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

アンマッチ病名における手付符号付け作業の効率化検討に関する研究

研究分担者 來島裕太（山口県立総合医療センター 医療企画室 室長補佐）

研究要旨

本研究は、統計法に基づく基幹統計である患者調査において、辞書ファイル適用後も符号付けができなかったアンマッチ病名の傾向と課題を明らかにし、今後の辞書ファイル改善に資することを目的とした。令和 5 年の患者調査では、全体の 14%にあたる約 1 万 6 千語が目視による手付け符号付けを要しており、その要因分析が課題であった。

今回、約 1 万 6 千語について、辞書ファイル適用後でも符号付ができなかったアンマッチ病名 75,218 件を対象に、ICD1 桁および 3 桁別の出現状況を集計・分析し、出現率の高い分類に着目した。また、新型コロナウイルス関連病名について表記揺らぎの検証を行った。

集計した結果、ICD1 桁では悪性新生物（C コード）、外傷（S コード）、筋骨格系疾患（M コード）、循環器系疾患が全体の約 4 割を占めた。ICD3 桁では脳梗塞、肺癌、乳癌などが多く、癌を含む病名の大半は C コードに該当していた。また、左右や両側といった側性表現を含む病名が多く確認された。悪性新生物に関する表記や側性表現の整理・除去など、クリーニング作業を追加実施することにより自動符号付け率の向上が期待できることが示唆された。

A 研究目的

患者調査は、統計法（第 2 条第 4 項）に基づく基幹統計であり、病院及び診療所を利用する患者の傷病状況を明らかにし、医療行政の基礎資料を得るための、日本における唯一の網羅的な疾病統計である。厚生労働省内における患者調査の傷病符号付けプロセスは三段階で構成されている。令和 5 年患者調査では、使用した辞書ファイル等約 11 万 2 千語のうち、厚生労働省の共通サービス及び職員独自作成プログラムによる「辞書ファイルを用いた自動符号付け」により 33%（約 3 万 7 千語）、患者調査オートコーディングシステムによる「形態素解析を活用した手付け符号付け」により 53%（約 6 万語）が符号付けされたが、14%（約 1 万 6 千語）は目視による手付け符号付けが必要であった。

このうち、辞書ファイル適用後に、符号付ができなかった病名（以下：アンマッチ病名）について、どのような傾向や課題があるかを分析し、今後の辞書ファイルのあり方について検討する材料とするため、検証した。

B 研究方法

令和 5 年患者調査で使用した辞書ファイル等のうち、アンマッチ病名 75,218 件を対象として分析を行った。

まず、ICD コードの第 1 桁（アルファベット）および第 3桁分類別に出現状況を集計した。その上で、出現頻度の高い分類に焦点を当て、傾向の分析を実施した。

また、令和5年患者調査の実施時期は、新型コロナウイルス感染症の国内初発例の報告から約 3 年が経過した時点で実施されている。この点を踏まえ、新型コロナウイルス感染症が付与された病名について、アンマッチの要因および表記揺らぎの実態について検証を行った。

（倫理面への配慮）

本研究で使用したデータには個人情報は含まれておらず、倫理的配慮に基づき適切に管理・利用した。

C 研究結果

アンマッチ病名の ICD1 桁目のアルファベット別の出現割合は、Cコード(悪性新生物)が 9,395 件(12.5%)と最も多かった。続いて S コード(損傷 中毒及びその他の外因の影響)のうち外傷に該当するコードが 8,203 件(10.9%)、M コード:筋骨格系及び結合組織の疾患が 7,248 件(9.6%)、循環器系の疾患が 6,867 件(9.1%)であり、これら上位4つは 42.16%を占めた。

アンマッチ病名の ICD3桁目別の出現割合の検証では、I63:脳梗塞が 1,442 件(1.9%)と最も多かった。続いて、C34:肺癌が 1,198 件(1.6%)、C50:乳癌が 865 件(1.1%)L30;湿疹が 760 件(1.0%)、D48:骨腫瘍が 750 件(1.0%)であった。

悪性腫瘍が多く含まれることから、アンマッチ病名に「癌」が含まれるものは、6,339 件であり、そのうち 5,956 件(93.9%)は C コード:悪性新生物であった。C コード(悪性新生物)にならない病名については、上皮内癌・非浸潤癌が 102 件(1.6%)、複数病名による別の疾患へのコーディングが、177 件(2.8%)が多く占めた。この複数病名については、急性肺炎、脳梗塞、狭心症などの疾患に加えて、癌の術後の病名付与していた。残りの 104 件(1.7%)については、「がん検診」など本来は病名ではない用語の付与や、「境界悪性卵巣腫瘍」などの境界病変、さらに「解離性大動脈瘤」や「右眼内レンズ挿入後」等における誤記・表記誤りなど、複数の要因により別のコードへ分類されていた。

アンマッチ病名に含まれる側性表現について、語単位での出現回数を集計した。「右」が 12,555 件(16.7%)、「左」が 12,239 件(16.3%)、「両」が 5,894 件(7.8%)、「右側」が 569 件(0.8%)、「左側」が 662 件(0.9%)、「両側」が 1,925 件(2.6%)、「左右」が 128 件(0.2%)、「ミギ」が 111 件(0.1%)、「ヒダリ」が 102 件(0.1%)であった(いずれもアンマッチ病名全体に対する割合)。なお、これらの語句は同一病名内に複数含まれる場合があるため、重複を許容して集計している。

新型コロナウイルス感染症に関係する病名表記は、916 件であり、そのうち 715 件(78.1%)は、U コードにコーディングされており、それ以外は、感染後の病態が 54 件(7.6%)、関連する多系統炎症性症候群が 2 件(0.3%)であった(表4)。病名表記には多様な揺らぎが認められ、「コヴィッド 19」、「COVID」、「コロナウイルス感染症 2019」、「新型コロナウイルス感染症」など、略語、カタカナ表記、数字の組み合わせによって多様な表現が用いられていた(図2)。

新型コロナウイルス感染症の活動性の状態である U07.1 は、641 件、U07.2 は 18 件であった(表5)。

D 考察

C コード(悪性新生物)、S コード(損傷 中毒及びその他の外因の影響)のうち外傷に該当するコード、Mコード:筋骨格系及び結合組織の疾患、循環器系の疾患が大半に占めることから、この4つの分類について改善されることが有効と考えられた。

Cコード(悪性新生物)については、病名に「癌」と記載されている場合、その多くが悪性新生物として分類され、上皮内新生物を除くカテゴリーに該当する。このことから、マッチング精度向上の観点では、C コード(悪性新生物)に分類されない病名表記のクリーニングが重要であると考えられる。

実際に、「癌」と診断された症例のうち、非浸潤性あるいは上皮内にとどまる病変(例:非浸潤性膀胱癌や結腸上皮内癌など)を除外することで、悪性新生物以外への誤分類は概ね整理可能であることが示唆された。さらに、それ以外のアンマッチ病名としては、境界悪性を示す病名や、「癌検診」など本来は病名ではない用語が付与されている事例も確認された。これらについても、一定のルールに基づくデータクリーニングを実施することで、対応可能であると考えられる。

I コードおよび S コードに関しては、左右情報の取扱いがマッチング結果に影響している可能性が示唆された。左右を含む病名について分析したところ、アンマッチ病名において「右」「左」「両」の順に語句を含む割合が高い傾向が認められた。この結果から、ICD コードに直接影響を及ぼさない左右情報については、前処理として除去または正規化を行うことで、マッチング件数の向上が期待される。

一方、標準病名マスター(ver.516)における左右表

記に起因するコード変更事例は24件にとどまっていた（表 3）。このことは、左右情報がコーディング全体に及ぼす影響は限定的である可能性を示している。

ただし例外として、変形性膝関節症や鼠径ヘルニアのように、左右情報ではなく、片側性あるいは両側性の違いが分類に影響する疾患も存在するため、左右情報の一律な除去には慎重な対応が求められる。

新型コロナウイルス感染症については、その多くがUコードに分類されることを踏まえると、全アンマッチ病名に占める割合は約 1%と低値であり、全体への影響は限定的であると考えられる。一方で、当該疾患に関する表記の揺らぎは相対的に大きいことが明らかとなった。英語表記とカタカナ表記やその混在が存在する多様な表記は、アンマッチの一因となっている可能性が高い。

また、新型コロナウイルス感染症は、臨床的・疫学的にはインフルエンザなど既存の呼吸器感染症に近似しつつあるものの、入院患者において主病名または併存症として一定数存在し、データの解析精度に影響を及ぼす可能性がある。

以上より、新型コロナウイルス感染症に関しては、影響の大きさ自体は限定的であるものの、マッチング精度向上の観点から、表記の標準化や同義語の統合といったデータクリーニングの必要性が示唆された。

E 結論

アンマッチ病名の傾向を把握することで、手付付与の ICD 作業を改善することに期待できる。さらに ICD 分類の性質や影響度合いから、クリーニング作業やマッチング率向上を優先的にすることが必要である。

F 健康危険情報

なし

G 研究発表

1.論文発表

なし

2.学会発表

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

なし

2.実用新案登録

なし

3.その他

なし

表 1　アンマッチ 75,218 件に対する ICD1 桁目（アルファベット）分類による集計結果

A	感染症及び寄生虫症	711	0.9%
B	感染症及び寄生虫症	1,478	2.0%
C	悪性新生物	9,395	12.5%
D	上皮内・良性新生物／血液及び造血器の疾患並びに免疫機構の障害	4,188	5.6%
E	内分泌，栄養及び代謝疾患	2,271	3.0%
F	精神及び行動の障害	1,798	2.4%
G	神経系の疾患	2,460	3.3%
H	眼及び付属器／耳及び乳様突起の疾患	3,763	5.0%
I	循環器系の疾患	6,867	9.1%
J	呼吸器系の疾患	3,132	4.2%
K	消化器系の疾患	5,580	7.4%
L	皮層及び皮下組織の疾患	5,391	7.2%
M	筋骨格系及び結合組織の疾患	7,248	9.6%
N	腎尿路生殖器系の疾患	2,285	3.0%
O	妊娠，分娩及び産褥	2,224	3.0%
P	周産期に発生した病態	287	0.4%
Q	先天奇形，変形及び染色体異常	1,305	1.7%
R	症状，徴候及び異常臨床所見・異常検査所見で他に分類されないもの	1,852	2.5%
S	損傷 中毒及びその他の外因の影響	8,203	10.9%
T	損傷 中毒及びその他の外因の影響	3,829	5.1%
U	特殊目的用コード	715	1.0%
Z	健康状態に影響を及ぼす要因及び保健サービスの利用	236	0.3%
総計		75,218	100.0%

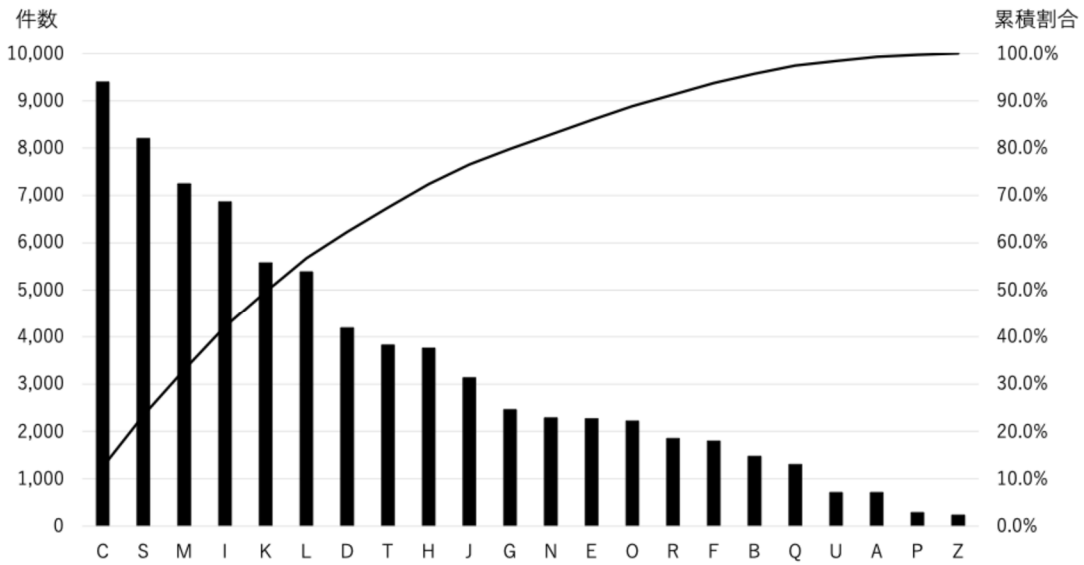


図 1：ノンマッチ75,218件に対するICD1桁目（アルファベット）分類による分布

表 2　アンマッチ 75,218 件に対する ICD3桁分類による集計結果（上位 25 位）

順位	ICD3 桁	病名	件数	割合	累積割合
1	I63	脳梗塞	1,442	1.9%	1.9%
2	C34	肺癌	1,198	1.6%	3.5%
3	C50	乳癌	865	1.1%	4.7%
4	L30	湿疹	760	1.0%	5.7%
5	D48	骨腫瘍	750	1.0%	6.7%
6	C16	胃癌	697	0.9%	7.6%
7	U07	新型コロナウイルス	659	0.9%	8.5%
8	S06	頭部外傷	649	0.9%	9.3%
9	C18	結腸癌	635	0.8%	10.2%
10	S32	腰椎損傷	607	0.8%	11.0%
11	S82	膝損傷	569	0.8%	11.7%
12	I61	脳出血	542	0.7%	12.5%
13	T02	多発骨折	525	0.7%	13.2%
14	S72	大腿骨折	517	0.7%	13.8%
15	L89	褥瘡	507	0.7%	14.5%
16	S62	指骨折	505	0.7%	15.2%
17	L03	蜂窩織炎	487	0.6%	15.8%
18	T81	術後合併症	481	0.6%	16.5%
19	I10	高血圧	457	0.6%	17.1%
20	I50	心不全	446	0.6%	17.7%
21	E11	２型糖尿病	442	0.6%	18.3%
22	M17	変形性膝関節症	431	0.6%	18.8%
23	H35	網膜症	428	0.6%	19.4%
24	E14	糖尿病	420	0.6%	20.0%
25	S92	足指骨折	410	0.5%	20.5%

表３：標準病名マスター(ver.516)における左右の記載がある病名一覧

病名表記（右）	ＩＣＤ（右）	病名表記（左）	ＩＣＤ（左）	比較
右心室粘液腫	D151	左心室粘液腫	D151	同一
右心房粘液腫	D151	左心房粘液腫	D151	同一
肝右葉腫瘍	D376	肝左葉腫瘍	D376	同一
右半球器質性感情障害	F078			片側のみ
右眼失明	H544	左眼失明	H544	同一
急性右室梗塞	I212			片側のみ
右室自由壁破裂	I219	左室自由壁破裂	I219	同一
不整脈原性右室心筋症	I428			片側のみ
		左脚前枝ブロック	I444	片側のみ
		左脚後枝ブロック	I445	片側のみ
		左脚ヘミブロック	I446	片側のみ
完全右脚ブロック	I451	完全左脚ブロック	I447	異なる
右脚ブロック	I451	左脚ブロック	I447	異なる
不完全右脚ブロック	I451	不完全左脚ブロック	I447	異なる
		左室駆出率が軽度低下した心不全	I501	片側のみ
		左室駆出率の保たれた心不全	I501	片側のみ
		左室駆出率の低下した心不全	I501	片側のみ
右室不全	I500	左室不全	I501	異なる
右心不全	I500	左心不全	I501	異なる
後天性左室右房交通症	I510	後天性左室右房交通症	I510	同一
		巨大左心房	I517	片側のみ
右室肥大	I517	左室肥大	I517	同一
右胃動脈瘤	I728	左胃動脈瘤	I728	同一
下顎右側偏位	K071	下顎左側偏位	K071	同一
上顎右側偏位	K071	上顎左側偏位	K071	同一
		潰瘍性大腸炎・左側大腸炎型	K515	片側のみ
両大血管右室起始症	Q201	両大血管左室起始症	Q202	異なる
右室型単心室症	Q204	左室型単心室症	Q204	同一
		左旋転位	Q205	片側のみ
右室流出路狭窄	Q208	左室流出路狭窄	Q208	同一
		大動脈左室トンネル	Q208	片側のみ
左室右房交通症	Q210	左室右房交通症	Q210	同一
右室低形成	Q226			片側のみ
右室低形成症候群	Q226			片側のみ
右心低形成症候群	Q226	左心低形成症候群	Q234	異なる
右胸心	Q240	左胸心	Q241	異なる
右心症	Q240	左心症	Q241	異なる
右室漏斗部狭窄	Q243			片側のみ
右冠動脈肺動脈起始症	Q245	左冠動脈肺動脈起始症	Q245	同一
右室二腔症	Q248			片側のみ
右側大動脈弓	Q254	先天性左室憩室	Q248	異なる
左肺動脈右肺動脈起始症	Q257			片側のみ
右肺動脈形成不全症	Q257	左肺動脈形成不全症	Q257	同一
右肺動脈欠損	Q257	左肺動脈欠損	Q257	同一
右肺動脈上行大動脈起始症	Q257	左肺動脈上行大動脈起始症	Q257	同一
		左上大静脈遺残	Q261	片側のみ
		左側胆のう	Q441	片側のみ
心電図右軸偏位	R943	心電図左軸偏位	R943	同一

表 4　アンマッチのうち、新型コロナウィルス感染症に関係する病名表記におけるコーディング状況

ｱﾙﾌｧﾍﾞｯﾄ	章	件数	割合
A	感染症及び寄生虫症	2	0.2%
B	感染症及び寄生虫症	0	0.0%
C	悪性新生物	4	0.4%
D	上皮内・良性新生物／血液及び造血器の疾患並びに免疫機構の障害	2	0.2%
E	内分泌，栄養及び代謝疾患	19	2.1%
F	精神及び行動の障害	4	0.4%
G	神経系の疾患	10	1.1%
H	眼及び付属器／耳及び乳様突起の疾患	0	0.0%
I	循環器系の疾患	12	1.3%
J	呼吸器系の疾患	53	5.8%
K	消化器系の疾患	10	1.1%
L	皮層及び皮下組織の疾患	1	0.1%
M	筋骨格系及び結合組織の疾患	4	0.4%
N	腎尿路生殖器系の疾患	10	1.1%
O	妊娠，分娩及び産褥	20	2.2%
P	周産期に発生した病態	12	1.3%
Q	先天奇形，変形及び染色体異常	0	0.0%
R	症状，徴候及び異常臨床所見・異常検査所見で他に分類されないもの	5	0.5%
S	損傷 中毒及びその他の外因の影響	11	1.2%
T	損傷 中毒及びその他の外因の影響	14	1.5%
U	特殊目的用コード	715	78.1%
Z	健康状態に影響を及ぼす要因及び保健サービスの利用	8	0.9%
総計		916	100.0%

- コヴイツト－１９
- コヴイツドナインテイーン
- コビツト
- コビツド１９
- ＣＯＶ－１９
- ＣＯＶ－１９感染症
- ＣＯＶＩＤ
- ＣＯＶＩＤ－１９
- コロナ感染
- コロナ感染者
- コロナ感染症
- コロナ陽性
- コロナウイルス
- コロナウイルス・２０１９
- コロナウイルス感染症
- コロナウイルス感染症２０１９
- 新型コロナウイルス
- 新型コロナウイルス感染症

など

図２：新型コロナウイルス感染症の表記例

表 5 アンマッチのうち、新型コロナウィルス感染症のコーディング状況

ICD	名　　称	件　数
U07.1	コロナウイルス感染症 2019，ウイルスが同定されたもの	641
U07.2	コロナウイルス感染症 2019，ウイルスが同定されていないもの	18
U08.9	コロナウイルス感染症 2019 の既往歴，詳細不明	0
U09.9	コロナウイルス感染症 2019 後の病態，詳細不明	54
U10.9	コロナウイルス感染症 2019 に関連する多系統炎症性症候群，詳細不明	2
U 11.9	コロナウイルス感染症 2019 に対する予防接種の必要性，詳細不明	0
U12.9	治療上の使用により有害作用を引き起こしたコロナウイルス感染症	0
	合計	715