

令和4年度厚生労働行政推進調査事業補助金
政策科学総合研究事業(政策科学推進事業)

「DPC制度の適切な運用及びDPCデータの活用に資する研究」

分担研究報告書

「DPC データ、様式 1 における欠損および不明登録等の実態と要因の検証」

研究分担者 藤森研司(東北大学 医学系研究科公共健康医学講座 医療管理学分野 教授)

研究協力者 桜澤邦男(同分野 講師)

研究協力者 宮崎大輔(同分野 修士課程大学院生)

研究要旨

DPC データ、様式 1 における欠損および不明登録等の実態を明らかにし、欠損または不明登録等につながる要因を症例単位および施設単位ごとに検証した。

様式 1 に含まれるデータについて、欠損となる割合は大半が 1%未満であり、不明登録等の割合は欠損に比べ高い出現頻度であった。欠損または不明登録等の割合は、予定外・緊急入院症例および相対的に症例数が少ない施設において有意に高かった。

全体として適切にデータ登録がなされている実態が明らかとなったが、欠損または不明登録等につながる要因に従い、データ登録に係る一層の体制強化が期待される。

A 研究目的

DPC データは日本全国の急性期病院等より収集され、薬剤や手術等の医療費請求に係る情報(E ファイル、F ファイル、D ファイル等)に加え、患者属性や病態等の詳細な情報を含む様式 1 が含まれている。DPC データではこれらの情報を組み合わせて分析することが可能であり、医療サービスへのアクセスや医療の質に関わる研究のみならず多くの臨床疫学研究に用いられている。

臨床疫学研究を行う際の課題の一つとして、様式 1 におけるデータの欠損および不明登録等の存在があげられる。多変量解析に組み入れるデータに欠損や不明登録等がみられる場合、当該症例を除外して分析する必要があるため、分析対象となる症例選択にバイアスが生じる懸念がある。欠損値を補完する分析手法として、多重代入法などの統計手法を用いる方法もあるが、推定値ではなく正しく登録された情報を基に分析することが極めて肝要で

ある。

東北大学大学院医学系研究科医療管理学分野では、入院患者の状態評価に有用な SOFA スコアの欠損および不明登録に係る分析を継続的に行ってきた¹⁾³⁾。しかしながら様式 1 における必須登録項目に対して、網羅的にデータの欠損や不明登録等の実態を明らかにしたものはない。加えて、データの欠損や不明登録等につながる患者や施設に起因する問題を明らかにし、欠損や不明登録等の発生につながる要因について検証した報告はみられない。そこで本報告書は、様式 1 におけるデータの欠損および不明登録等の実態を把握し、データの欠損および不明登録等につながる要因について論じる。

B 研究方法

1) 分析対象

本研究は、一般社団法人診断群分類研究支援機構を通して収集された DPC 調査データ

(各医療施設が厚生労働省に提出する様式 1) を分析対象とした。対象期間は、2020 年度および 2021 年度とした。様式 1 へのデータ登録の方法に関しては「2020 年度 DPC 導入の影響評価に係る調査 実施説明資料(2020 年 2 月 28 日時点)」を参照した。

2) 一次集計

分析対象に含まれる各種データについて、欠損および不明登録等に係る件数および割合を集計した。不明登録等は、不明または測定不能などの登録定義が明確であるものを対象とした。

集計対象とする項目は、患者属性(生年月日、性別、患者住所地域の郵便番号)、入院情報(入院年月日、入院経路、他院よりの紹介の有無、自院外来からの入院、予定・緊急医療入院、救急車による搬送の有無、入院前の在宅医療の有無、自傷行為・自殺企図の有無)、退院情報(退院年月日、退院先、退院時転帰、24 時間以内の死亡の有無、退院後の在宅医療の有無)、様式 1 対象期間(様式 1 開始日、様式 1 終了日)、診療科(診療科コード、転科の有無)、診療目的・経過(診療目的、治療実施の有無)、患者プロフィール(身長、体重、喫煙指数)、要介護度、診断情報/主傷病(ICD10 コード、主傷病名)、診断情報/入院契機(ICD10 コード、入院の契機となった傷病名)、診断情報/医療資源(ICD10 コード、医療資源を最も投入した傷病名)、ADL スコア(入院時の ADL スコア、退院時の ADL スコア)、JCS(入院時意識障害がある場合の JCS、退院時意識障害がある場合の JCS)、心疾患患者/NYHA、狭心症 CCS、急性心筋梗塞 Killip、急性膵炎の重症度分類、入院周辺の分娩の有無、分娩時出血量、SOFA スコア/特定集中

治療室(入室当日、退室日)、SOFA スコア/敗血症(治療開始日)をとりあげた。

3) 二次集計(クロス集計)

欠損または不明登録等の出現割合に与える影響を症例単位および施設単位ごとに確認するため、データの二次集計(クロス集計)を行った。症例単位として、予定・救急医療入院の影響を、施設単位として症例数の大小および大学病院の該当有無を、それぞれ取り上げた。予定・救急医療入院については、予定外・緊急入院および予定入院に群分けした。症例数の大小については、医療施設別に診療月 1 か月あたりのデータ提出件数を求め、中央値折半法により症例数が中央値以上および症例数が中央値未満に群分けした。大学病院の該当有無については、医療施設コードをもとに大学病院および大学病院以外に群分けした。

データの欠損または不明登録等の出現割合を評価する項目として、入院時の患者状態評価に用いる身長、体重、入院時の ADL スコアそして入院時意識障害がある場合の JCS をとりあげた。

4) 分析手法

一次集計は、データの欠損および不明登録等に係る件数および割合をもとめ記述統計学的に整理した。

二次集計(クロス集計)は、予定・救急医療入院、症例数の大小および大学病院の該当有無の各群別に、身長、体重、入院時の ADL スコア、入院時意識障害がある場合の JCS のデータ欠損および不明登録等の件数と割合を整理し比較した。比較に際し、割合の違いを明らかにするための統計学的手法として独立性の検定(カイ二乗検定)を用いた。本検定は

両側検定により実施し、統計学的有意水準を5%未満とした。

C 研究結果

表1は、様式1におけるデータ欠損および不明登録等の集計結果を示したものである。

欠損の割合は0.000%～11.967%の範囲で出現していたが、大半の項目は1%以内におさまっていた。不明登録等の割合は0.000%～21.972%の範囲で出現しており、欠損に比べて高い頻度であった。

表1 様式1、データ欠損および不明登録等の集計

種別	項目名	データ登録 対象件数	欠損 の件数	欠損 の割合	不明登録等 の件数	不明登録等 の割合
患者属性	生年月日	14,558,456	0	0.000%	0	0.000%
	性別	14,558,456	0	0.000%		
	患者住所地域の郵便番号	14,558,456	5	0.000%	3,665	0.025%
入院情報	入院年月日	14,558,456	0	0.000%		
	入院経路	14,558,456	12	0.000%		
	他院よりの紹介の有無	14,358,511	9	0.000%		
	自院外来からの入院	14,358,511	9	0.000%		
	予定・救急医療入院	14,358,511	2	0.000%		
	救急車による搬送の有無	14,358,511	0	0.000%		
	入院前の在宅医療の有無	14,358,511	35	0.000%	124,987	0.870%
	自傷行為・自殺企図の有無	76,946	163	0.212%		
退院情報	退院年月日	14,558,456	0	0.000%		
	退院先	14,558,456	79	0.001%		
	退院時転帰	14,558,456	31	0.000%		
	24時間以内の死亡の有無	14,558,456	12	0.000%		
	退院後の在宅医療の有無	14,558,356	96	0.001%	127,345	0.875%
様式1対象期間	様式1開始日	14,558,456	0	0.000%		
	様式1終了日	14,558,456	0	0.000%		
診療科	診療科コード	14,558,456	49	0.000%		
	転科の有無	14,558,456	4	0.000%		
診療目的・経過	診療目的	14,558,456	2	0.000%		
	治療実施の有無	14,558,456	1	0.000%		
患者プロフィール	身長	14,558,456	11	0.000%	704,046	4.836%
	体重	14,558,456	9	0.000%	533,662	3.666%
	喫煙指数	14,558,456	13	0.000%	1,598,129	10.977%
要介護度	要介護度	14,558,456	402	0.003%	454,101	3.119%
診断情報/主傷病	ICD10コード	14,558,456	0	0.000%		
	主傷病名	14,558,456	0	0.000%		
診断情報/入院契機	ICD10コード	14,558,456	3	0.000%		
	入院の契機となった傷病名	14,558,456	3	0.000%		
診断情報/医療資源	ICD10コード	14,558,456	0	0.000%		
	医療資源を最も投入した傷病名	14,558,456	0	0.000%		
ADLスコア	入院時のADLスコア	13,494,542	99,082	0.734%	876,002	6.492%
	退院時のADLスコア	12,853,338	139,629	1.086%	551,611	4.292%
JCS	入院時意識障害がある場合のJCS	14,558,456	599,960	4.121%		
	退院時意識障害がある場合のJCS	13,914,106	49	0.000%		
心疾患患者/NYHA	心疾患患者/NYHA	23,073	17	0.074%		
狭心症 CCS	狭心症 CCS	432,224	96	0.022%		
急性心筋梗塞 Killip	急性心筋梗塞 Killip	100,459	21	0.021%		
急性膵炎	急性膵炎の重症度分類	46,754	10	0.021%	6,268	13.406%
産科患者情報	入院周辺の分娩の有無	235,873	3	0.001%		
産科患者情報	分娩時出血量	147,566	0	0.000%	2,351	1.593%
SOFAスコア/特定集中治療室	入室当日	444,890	53,240	11.967%	80,000	17.982%
	退室日	444,890	53,248	11.969%	83,321	18.728%
SOFAスコア/敗血症	治療開始日	218,844	4,683	2.140%	48,084	21.972%

表 2 は、予定・救急医療入院と入院時の患者状態評価のデータ欠損または不明登録等の関連を示したものである。

身長に関して欠損または不明登録等がある割合は、予定外・緊急入院で 9.111%、予定入院で 1.405%であった。体重に関して欠損または不明登録等がある割合は、予定外・緊急入院で 6.882%、予定入院で 1.115%であった。入院時の ADL スコアに関して欠損または不明登録等がある割合は、予定外・緊急入院で 12.000%、予定入院で 3.441%であった。入院時意識障害がある場合の JCS に関して欠損がある割合は、予定外・緊急入院で 8.041%、予定入院で 1.032%であった。

いずれの項目も統計学的有意差がみられ、予定外・緊急入院で欠損または不明登録等がある割合が高かった。

表2 予定・救急医療入院と入院時の患者状態評価のデータ欠損および不明登録等の関連
身長

	欠損または不明登録等なし		欠損または不明登録等あり	
予定外・緊急入院	5,822,136	90.889%	583,629	9.111%
予定入院	7,841,047	98.595%	111,697	1.405%

$\chi^2=457267.1, df=1, p<0.001$

体重

	欠損または不明登録等なし		欠損または不明登録等あり	
予定外・緊急入院	5,964,908	93.118%	440,857	6.882%
予定入院	7,864,087	98.885%	88,657	1.115%

$\chi^2=332268.0, df=1, p<0.001$

入院時のADLスコア

	欠損または不明登録等なし		欠損または不明登録等あり	
予定外・緊急入院	5,228,534	88.000%	712,973	12.000%
予定入院	7,277,857	96.559%	259,381	3.441%

$\chi^2=363581.2, df=1, p<0.001$

入院時意識障害がある場合のJCS

	欠損なし		欠損あり	
予定外・緊急入院	5,890,698	91.959%	515,067	8.041%
予定入院	7,870,674	98.968%	82,070	1.032%

$\chi^2=437257.8, df=1, p<0.001$

表 3 は、症例数の大小と入院時の患者状態評価のデータ欠損または不明登録等の関連を示したものである。

身長に関して欠損または不明登録等がある割合は、症例数が中央値以上で 4.684%、中央値未満で 5.575%であった。体重に関して欠損または不明登録等がある割合は、症例数が中央値以上で 3.527%、中央値未満で 4.342%であった。入院時の ADL スコアに関して欠損または不明登録等がある割合は、症例数が中央値以上で 7.109%、中央値未満で 7.766%であった。入院時意識障害がある場合の JCS に関して欠損がある割合は、症例数が中央値以上で 3.751%、中央値未満で 5.921%であった。

いずれの項目も統計学的有意差がみられ、症例数が中央値未満で欠損または不明登録等がある割合が高かった。

表3 症例数の大小と入院時の患者状態評価のデータ欠損および不明登録等の関連
身長

	欠損または不明登録等なし		欠損または不明登録等あり	
症例数が中央値以上	11,510,506	95.316%	565,662	4.684%
症例数が中央値未満	2,343,893	94.425%	138,395	5.575%

$\chi^2=3553.3, df=1, p<0.001$

体重

	欠損または不明登録等なし		欠損または不明登録等あり	
症例数が中央値以上	11,650,290	96.473%	425,878	3.527%
症例数が中央値未満	2,374,495	95.658%	107,793	4.342%

$\chi^2=3881.4, df=1, p<0.001$

入院時のADLスコア

	欠損または不明登録等なし		欠損または不明登録等あり	
症例数が中央値以上	10,308,282	92.891%	788,913	7.109%
症例数が中央値未満	2,211,176	92.234%	186,171	7.766%

$\chi^2=1267.8, df=1, p<0.001$

入院時意識障害がある場合のJCS

	欠損なし		欠損あり	
症例数が中央値以上	11,623,186	96.249%	452,982	3.751%
症例数が中央値未満	2,335,310	94.079%	146,978	5.921%

$\chi^2=24539.6, df=1, p<0.001$

表4は、大学病院の該当有無と入院時の患者状態評価のデータ欠損または不明登録等の関連を示したものである。

身長に関して欠損または不明登録等がある割合は、大学病院で 2.135%、大学病院以外で 5.427%であった。体重に関して欠損または不明登録等がある割合は、大学病院で 1.673%、大学病院以外で 4.102%であった。入院時の ADL スコアに関して欠損または不明登録等がある割合は、大学病院で 5.504%、大学病院以外で 7.592%であった。入院時意識障害がある場合の JCS に関して欠損がある割合は、大学病院で 2.065%、大学病院以外で 4.571%であった。

いずれの項目も統計学的有意差がみられ、大学病院以外で欠損または不明登録等がある割合が高かった。

表4 大学病院の該当有無と入院時の患者状態評価のデータ欠損および不明登録等の関連
身長

	欠損または不明登録等なし		欠損または不明登録等あり	
大学病院	2,557,709	97.865%	55,792	2.135%
大学病院以外	11,296,690	94.573%	648,265	5.427%

$\chi^2=50505.3, df=1, p<0.001$

体重

	欠損または不明登録等なし		欠損または不明登録等あり	
大学病院	2,569,778	98.327%	43,723	1.673%
大学病院以外	11,455,007	95.898%	489,948	4.102%

$\chi^2=35819.4, df=1, p<0.001$

入院時のADLスコア

	欠損または不明登録等なし		欠損または不明登録等あり	
大学病院	2,236,679	94.496%	130,274	5.504%
大学病院以外	10,282,779	92.408%	844,810	7.592%

$\chi^2=12695.5, df=1, p<0.001$

入院時意識障害がある場合のJCS

	欠損なし		欠損あり	
大学病院	2,559,533	97.935%	53,968	2.065%
大学病院以外	11,398,963	95.429%	545,992	4.571%

$\chi^2=34080.0, df=1, p<0.001$

D 考察

本報告書では、様式 1 におけるデータの欠損および不明登録等の実態を把握し、データの欠損または不明登録等につながる要因を患者および施設の両面から検証した。

データの欠損に関して、欠損割合は 0.000%～11.967%の範囲で出現していたが、大半の項目は 1%以内におさまっていた。これにより、全体としては漏れなく登録されているケースが多数を占めている実態が明らかとなった。2020 年度より新たに必須登録項目として加わった「要介護度」についても、欠損となる割合は 0.003%であるため、新規導入となった項目にも適切な対応がなされていると考えられた。入院時意識障害がある場合の JCS は、欠損となる割合が 4.121%であった。当該項目は不明登録とする際の定義が存在しない項目であるため、入院時になんらかの原因で測定できなかった場合は欠損とせざるを得ないケースが含まれるものと考えられた。また、SOFA スコア/特定集中治療室は、欠損となる割合が入室当日で 11.967%、退室日で 11.969%であった。特定集中治療室管理料算定など、定められた診療報酬を算定する患者への登録に関して、登録を促すアナウンス等の改善活動を一層強化するなどの対策を要すると考えられた。なお DPC 調査データに関して、各医療機関は厚生労働省へ提出する際、データ欠損等の確認作業を行う必要がある。本報告書の分析に用いた DPC 調査データについては、この確認作業を完了しているかどうかを判別できない。一方、データの不明登録等については、提出時の確認作業に含まれていないため、本報告書においてその実態が示されているものと考えられる。

データの不明登録等について、不明登録等の割合は 0.000%～21.972%の範囲で出現しており、全体的に欠損に比べて高い頻度であった。本調査において、ADL や SOFA の各スコアならびに急性膵炎の重症度分類等については、登録すべき複数の項目のうち1つでも不明登録が含まれている場合に「不明登録等」と定義し抽出・集計を行った。そのため不明登録等として扱った症例についても、部分的には正確に登録されているデータを含む可能性がある。正確に登録されているデータについては分析への使用が可能となることから、本調査における集計方法では不明登録等の割合が高めに計上されている。今後は、データの全項目が不明なのか一部の項目が不明なのかを識別できる集計方法の採用を採り入れ、データの不明登録等についてさらなる実態把握を行う必要がある。

データの欠損または不明登録等につながる要因に関して、患者要因では予定入院に比べ、予定外・緊急入院で欠損または不明登録等となる割合が有意に高かった。これは、時間外、休日そして深夜など病院スタッフが少ない時間帯を含む予定外・緊急的な入院が相対的に多く含まれるため、入院時の手続きとして必要なデータ登録作業を行うことが困難であったことなどが要因として考えられた。

施設要因では、症例数が中央値未満および大学病院以外の施設で欠損または不明登録等がある割合が有意に高かった。これは、病床数の大小に起因する病院施設の規模感に応じて配置されるデータ登録担当者等の人数の相対的な違い(一般論として、大規模施設で多くの担当者が配置される)などが関連していると考えられた。

本調査で示した欠損または不明登録等に

つながる要因に従い、データ登録に係る一層の体制強化が期待される。

本調査は、年間あたり1,100を超える施設ならびに700万件を超える症例数からなるDPCデータを対象として実施した。本データは我が国の急性期入院医療に関して代表性を有する大規模医療ビッグデータであるが、悉皆性を有するものではない。そのため、本調査における結果の解釈には留意する必要がある。

E 結論

DPC データ、様式1における欠損および不明登録等の実態を明らかにし、欠損および不明登録等につながる要因について検証した。その結果、欠損は大半の項目で1%未満の出現割合であり、欠損または不明登録等となる要因として、予定外・緊急入院であること、相対的に症例数が少ない病院であることが挙げられた。

全体として適切にデータ登録がなされている実態が明らかとなったが、欠損または不明登録等につながる要因に従い、データ登録に係る一層の体制強化が期待される。

F 健康危険情報

なし

G 知的財産権の出願

特許取得

なし

実用新案登録

なし

その他

なし

H 利益相反

なし

I 研究発表

なし

引用文献

1) 桜澤邦男, 藤森研司: SOFA スコアの不明登録および活用に関する評価 大規模DPCデータを用いた ICU への緊急入院患者の分析, 診療情報管理, 32 巻 4 号, p22-30, 2021.

2) 桜澤邦男, 藤森研司: 特定集中治療室管理料算定病床における SOFA スコア登録漏れの傾向に関する研究～DPC 全国データを用いた時系列解析～, 日本診療情報管理学会学術大会抄録集, 2022.

3) 桜澤邦男, 藤森研司: DPC データにおける SOFA スコアの登録と臨床疫学的活用 不明登録の状況と DIC 治療薬の有効性, 日本診療情報管理学会学術大会抄録集, 2021.

