厚生労働行政推進調査事業費補助金(厚生労働科学特別研究事業)
分担研究報告書

患者調査と標準病名マスターの符号化方針相違事例における診療情報管理士による盲検下符号化の予備的検証

研究代表者　荒井　康夫　北里大学 未来工学部データサイエンス学科　准教授
研究分担者　石割　大範　国立健康危機管理研究機構 システム基盤整備局 係長

研究要旨

本研究班の前報により、令和 5 年（2023 年）患者調査で使用した辞書ファイル等と標準病名マスターの索引用語に対応する ICD コードとが相違する事例（符号化方針相違事例）の存在が確認された。本予備的検証では、ICD-10 章別の構成比を反映するように抽出した 124 症例を対象に、診療情報管理士 4 名（3 組）が、標準病名マスターを参照せず、患者調査 ICD と標準病名マスターICD の双方を併記したワークシート上で盲検下に ICD-10 コードを付与した。事例単位（n=124）の多数決一致先は、標準病名マスターICD が 80 件（64.5%、95%CI　55.8-72.4%）、患者調査 ICD が 34 件（27.4%、同：20.3-35.9%）、多数決不成立が 10 件 8.1%）であった。独立評価者対である A-B ペアの Cohen の κ 係数は 3 桁レベル 0.577（中等度）、章レベル 0.778（かなりの一致）であった。探索的所見として、術後状態 13 事例における原疾患章と Z90-Z99 との符号化方針の相違、および「コロナウイルス感染症」事例の方針差異が観察された。両方針は運用目的の違いに由来する構造的差異であり、本予備的検証は両方針の優劣を示すものではなく、今後検証規模および評価設計を拡張する必要がある。

研究協力者

小泉　麻美
　北里大学病院 医療の質評価室 診療情報管理士
村上　沙織
　北里大学病院 診療情報管理室 診療情報管理士
香取　麻紀子
　北里大学病院 診療情報管理室 診療情報管理士

A．研究目的

厚生労働省が実施する患者調査は、統計法（平成 19 年法律第 53 号）に基づく基幹統計のひとつであり、推計患者数や受療率の算出を通じて医療政策の立案・評価に用いられる重要な統計である[1]。患者調査では、医療機関が報告した自由記述の傷病名が WHO の「疾病及び関連保健問題の国際統計分類 第 10 回修正（ICD-10）」に準拠して符号化される[2]。一方、医療機関で広く用いられている標準病名マスター[3]は臨床用途を前提に整備されており、ICD-10 付与の判断基準が患者調査の符号化方針と一致しない場合がある。本研究班は前報により、令和 5 年（2023 年）患者調査で使用した辞書ファイル等と標準病名マスターの索引用語に対応する ICD コードとの突合により両者が一致しない事例の存在を確認した。

本予備的検証は、これら方針差異事例について、医療機関で符号化方針を担う専門家である診療情報管理士が標準病名マスターを参照しない条件下でいかなる符号を付与するかを実証的に検討することを目的とする。具体的には、（1）方針差異の

もとでの判定がいずれの方針側に偏る傾向にあるか、(2) 専門家間での判定の一致度はどの程度か、を明らかにする。本検証は、今後の本格的な検証研究に向けた実施可能性検討の位置づけで行った。

B．研究方法

1．対象事例の選定

令和 5 年患者調査で得られた傷病名の種類数(111,984 件)から、本研究班の前報の方法に従い、(i) 医療機関から報告された傷病名（以下、収集された傷病名）の正規化前文字列が標準病名マスターの索引用語と一致し、(ii) 患者調査において符号化された診断 ICD-10 準拠統計基準のコード（以下、符号化済み ICD コード）と当該索引用語に対応する標準病名マスター由来の ICD-10 コードとが相違する事例を「符号化方針相違事例」として抽出した。当該事例は前報において 1,184 件（索引一致かつ符号化済み ICD コードと ICD コード本体が相違する事例 746 件、および左右の指定差異により符号化済み ICD コードと ICD コード本体が相違する事例 438 件の合計）であった。

本検証では母集団 1,184 件の ICD-10 章別構成比を反映するように抽出した 124 症例を対象とし(標本割合 10.6%)、各事例には (ア) 患者調査 ICD (患者調査の符号化済み ICD コード) と (イ) 標準病名マスターICD (標準病名マスターの索引用語に対応する ICD-10 コード) の二つを併記したワークシートを作成した。なお、両者のいずれが患者調査由来かは、評価者には開示しなかった。

2．評価者と盲検性

診療情報管理士 4 名（うち 1 名は本分担研究者であるが、本検証用テストデータの作成には関与していない）が、標準病名マスターを参照しない条件下で、各事例に ICD-10 コードを手付けで付与した。担当区分は 3 組（うち 1 組は 2 名共同付与）とし、本報告では組の特定を避けるため、結果の記述では 3 組をそれぞれ「専門家 A」「専門家 B」「専門家 C」と表記する（割当順序は無作為)。

評価者には、各事例について「入力語」「索引用語」「正規化後表記」「病名表記」「参考 ICD-A」「参考 ICD-B」を併記したワークシートを配布し、別途設けた「確認結果」列に評価者が ICD-10 コードを記入する形式とした。参考 ICD-A と参考 ICD-B は、それぞれ患者調査の符号化済み ICD コードおよび標準病名マスター由来の ICD コードに対応するが、評価者にはいずれが患者調査由来かを開示しなかった（盲検評価)。評価者は標準病名マスターを参照せず、各自の判断により ICD-10 コードを付与することとした。なお、本ワークシート構成では参考 ICD-A・参考 ICD-B の双方が評価者に提示されているため、評価者の付与結果は両参考 ICD による選択の影響を受けうる。実際の集計でも、評価者の付与結果は対象 124 事例の約 90%（B：92.7%、C：91.9%、A：89.5%）が 4 桁レベルで参考 ICD-A または B のいずれかに完全一致しており、本検証は実質的に 2 候補併記下での盲検選択評価としての性格を有する。

3．評価指標と統計解析

主要評価指標は事例単位（n=124）の多数決一致先とした。各事例について 3 組の付与結果のうち 2 組以上が同一の参考符号と 3 桁レベルで一致した場合に多数決成立とみなし、対象 124 事例全体を分母として、患者調査 ICD 多数、標準病名マスター ICD 多数、多数決不成立の件数と割合および Wilson 法による 95%信頼区間を集計した。

本検証は予備的検証であり、仮説検定は行わない。多重比較の調整も行っていない。

副次評価指標として、各専門家による手付け ICD-10 コードと患者調査 ICD・標準病名マスター ICD との一致を 4 桁完全一致・3 桁一致・ICD-10 章一致の 3 水準で延べ観測（最大 370 観測）として集

計した。これらの観測は同一事例に対する反復測定であり独立性が成立しないため、95%信頼区間はWilson 法による粗推定に加え、事例クラスタを考慮したノンパラメトリック・ブートストラップ法（事例単位再抽出、評価者効果は固定、2,000 回反復）による区間を併記した。

専門家間の一致度はペアごとの Cohen の κ 係数および3組間の Fleiss の κ 係数で評価した[4][5]。ただし、共同付与の専門家 C は 2 名の合意ベース判定であるため、C を含む解析の κ 係数は独立評価者を仮定する前提に対し過大評価方向に偏りうる。本報告では独立評価者対である A-B ペアの Cohen の κ を主要解釈の対象とし、Fleiss の κ は参考値として併記する。κ 係数の 95%信頼区間はノンパラメトリック・ブートストラップ法（2,000 回反復）により算出し、解釈には Landis & Koch の基準[6]を参照した。一致率の 95%信頼区間は Wilson 法[7]により算出した。

感度分析として、共同付与の専門家 C を除外した A-B ペア（独立な評価者対）における両者一致事例の集計を併せて実施した。なお、本感度分析は両者一致事例における方針側分布の確認であり、全体集団における選好率を直接示すものではない。サブグループ別の集計（章別、術後状態、感染症章単一事例）は事後的に同定された記述統計であり、仮説検定は行わない。解析には Python 3.12（pandas、scikit-learn、statsmodels、NumPy、SciPy）を使用した。

（倫理面への配慮）本検証で用いたデータは令和 5 年（2023 年）患者調査で使用した辞書ファイル等であり、個人を特定し得る情報は含まれない。検証結果の集計にあたっては、専門家個人を特定可能な形での結果記述を避けるため、専門家 A・B・C の匿名表記とした。

C．研究結果

１．主要結果（事例単位、n=124）

対象 124 事例における事例単位の多数決一致先を表 1 に示す。多数決一致先が標準病名マスター ICD コードであった事例が 80 件（64.5%、95%CI：55.8-72.4%）、患者調査 ICD-10 準拠統計基準のコードであった事例が 34 件（27.4%、同：20.3-35.9%）、3 組すべてが 3 桁レベルで異なる付与（多数決不成立）が 10 件（8.1%）であった。本標本では多数決一致先として標準病名マスターICD コードが選ばれた事例の割合が患者調査 ICD の 2 倍を超える分布を示したが、これは予備標本における選択分布の記述にとどめ、両方針の優劣を示すものではない。

２．副次結果：評価者間の一致度

専門家ペアごとの一致率および Cohen の κ 係数を表 2 に示す。独立な評価者対である A-B ペアの Cohen の κ 係数は 3 桁レベル 0.577（中等度の一致）、章レベル 0.778（かなりの一致）であった。A-B ペアの一致率は 3 桁レベル 58.1%、章レベル 79.8%であった。専門家 C を含むペア（A-C、B-C）の κ 係数（3 桁レベル 0.661-0.694、章レベル 0.829-0.837）は A-B ペアより高かったが、専門家 C が 2 名共同による合意ベース判定であることに留意する必要がある。3 組間の Fleiss の κ 係数は 3 桁レベル 0.647、章レベル 0.819（参考値）であった。Fleiss の κ は独立な評価者を仮定するため、本検証では A-B ペアの Cohen の κ を主要解釈の対象とする。

専門家別の延べ観測（最大 370 観測）における標準病名マスターICD への一致は 246 件（66.5%）、患者調査 ICD への一致は 108 件（29.2%）であった。これらは同一事例に対する反復測定であり観測の独立性が成立しないため、Wilson 法による粗推定（標準病名マスターICD：61.5-71.1%／患者調査 ICD：24.8-34.0%）に加え、事例クラスタを考

慮したブートストラップ 95%信頼区間（標準病名マスターICD：60.2-72.2%／患者調査 ICD：23.5-35.0%）を併記する。クラスタを考慮した区間は粗推定よりも幅が広く、独立性の前提を置いた粗推定は不確実性を過小評価する。なお、本ブートストラップは事例クラスタのみを考慮し評価者効果は固定したため、評価者間の個性に伴う変動は反映されていない。

３．感度分析（副次解析：A-B ペアによる独立評価者対）

本感度分析は、独立評価者対である A-B ペアにおける主要結果の頑健性確認として位置づけられる副次解析である。A-B 両者の判定が 3 桁レベルで一致した事例は 72 件あり、両者が標準病名マスターICD で一致した事例は 57 件、両者が患者調査 ICD で一致した事例は 15 件であった。両者一致事例（n=72）における方針側分布の確認としては、両者標準病名マスターICD が 57/72 件（79.2%、95%CI: 68.4-86.9%）であった。なお、本値は両者一致事例における方針側分布の確認であり、両者不一致事例（52 件）の情報は反映していないため、全体集団における選好率を直接示すものではない。A・B の延べ観測（n=248）における標準病名マスターICD 一致は 161 件（64.9%）、患者調査 ICD 一致は 78 件（31.5%）であり、3 組合計の集計と整合する傾向を示した。

４．探索的所見

以下は事後的に同定された記述統計であり、仮説検定は行っていない。

＜対象事例の章別構成＞　対象 124 事例の患者調査 ICD に基づく ICD-10 章構成は、母集団 1,184 件の章別比率を反映するように抽出した結果に由来する。最も事例数の多い章は第Ⅱ章（新生物）28 件、第ⅩⅨ章（損傷・外因）23 件、第Ⅸ章（循環器系）10 件、第ⅩⅧ章（症状・徴候等）10 件、第ⅩⅠ章（消化器系）9 件であった。1-3 件の章が 8 章あり、第Ⅷ章（耳）、第ⅩⅥ章（周産期）、第ⅩⅩ章（外因）、第ⅩⅪ章（健康状態に影響を及ぼす要因）に該当する事例は対象に含まれなかった。各章の事例数の偏りは標本設計に由来し、本研究独自の発見ではない。章単位の統計的推測は本検証の標本設計の射程外である。

＜3 組の付与が分かれた事例＞　3 組の付与が 3 桁レベルですべて異なった事例は 10 件であった。これらには、（１）原疾患の章と症状の章の選択に幅があるもの（例：「肝門部癌」C22 と C24、「局所性痙攣」G40 と R25・R56）、（２）腫瘍の良悪性判定の差異（例：「椎間関節性腰痛症」M47 と M51）、（３）先天性と後天性の判定差異（例：「肺動脈弁下狭窄症」I37 と Q24、「副腎異常」E27 と Q89）、（４）損傷の章と臓器系の章の差異（例：「会陰部血腫」S30 と N30 と O71）が含まれた。専門家 C においては判定困難を理由とする自由記述記載が 3 件あり、いずれも判定不能の意思表示と解釈できる。

＜術後状態に関する事例＞　対象事例には、入力語に「○○術後」「○○術後状態」「○○術後遺症」「○○移植後」「○○留置（状態）」等の術後状態を記述する事例が 13 件存在した。これらの事例における患者調査 ICD は原疾患の章（第Ⅱ章 C 章、第Ⅸ章 I 章、第ⅩⅠ章 K 章等）または第ⅩⅧ章 R 章に符号化されている一方、標準病名マスターICD は多くの場合、第ⅩⅪ章 Z90-Z99（器官又は組織の後天的欠損及び植込み・置換術後の状態）に該当するコードに符号化されていた（例：「冠動脈バイパス術後」I251／Z951、「早期胃癌術後」C169／Z080、「前立腺癌術後状態」C61／Z988）。本検証における 3 組の付与は、これら術後事例の大半で標準病名マスターICD（Z 章）への一致を示した。

＜感染症章の単一事例＞　対象事例には、入力語「コロナウイルス感染症」が 1 件含まれた。当該事

例は、患者調査では新型もしくはCOVID-19の明記がなくともU071（COVID-19、ウイルスが特定されたもの）として符号化されていた。一方、標準病名マスターICDは入力文字列に対応する索引用語の登録に従いB342（コロナウイルス感染症、部位不明）であった。3組すべてが標準病名マスターICD側（B342系統）を選択した。

D．考察

１．方針差異事例における付与傾向

本予備的検証では、対象124事例における事例単位の多数決一致先のうち、標準病名マスターICDであった事例が64.5%（95%CI 55.8-72.4%）、患者調査ICDであった事例が27.4%（同：20.3-35.9%）、多数決不成立が8.1%であり、両割合の差は約37ポイントであった。本割合差は予備標本における選択分布の記述にとどまり、（１）一般集団における方針差異の存在を示すものではなく、（２）両方針のいずれが妥当であるかの判定材料となるものでもない。

本検証の評価者は標準病名マスターを参照しない条件下で判定したが、ワークシートには参考ICD-A（患者調査の符号化済みICDコード）と参考ICD-B（標準病名マスター由来のICDコード）が併記されていたため、評価者の判定はこれら2候補に強く影響された可能性がある。実際の付与結果では、評価者の付与の約90%が4桁レベルで参考ICD-AまたはBのいずれかと完全一致しており、本検証は実質的に2候補併記下での盲検選択評価としての性格を持つ。したがって、本検証で観察された割合差は、「専門家が独立に思考した結果としてB選好が出た」という解釈には直接つながらず、「2候補併記下において評価者がBを選びやすい傾向にあった」という限定的な解釈にとどめるのが妥当である。

患者調査の符号化方針は、統計法に基づく基幹統計として、WHOのICD-10分類原則に準拠しつつ、疾病別の医療需要および受療動向の把握を目的として整備されたものである。これに対し、標準病名マスターは電子カルテ等での臨床用途を主目的に整備された病名集であり、索引用語とICD-10との対応関係はその目的に沿って構築されている。両方針はそれぞれの目的に照らせばいずれも妥当であり、方針差異は優劣の問題ではなく、運用目的の違いに由来する構造的な差異である。

本検証で観察された標準病名マスターICDへの傾きは、診療情報管理士が日常業務において標準病名マスターを介してICD-10コードを取り扱うこととの関連で理解できる所見であり、いずれの方針が妥当であるかを示すものではない。

２．専門家間の一致度

本検証では、独立評価者対であるA-BペアのCohenのκ係数を主要指標とし、これは3桁レベル0.577（中等度の一致）、章レベル0.778（かなりの一致）であった[6]。ただし、本検証は独立評価者2名（A・B）に基づく予備的観察であり、評価者数および標本数の制約から、専門家間の一致度に関する本検証の評価は暫定的なものにとどまる。専門家Cを含むペア（A-C、B-C）のκ係数（3桁レベル0.661-0.694）はA-Bペアより高かったが、Cが2名の合意ベース判定であるため一致度が高めに出る方向の偏りが含まれうる。Fleissのκ係数（3桁レベル0.647、章レベル0.819）は独立な評価者を仮定する前提に対し過大評価方向に偏りうるため、参考値として位置づけられる。

章レベルの高い一致（A-Bペアκ=0.778）に対し3桁レベルでは0.577と相対的に低下しており、章の選択（疾患群の所属）には共通した判断枠組みが存在する一方、3桁分類への落とし込み（特定病態の選択）では複数の妥当な選択肢が併存し判断が分かれ得ることを示している。3組すべてが異なる

付与をした 11 事例には、原疾患か症状か、先天性か後天性か、損傷の章か臓器系の章かといった ICD-10 運用上の論点が含まれており、これは方針差異とは別の、原資料の情報不足や複数の合理的解釈の併存に由来する判定の困難性を示す。

これらの所見は、収集された傷病名のみからは適切な ICD-10 コードの選択が困難な事例が存在することを示すものであり、より適切な符号化のためには原資料となる診療情報を確認する必要性が示唆される。

３．標本設計と章別分析の解釈

対象事例の章別構成は、母集団 1,184 件の章別出現比率を反映するように抽出して決定した。これは母集団全体の傾向を縮約的に観察し、実態に近い条件下で予備的所見を得ることを優先した結果である。この設計により、第Ⅰ章、第Ⅲ章、第ⅩⅫ章等は事例数が 1 件にとどまり、第Ⅴ章・第Ⅵ章・第Ⅶ章・第Ⅹ章・第Ⅻ章は 2-3 件と少数となった。これらの章における選択件数は当該章の傾向を一般化する根拠としては不十分であり、章単位での統計的推測（章ごとの選好割合の信頼区間推定や章間比較等）は本検証の標本設計では行わない。章別の傾向を統計的に評価するには、章ごとに必要標本数を確保した次段階の検証が必要である。

４．術後状態に関する方針差異

対象事例には、入力語に術後状態を記述する事例が 13 件含まれた。これらの事例において、患者調査 ICD は原疾患の章（第Ⅱ章 C 章、第Ⅸ章Ⅰ章、第Ⅺ章 K 章等）または第ⅩⅧ章 R 章に符号化されていた一方、標準病名マスターICD は第ⅩⅪ章 Z90-Z99（器官又は組織の後天的欠損及び植込み・置換術後の状態）に該当するコードに符号化されていた。

患者調査における原疾患章への符号化は、基幹統計として疾病別の医療需要および受療動向を把握するという当該調査の目的に沿った方針運用と理解できる（例えば「早期胃癌術後」を C169 に符号化することで、胃癌に関する受療動向の集計に当該症例が反映される）。標準病名マスターにおける Z90-Z99 への符号化は、電子カルテ等の臨床用途における手術後の継続的な状態を表す概念分類として運用されている。両者はそれぞれ異なる目的・前提に基づく方針であり、いずれもその方針内で整合する符号化結果である。本検証における 3 組の手付け付与は、これら術後事例の大半で標準病名マスターICD への一致を示した。

５．コロナウイルス感染症の符号化

入力語「コロナウイルス感染症」の事例は、患者調査では新型もしくは COVID-19 の明記がなくとも U071（COVID-19、ウイルスが特定されたもの）として符号化されていた。一方、標準病名マスターICD は入力文字列に対応する索引用語の登録に従い B342（コロナウイルス感染症、部位不明）であった。本検証のワークシートでは U071 と B342 の双方が参考 ICD として併記された状態で評価者に提示され、3 組すべてが参考 ICD-B 側（B342）を選択した。

患者調査における U071 への符号化は、新興感染症の発生状況を統計上把握するという当該調査の方針運用に基づく。標準病名マスターにおける B342 への符号化は、入力文字列に対応する索引用語との関係に基づく方針運用である。両者はそれぞれの方針および目的に基づく符号化結果であり、本検証ではその相違が観察された。本事例は単一事例であり、検証規模を拡張した上で、同種事例の出現状況および各方針における取り扱いを引き続き観察することが、今後の検討項目となる。

６．本検証の限界と今後の展望

本検証は予備的なものであり、以下の限界がある。

＜標本に関する限界＞　対象は母集団 1,184 件のうち章別構成比を反映するように抽出した 124

件にとどまり、母集団全体の傾向を統計的に推定するには標本数が十分でない。事例数 1-3 件の章が 8 章を占めるため章別の統計的推測は本検証の標本設計の射程外であり、章単位での結論は導出しない。

＜評価者に関する限界＞　評価者は 3 組（4 名）に限られ評価者数による不確実性が大きく、Fleiss の κ 係数は評価者の代表性を仮定するが、本検証は便宜標本による評価者対であり、より広範な診療情報管理士集団への一般化可能性は本検証では確立されない。3 組のうち 1 組は 2 名共同付与のため、他の 2 組（各 1 名）と判定の独立性が完全には等価でなく、評価者単位の解析（Fleiss の κ）にも影響しうる。共同付与の専門家 C は 2 名の合意ベース判定であり、個別データが記録されていないため、4 者として分離した感度分析は本検証では不能であった。

＜統計処理に関する限界＞　本検証の主要結果は事例単位（n=124）の多数決一致先による割合と Wilson 法 95%信頼区間で構成されるが、専門家別の延べ観測（最大 370 観測）は同一事例に対する反復測定であり観測の独立性が成立しないため、独立性を仮定した Wilson 法による信頼区間は不確実性を過小評価する。本報告では延べ観測について事例クラスタを考慮したブートストラップ 95%信頼区間を併記したが、本ブートストラップは事例クラスタのみを再抽出し評価者効果は固定したため、評価者間の個性に伴う変動は反映されていない。より厳密には階層モデル（GLMM）による評価者効果と事例効果の同時推定が必要であり、これは次段階の課題である。

本検証では仮説検定は行わず、効果量（割合）と 95%信頼区間による記述統計的提示にとどめた。本検証は評価者が日常業務で標準病名マスターを介して ICD-10 コードを扱う診療情報管理士による評価であり、かつワークシート上に参考 ICD-A および B が併記されているため、「両方針への一致率が等しい」とする帰無仮説は構造上成立しにくく、形式的な仮説検定は実質的な意味を限定的にしか持たない。

感度分析（A-B ペアによる独立評価者対）は、両者一致した 72 事例における方針側分布の確認であり、両者不一致 52 事例の情報は反映していない。したがって、当該感度分析の数値（両者 B 79.2%）は両者一致事例における選好分布の記述であって、全体集団における選好率を直接示すものではない。Fleiss の κ 係数（0.647）は専門家 C が 2 名共同付与であるため独立評価者を仮定する前提に対し過大評価方向に偏りうる。本検証では独立評価者対である A-B ペアの Cohen の κ（3 桁 0.577、章 0.778）を主要指標とした。

本検証は予備的検証のため多重比較の調整は行っておらず、サブグループ別の集計（章別、術後状態、感染症章単一事例）は事後的に同定された記述統計として位置づけられる。多数決不成立事例（n=10）等の小数事例における信頼区間の精度は限定的である。

＜評価設計に関する限界＞　評価者は標準病名マスターを参照しない条件としたが、長年の業務を通じて獲得した符号化判断の枠組みそのものを統制することは方法論上困難である。また本検証では、ワークシートに参考 ICD-A／参考 ICD-B の二候補を併記し評価者にいずれが患者調査由来かを伏せた盲検評価を採用したが、二候補の併記自体が評価者の連想を誘導する可能性は残る。自由記述方式での独立評価の併用や、別形式での提示による感度分析は今後の課題である。

＜今後の展望＞　今後は対象事例数および評価者数を拡大し、独立評価を原則とする評価設計のもとで、章別の事例数バランスの改善、診療情報管

理士以外の関連職種との比較、方針差異事例における判定根拠の質的分析、術後状態およびコロナウイルス感染症等のように方針差異が生じる領域における観察の継続、を含む本格的な検証研究が必要である。

E．結論

令和 5 年（2023 年）患者調査で使用した辞書ファイル等と標準病名マスターの符号化方針が乖離する 124 事例を対象に、診療情報管理士 4 名(3 組)が標準病名マスターを参照しない条件下で、参考 ICD-A（患者調査 ICD）および参考 ICD-B（標準病名マスターICD）の双方が併記されたワークシート上で盲検下に ICD-10 コードを付与した結果、対象 124 事例における事例単位の多数決一致先は、標準病名マスターICD が 80 件（64.5%、95%CI 55.8-72.4%）、患者調査 ICD が 34 件（27.4%、同：20.3-35.9%）、多数決不成立が 8.1%であった。独立評価者対（A-B ペア）における Cohen の κ 係数は 3 桁レベル 0.577（中等度の一致）、章レベル 0.778（かなりの一致）であった。なお、評価者の付与結果は対象事例の約 90%が 4 桁レベルで参考 ICD-A または B のいずれかに完全一致しており、本検証は実質的に 2 候補併記下での盲検選択評価としての性格を有する。したがって本結果は、2 候補併記下で評価者が B を選びやすい傾向にあったという限定的解釈にとどめ、両方針の優劣を示すものではない。探索的所見として、術後状態に関する記述 13 件における原疾患章と Z90-Z99（第ⅩⅪ章）との符号化方針の相違、および入力語「コロナウイルス感染症」における符号化方針差異の事例（n=1）が観察された。これらは患者調査と標準病名マスターのそれぞれの方針および目的に沿った符号化結果である。両方針はそれぞれの目的に照らせばいずれも妥当であり、方針差異は優劣の問題ではなく、運用目的の違いに由来する構造的差異として位置づけられる。本検証は予備的なものであり、2 候補併記による選択誘導の可能性、観測の独立性、評価者の代表性、評価設計の限界などを踏まえ、今後さらに多数の専門家の協力を得て検証規模および評価設計を拡張する必要がある。

F．健康危険情報

該当なし。

G．研究発表

1．論文発表

該当なし。

2．学会発表

該当なし。

H．知的財産権の出願・登録状況

（予定を含む。）

1．特許取得

該当なし。

2．実用新案登録

該当なし。

3．その他

該当なし。

参考文献

[1] 厚生労働省. 令和 5 年（2023 年）患者調査の概況. 2024.

[2] 世界保健機関（WHO）. International Statistical Classification of Diseases and Related Health Problems, 10th Revision (ICD-10). 厚生労働省 監訳. 疾病、傷害及び死因の統計分類提要 ICD-10（2013 年版）. 2016.

[3] MEDIS-DC. ICD10 対応 標準病名マスター. Ver.5.17.

[4] Cohen J. Educational and Psychological Measurement. 1960; 20(1): 37-46.

[5] Fleiss JL. Psychological Bulletin. 1971; 76(5): 378-382.

[6] Landis JR, Koch GG. Biometrics. 1977; 33(1): 159-174.
[7] Wilson EB. Journal of the American Statistical Association. 1927; 22(158): 209-212.

表 1　事例単位の多数決一致先（対象 124 事例を分母）

多数決一致先	件数	割合（%）	95%CI (%)
標準病名マスターICD	80	64.5	55.8–72.4
患者調査 ICD	34	27.4	20.3–35.9
多数決不成立（3 組とも異なる）	10	8.1	4.4–14.2
いずれとも不一致	0	06	0.0–3.0

割合は対象 124 事例を分母とした。CI は Wilson 法。

表 2　専門家ペアごとの一致率および Cohen の κ 係数

ペア	n	4 桁 一致率	3 桁 一致率	章 一致率	κ（3 桁）	κ（章）
A–B（独立対）	124	53.2	58.1	79.8	0.577	0.778
A–C	122	62.3	66.4	84.4	0.661	0.829
B–C	122	66.4	69.7	85.2	0.694	0.837

数値は一致率(%)。C は C のみ 2 名共同付与。A–B ペア（太字）は独立な評価者対。

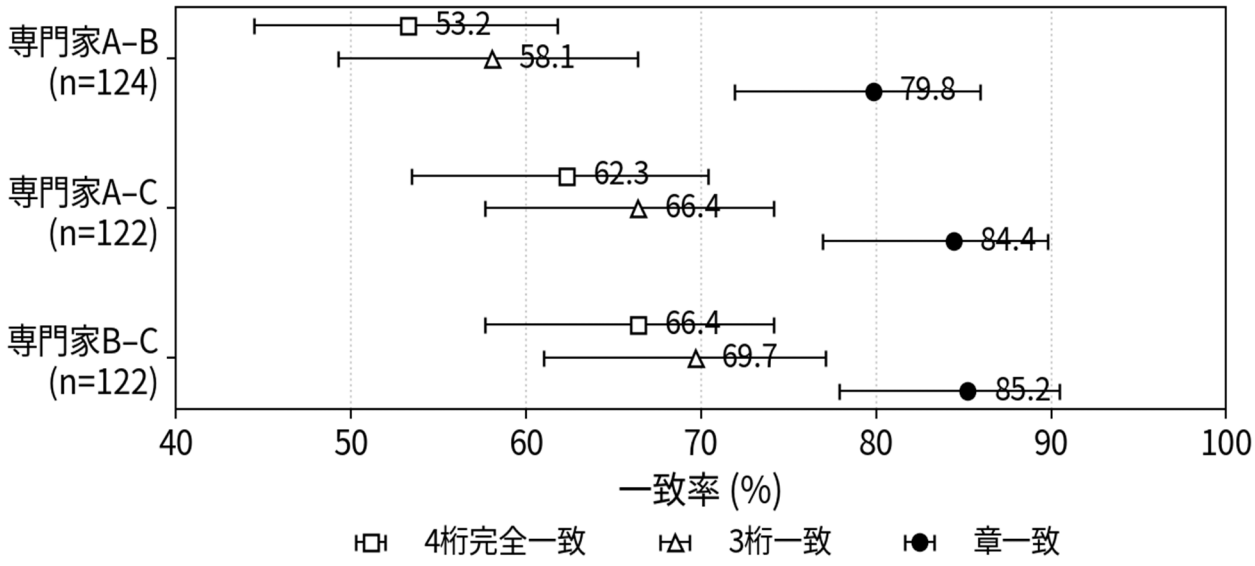


図1　ペアごとの一致率と 95%信頼区間（Wilson 法）