数が浅い人にもコーディングの正解率が上がるように, 正解率の高い項目から回答を始めるようなコーディングの問題を作成する工夫が必要と思われる.

言語の壁については, ドイツのICD-11フィールドトライアルの論文においても, 研究の限界として挙げられている[26]. 我が国のICD-11フィールドトライアルでは, 研究班においてローカライズを実施し, ICD-11リファレンスガイドを翻訳し, フィールドトライアルの評価者に日本語訳のPDFファイルを配布した. ICD-FITの個別評価項目と診断用語の英語と日本語の対照表を作成して提供したが, ICD-FITは英語環境下での実施であったため, その影響が大きかったと考えられる. 短い診断用語であれば, その語句をコーディングツールの検索ウィンドーに入力すればコードを検索できる. また, その語句で検索できなくても, 適切な医学英語に読み替えることも容易である. 複数の単語で構成される疾患名に関しては, どの単語を検索語として選択するかについて迷い, 誤った判断に繋がる可能性がある. つまり適切なキーワードを選択できるかに大きく影響を受けたことが考えられる.

ICD-11では, オンラインのコーディングツール機能が提供されることが, 特徴の一つである. オーストラリアとニュージーランドの追加のフィールドトライアルでは, ICD-11のコーディングツール環境を強化し, 検索ルールの厳格さを軽減したことで, コーディングエラーが減少された[27], と報告されている. WHOからICD-11 Browser [28]とICD-11 coding tool[29]が公開され, 日々更新されており[30], キーワードの選択の影響については今後の改善が期待できる.

韓国のICD-11フィールドトライアルでは, ICD-11韓国語プログラムを提供してフィールドトライアルを実施されている. 評価者は, 母国語でコーディングができる環境であったにも関わらず, コーディングルールの変更や関連情報がない場合の正解率が低かった[24], と報告されている. これは, 言語の問題だけでなく, 詳細なガイドラインの提供や, 研修が必要であることを示している. WHOではICD-FITを使用した教育ツールの開発が進んでいる[3,31].

政策科学総合研究事業(政策科学推進研究事業)「社会構造の変化を踏まえた保健医療にかかる施策立案に資する国際疾病分類の国内導入のための研究(研究代表者小川俊夫)」では, わが国での翻訳作業が進みつつある. 現在, 各学会による意見等を集約し, 全体の和訳案

について, 統一的な方向性と整合性や学会間の誤差等を確認している[1]. それと同時に日本語での教育体制の構築も必要である. 更に, ICD-11のウェブツールすべて(ICD-11 Browser, ICD-11 Coding Tool, Foundation, ICD-FIT)の日本語化も必要である.

本研究の限界として, 4点を挙げる.

第一に, 我が国のICD-11フィールドトライアルで使用したICD-11のコードは2017年8月時点のものであり現在のコードと一致しない点である. つまり, 現在WHOから提供されているICD-11 Browserは, 情報が更新されており, 現段階の情報にて調査をした場合の結果が今回の結果と同様になるとは限らない.

第二に, 検索ツールであるICD-FITが英語環境下での実施であったため, 日本語にて実施した場合に同様の結果になるかは不明であるという点である.

第三に, WHOから示されたGSはわずかに19診断用語のみであり, その他の診断用語のコーディング結果の一致性については, 分析できていない点である. また, 19の診断用語は感染症分野(第1章), 新生物分野(第2章)のみであったため, 章ごとのコーディング結果の一致性も分析できていない. オーストラリアとニュージーランドのパイロットテストでは, ICD-10の一致率がわずかに優れていたが, 章によってばらつきがみられた[32], と報告されている. 章ごとのコーディング結果の一致性の分析も重要である.

第四に, WHOよりGSが提供された19の診断用語に対する総回答数は3,959件であった. 診断用語により回答数にばらつきがあり, データ数に偏りが生じ, 正解率に影響を及ぼした可能性がある. また, 378名の評価者のうち80名が未回答であった点でも同様の影響を及ぼした可能性がある. 各国のICD-11フィールドトライアルの報告では, ICD-10と大きく異なるため詳細なガイドラインが必要[24], 言語の壁 [26], ICD-11のトレーニングを検討する場合は, 対面とオンラインの併用が適切である[33], と挙げられている. 80名が未回答であった理由も同様であるとは限らない. しかし, ICD-11の導入に向けた正確なコーディングの実施のために必要な対策としては, コーディングルールの整備[9], 日本語での教育体制, 対面とオンラインの併用など事前教育を充実させることが重要となる.

ICD-11の国内適用の前に医療現場でのコーディング結果の信頼性を評価し, コーディング実務者の教育をすることで, 死因統計を含む公衆衛生や各種疫学研究などにICD改訂が与える影響を最小限にすることにより, 適切な移行を進めることができる.

その移行の為には, 今回の研究のような実務者による実証評価が欠かせない. この研究の結論として, 詳細な分類が可能となった診断用語は, 正解率が高く, ICD-11への改訂が, コーディング結果への向上に繋がっていることが見いだされた. 一方で, 複数のコードやエクステンションコードを必要とする診断用語は, 正解率が低

く、今後の研修等が必要であると考えられた。国内適用の前には、十分な研修と完全な日本語環境によるフィールドトライアルの実施が求められる。

本研究は、平成29年度厚生労働研究費補助金・政策科学総合研究事業（政策科学推進研究事業）「国際統計分類ファミリーに属する統計分類の改善や有用性の向上に資する研究」（201901001A 研究代表者：今村知明、研究分担者：水島洋。）による成果の一部である

この論文はMEDINFO2019 The Survey for Determining Knowledge-Related Problems in the Dissemination of ICD-11. NISHIO Akemi, KIMURA Eizen, SETO Ryoma, SATO Yoko, MIZUSHIMA Hiroshiの発表内容を論文化したものである。

利益相反

本稿において開示すべきCOI（利益相反）はない。

引用文献

- [1] 渡三佳, 山口雪子, 柳川侑子, 及川恵美子, 高橋恵介. 統計基準のICD-11 準拠のための改正対応の現状について. 第40回医療情報学連合大会（第21回日本医療情報学会学術大会）; 2020.11.18-23; 浜松. 医療情報学. 2020;40:305-306.
Watari M, Yamaguchi Y, Oikawa E, Yanagawa Y, Takahashi K. [The current situation of the revision work of the statistical standard conforming to ICD-11.] 40th Japan Association for Medical Informatics (21th Nihon Iryo Joho Gakkai Gakujutsu Taikai); 2020.11.18-23; hamamatsu. An Official Journal of the Japan Association for Medical Informatics. 2020;40:305-306.
- [2] 厚生労働省. 疾病, 傷害及び死因の統計分類提要 ICD-10 (2013年版) 準拠 第2巻 総論. 東京: 厚生労働統計協会; 2013.
Ministry of Health, Labour and Welfare. [Shippei, shogai oyobi shiin no tokei bunrui teiyō ICD-10(2013 nen ban) junkyo dai2kan soron.] Tokyo: Kosei Rodo Tokei Kyokai; 2013. (in Japanese)
- [3] 厚生労働省. 第7回社会保障審議会統計分科会疾病, 傷害及び死因分類部会資料2(2018.8.8). https://www.mhlw.go.jp/stf/shingi/shingi-hosho_164149.html (accessed 2020-12-27)
Ministry of Health, Labour and Welfare. [Dai 7 kai shakai hosho shingikai tokei bunkakai shippei, shogai oyobi shiin bunrui bukai shiryō (2018 nen 8 gatsu 8nichi).] https://www.mhlw.go.jp/stf/shingi/shingi-hosho_164149.html (in Japanese)(accessed 2020-12-27)
- [4] 佐藤洋子, 水島洋. ICD-11フィールドトライアルについて. 保健医療科学. 2018;67(5):508-517.
Sato Y, Mizushima H. [Field trial of ICD-11.] Journal of

the National Institute of Public Health. 2018; 67(5):508-517. (in Japanese)

- [5] 水島洋, 研究分担者. 厚生労働科学研究費補助金政策科学総合研究事業（政策科学推進研究事業）「国際統計分類ファミリーに属する統計分類の改善や有用性の向上に資する研究」平成29年度～令和元年度総括研究報告書. 2019.
Misuzhima H, kenkyu buntansha. Research on Policy Planning and Evaluation. Health, Labour and Welfare Policy Research Grants. [Kokusai tokei bunrui family ni zokusuru tokei bunrui no kaizen ya yuyosei no kojo ni shisuru kenkyu. Heisei 29nendo ~ reiwa1nendo sokatsu/buntan kenkyu hokokusho.] 2019. (in Japanese)
- [6] WHO. Report of ICD-11 Revision Review 14 April 2015. <https://www.who.int/classifications/icd/reportoftheicd11review14april2015.pdf> (accessed 2020-12-27)
- [7] Ellis PM, Welch G, Purdie GL, Mellsop GW. Australasian field trials of the mental and behavioural disorders section of the draft ICD-10. Aust New Zeal J Psychiatry. 1990;24(3):313-321.
- [8] Donada M, Kostanjsek N, Mea VD, Celik C, Jakob R. Piloting a collaborative web-based system for testing ICD-11. Stud Health Technol Inform. 2017;235:466-470.
- [9] 佐藤洋子, 水島洋, 木村映善, 西大明美, 緒方裕光. 我が国におけるICD-11フィールドトライアル—診断用語コーディングにおけるゴールドスタンダードの解析— 第38回医療情報学連合大会（第19回日本医療情報学会学術大会）; 2018.11.22-25; 福岡. 医療情報学. 2018;38(suppl):800-802.
Sato Y, Mizushima H, Kimura E, Nishio A, Ogata H. [Waga kuni ni okeru ICD-11 field trial: shindan yogo coding ni okeru gold standard no kaiseki.] 38th Japan Association for Medical Informatics. (19th Nihon Iryo Joho Gakkai Gakujutsu Taikai); 2018.11.22-25; Fukuoka. An Official Journal of the Japan Association for Medical Informatics. 2018;38(suppl):800-802. (in Japanese)
- [10] Raghuram R, Shamasundar C. ICD-10 field trials in India: A Report. Indian J Psychiatry. 1992;34(3):198-221.
- [11] ICD-11ReferenceGuide. <https://icd.who.int/icd11refguide/en/index.html> (accessed 2020-12-04)
- [12] ICD-FiT. v2.10.3. <https://icdfit.uniud.it/login.php>(accessed 2020-12-27)
- [13] Sato Y, Nishio A, Kimura E, Ogata H, Mizushima H. An evaluation of gold standards in line coding -ICD-11 field trial in Japan-. WHO-FIC Network Annual Meeting; 2019.10.6-11; Canada. Poster. C319.
- [14] Sato Y, Mizushima H, Ogata H, Mori K. ICD-11 Filed trial in Japan: An evaluation of a line coding. WHO-FIC Network Annual Meeting; 2018.10.22-27; Korea. Poster. C310.
- [15] 佐藤洋子, 水島洋. 我が国におけるICD-11フィール

- ドトライアル—診断用語コーディングの分野別解析— 厚生指標. 2019;66(8):31-37.
- Sato Y, Mizushima H. [Waga kuni ni okeru ICD-11 field trial sindan yogo coding no bunyabetu kaiseki]. kousei no sihyou. 2019;66(8):31-37. (in Japanese)
- [16] Gwet KL. Computing inter-rater reliability and its variance in the presence of high agreement. *British Journal of Mathematical and Statistical Psychology*. 2008;61:29-48.
- [17] Wongpakaran N, Wongpakaran T, Wedding D, Gwet KL. A comparison of Cohen's Kappa and Gwet's AC1 when calculating inter-rater reliability coefficients: a study conducted with personality disorder samples. *BMC Med Res Methodol*. 2013;13:61. doi: 10.1186/1471-2288-13-61.
- [18] 西浦博. 観察者間の診断の一致性を評価する頑健な統計量AC₁について. *日本放射線技術学会雑誌*. 2010;66(11):1485-1491.
- Hiroshi N. [A robust statistic AC1 for assessing Inter-observer Agreement in reliability studies.] *Nihon Hoshasen Gijutsu Gakkai Zasshi*. 2010;66(11):1485-1491. (in Japanese)
- [19] Wongpakaran N, Wongpakaran T, Wedding D, Gwet KL. A comparison of Cohen's Kappa and Gwet's AC1 when calculating inter-rater reliability coefficients: a study conducted with personality disorder samples. *Wongpakaran et al. BMC Medical Research Methodology*. <https://bmcmedresmethodol.biomedcentral.com/articles/10.1186/1471-2288-13-61> (accessed 2021-06-02)
- [20] Robitzsch A. sirt: Supplementary Item Response Theory Models. R package version 3.9-4. 2020. <https://CRAN.R-project.org/package=sirt>. (accessed 2020-12-04)
- [21] Tanno LK, Molinari N, Bruel S, Bourrain JL, Calderon MA, Aubas P, et al. Field-testing the new anaphylaxis' classification for the WHO International Classification of Diseases-11 revision. *Allergy*. 2017;1472(5):820-826.
- [22] Cullen D, Speckeen J, Burelle L, Wang J, Henderson T. ICD-11 Field Trials: A Canadian perspective on post-coordination. WHO-FIC Network Annual Meeting; 2017.10.16-21; Mexico. Poster. C306.
- [23] Macpherson B, Katte J, Hargreaves J. Phase 2 morbidity field testing of ICD-11MMS in Australia. WHO-FIC Network Annual Meeting; 2017.10.16-21; Mexico. Poster. C304.
- [24] Lee H, Lee Y. Report on field trial study of ICD-11(KM), Year3 in Korea. WHO-FIC Network Annual Meeting; 2020.10.19-23; virtual. Poster. C308
- [25] Kyung BS, Hyun HJ, Kyung BY, Woo HS, Hyun YJ, Ryon BS, et al. 2017 ICD-11 MMS field trial in Korea. WHO-FIC Network Annual Meeting; 2017.10.16-21; Mexico. Poster. C310.
- [26] Stausberg J. Quality of the ICD-11 beta draft from the German perspective: Evaluation based on the alphabet of ICD-10-GM 2017. *Methods of Information in Medicine*. 2018;57(4):168-176.
- [27] Katte J, Macpherson B, Hargreaves J. ICD-11 MMS field testing for morbidity phase 3 preliminary findings from Australia. WHO-FIC Network Annual Meeting; 2018.10.22-27; Korea. Poster. C309.
- [28] WHO. ICD-11 browser. <https://icd.who.int/browse11/l-m/en> (accessed 2020-12-27)
- [29] WHO. ICD-11 coding tool. https://icd.who.int/ct11/icd11_mms/en/release (accessed 2020-12-27)
- [30] WHO. ICD-11-Mortality and morbidity statistics). <https://icd.who.int/dev11/l-m/en/> (accessed 2020-12-27)
- [31] 水島洋, 研究分担者. 厚生労働科学研究費補助金政策科学総合研究事業 (政策科学推進研究事業)「国際統計分類ファミリーに属する統計分類の改善や有用性の向上に資する研究」令和元年度総括・分担研究報告書. 2020.
- Misuzhima H, kenkyu buntansha. Research on Policy Planning and Evaluation. Health, Labour and Welfare Policy Research Grants. [Kokusai tokei bunrui family ni zokusuru tokei bunrui no kaizen ya yuyosei no kojo ni shisuru kenkyu. Reiwa 1 nendo sokatsu / buntan kenkyu hokokusho.] 2020. (in Japanese)
- [32] Steinum O, Macpherson B, Boreklev M, Katte J. Pilot testing ICD-11: Reasons for disagreement in code allocation using the results from two field trial centres. WHO-FIC Network Annual Meeting; 2017.10.16-21; Mexico. Poster. C305.
- [33] Hinds A, Quesnel-Crooks S, Martin J. ICD-11 field trials in the Englishspeaking Caribbean. WHO-FIC Network Annual Meeting; 2017.10.16-21; Mexico. Poster. C307.