

平成 30 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)

「診断群分類を用いた急性期等の入院医療の評価とデータベース利活用に関する研究」
分担研究報告書

DPC 分析用データセットの作成・開発について

堀口 裕正 国立病院機構本部総合研究センター 診療情報分析部 副部長

研究要旨

本研究班において、収集した DPC データは、データセットの量が大きく、一般的な研究者が保有する分析環境（コンピュータの能力やデータを保管するストレージの量等）では処理が行えない状況となっている。また、その膨大なデータのうち、矛盾するレコードや、研究で使用するには留意が必要なデータも混じっている。

そこで、いくつかのデータ処理を行うことによって、データを分析可能なものに絞り込み、さらに分析に必要な様々な処理を加えてデータセットを作成し、さまざまな研究が実施しやすい環境を構築することを行った。

本年度、平成 22 年～29 年のデータについては新たに利用承諾がとれた医療機関のデータを加えて、分析用データセットの再作成を行うとともに、平成 30 年のデータについて分析用のデータセットの作成を行い、分析に供することができた。その際、キー情報の重複や必要なデータの欠損のある症例などは、分析に影響度が大きいので、それらの症例情報については確実に除去を行った。

このデータセットの完成で、DPC データの精度を向上させ、より高度な分析を実施することが可能となると考えられる。

A. 目的

本研究班において、収集した DPC データは、まず基本調査として入院情報の D/E/F/H ファイル・様式 1・3・4 ファイル（以下それぞれ FF1/FF3/FF4 ファイルと表記）と多岐にわたっている。また各病院のデータ収集が通年化され収集時期の範囲も広く、多数の医療機関が参加している。さらに、付随する調査として ICU/ER・リ

ハビリ・CP・外来機能といった各種調査が付随し、それぞれが、基本調査の集計データを分析にあたって必要としている。

平成 29 年 3 月現在、本研究班では平成 15 年度調査から平成 29 年度調査までの 14 年度の調査データを医療機関との契約の元で保有し、分析に活用しているが、平成 30 年調査データの分量はすべてテキストのデータで、1TB を超えた量となっており、一

一般的な研究者が保有する分析環境（コンピュータの能力やデータを保管するストレージの量等）では処理が行えない状況となっている。

そこで、本研究では、平成 21 年度より本研究班の保有する DPC 調査データから、

- 1、 利用できるデータを選択し、
- 2、 その 1 つ 1 つのデータが研究に利用する際に留意すべきポイントがあるかをチェックして情報を付加し
- 3、 分析に必要な情報を容易にとりだせるよう用途別に分割し、

さまざまな統計処理を行うための数値を計算しやすいよう整理して共通分析用データセットとして作成し、本研究班においてより精度の高い、高度な分析を行うための環境を整備することを試みている。本年度は平成 30 年度(12 ヶ月)分のデータを通年で 1 つのデータセットとして取り扱い、各種の分析処理を行うことが出来る基盤作成を目的とした。

B.方法

以下の処理を、研究班保有の DPC データに対して行い、分析用データセットを作成することとした。

1、データの取り込み

参加医療機関から提供された DPC データ (FF1/3/4/D/E/F/H/外来ファイル) を DB に取り込む。その際、平成 26 年度データからの様式 1 ファイル縦持ち化に対応した処理を追加した。また、平成 28 年度フォーマット改正および H ファイルの追加に伴う処理を追加した。

2、エラーチェック

提出されたデータのエラーチェックを行

い、エラーデータをデータセット内から除外する。また研究に使用する際留意する項目に対してチェックシフラグを付与する。

どのような条件についてエラー・留意とするかについては本年の研究として検討を行う。

3、DPC コード情報の一体的保有

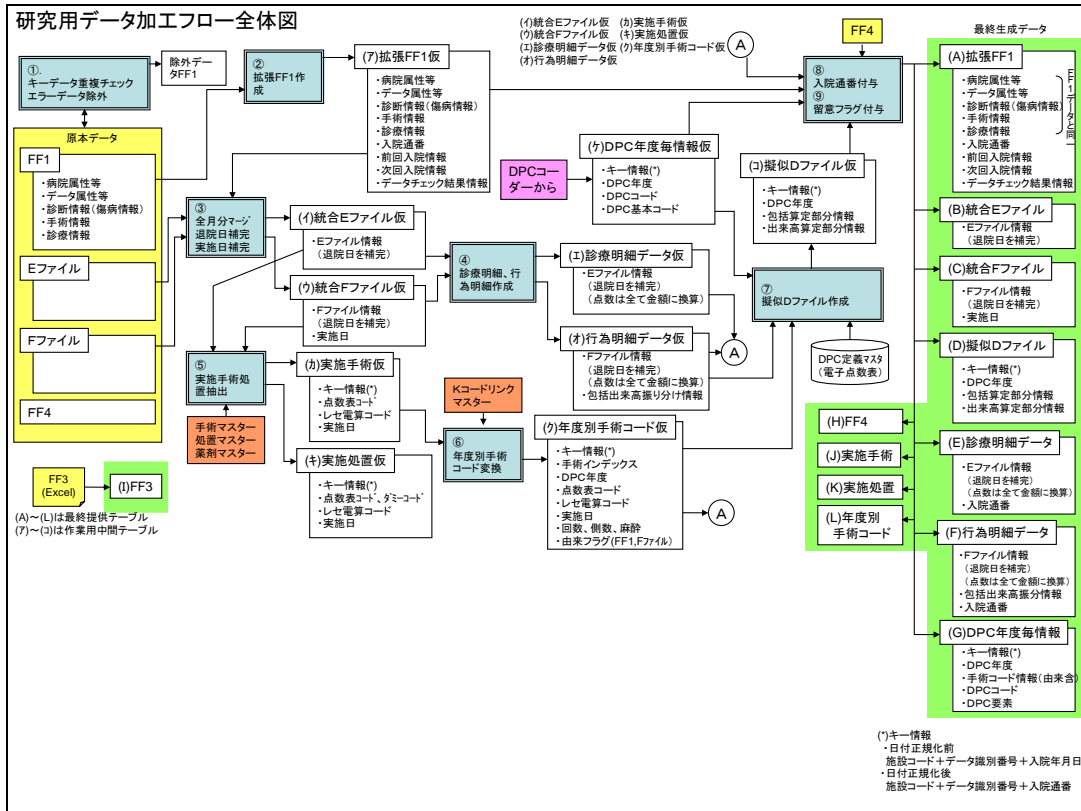
平成 20 年度「DPC 松田研究班版 DPC コーダーの開発について」で作成した DPC コーディングツールを利用した DPC コードを分析用データセット内に取り込み、他のデータと一体的に保有する

以上の処理を行い、分析用データセットを作成する。

分析用データセットには次のファイルが含まれる

- (A) 拡張 FF1 様式 1 (FF1) 情報に留意フラグ等の情報を付加したものの
- (B) 統合 E ファイル E ファイルに退院日及び期間内入院回数を付加したものの
- (C) 統合 F ファイル F ファイルに退院日・実施日等を付加したものの
- (D) 疑似 D ファイル D ファイル作成ルールに基づいて擬似的に D ファイルを生成
- (E) 診療明細データ 統合 E ファイルの点数部分を金額に置き換えたものの
- (F) 行為明細データ 統合 F ファイルの点数部分を金額に置き換えたものの
- (G) DPC 年度毎情報 運用時期別に生成された DPC コード情報

- (H) FF4 様式 4 出したもの
- (J) 実施手術 F ファイルから 手術関係のレコードのみを抽出したものの
- (L) 年度別手術コード 手術コードを診療報酬の運用年度別に変換したものの
- (K) 実施処置 F ファイルから 処置・薬剤関係のレコードのみを抽出し、データ処理のフローは下に示す図のとおりである。



C. 結果

本年度の研究期間においては平成 29 年度時点で伏見班保有のデータに対して B.方法で示した処理を行なうと同時に平成 30 年度データについても収集が終わった分について仮処理をおこなった。

以下特記すべき事項に関して記載する。

1. エラーチェックについて

本研究データセット作成については、以下のルールの下で、データエラーチェックおよび留意フラグ付与を行った。

表 1 エラーチェック仕様

No	エラー番号	チェック内容	チェック論理(NG条件)	対象データ	種別	備考	フラグ内容等
1	ERR010	FF1のキーが重複している	FF1 縦にて「施設コード+データ識別番号+入院年月日+回数管理番号+診療情報番号+コード+バージョン+連番」が重複(過去のFF1との重複もチェックする。)	FF1 FF1T	キー重複	<ul style="list-style-type: none"> ・重複データはFF1データを除外データFF1テーブルに、エラー情報と共に格納し、以降の処理には使用しない。 ・親様式1(診療情報番号=0)がエラーの場合、子様式1(診療情報番号≧1)のレコードも一緒に除外する。 ・子様式1の重複は該当の子様式1のみ除外とし、親様式1および他の子様式1には影響を及ぼさない。 ・「日帰り入院+同一日再入院」および「日帰り入院+日帰り入院+退院」の場合は、同一日再入院および最初の日帰り入院を採用する(回数管理番号が最小のもののみ採用する)。この場合は ERR010 ではなく、ERR20 として、非採用の様式1情報のみ除外とする。 	

2	ERROR20	FF4のキーが重複している	FF4にて「施設コード+データ識別番号+入院年月日」が重複	FF4	キー重複	<ul style="list-style-type: none"> 重複データはFF1データを除外データFF1テーブルに、エラー情報と共に格納し、以降の処理には使用しない。 親様式1と子様式1の両方を除外する。(過去に仮確定した子様式1も除外する。) 	
3	ERROR30	Eファイルのキーが重複している	Eファイルの「施設コード+データ識別番号+入院年月日+データ区分+順序番号+該当月」が重複	Eファイル	キー重複	<ul style="list-style-type: none"> 重複データはFF1データを除外データFF1テーブルに、エラー情報と共に格納し、以降の処理には使用しない。 親様式1と子様式1の両方を除外する。(過去に仮確定した子様式1も除外する。) 	
4	ERROR40	Fファイルのキーが重複している	Fファイルの「施設コード+データ識別番号+入院年月日+データ区分+順序番号+行為明細番号+該当月」が重複	Fファイル	キー重複	<ul style="list-style-type: none"> 重複データはFF1データを除外データFF1テーブルに、エラー情報と共に格納し、以降の処理には使用しない。 親様式1と子様式1の両方を除外する。(過去に仮確定した子様式1も除外する。) 	

5	ERR060	EF 統合ファイルのキーが重複している	EF 統合ファイルの「施設コード+データ識別番号+入院年月日+データ区分+順序番号+行為明細番号+該当月」が重複	EF 統合	キー重複	<ul style="list-style-type: none"> ・重複データはFF1 データを除外データFF1 テーブルに、エラー情報と共に格納し、以降の処理には使用しない。 ・親様式1と子様式1の両方を除外する。(過去に仮確定した子様式1も除外する。)
6	ERR060	外来Eファイルのキーが重複している	外来Eファイルの「施設コード+データ識別番号+入院年月日+データ区分+順序番号+該当月」が重複	外来E	キー重複	<ul style="list-style-type: none"> ・重複データはFF1 データを除外データFF1 テーブルに、エラー情報と共に格納し、以降の処理には使用しない。 ・親様式1と子様式1の両方を除外する。(過去に仮確定した子様式1も除外する。)
7	ERR070	外来Fファイルのキーが重複している	外来Fファイルの「施設コード+データ識別番号+入院年月日+データ区分+順序番号+行為明細番号+該当月」が重複	外来F	キー重複	<ul style="list-style-type: none"> ・重複データはFF1 データを除外データFF1 テーブルに、エラー情報と共に格納し、以降の処理には使用しない。 ・親様式1と子様式1の両方を除外する。(過去に仮確定した子様式1も除外する。)

	ERR080	Dファイルのキーが重複している	Dファイルの「施設コード+データ識別番号+入院年月日+データ区分+順序番号」が重複	Dファイル	キー重複	<ul style="list-style-type: none"> ・重複データはFF1データを除外データFF1テーブルに、エラー情報と共に格納し、以降の処理には使用しない。 ・親様式1と子様式1の両方を除外する。(過去に仮確定した子様式1も除外する。) 	
ERR080		Hファイルのキーが重複している	Hファイルの「施設コード+病棟コード+データ識別番号+退院年月日+入院年月日+実施年月日+コード+バージョーン+連番」が重複	Hファイル	キー重複	<ul style="list-style-type: none"> ・重複データはFF1データを除外データFF1テーブルに、エラー情報と共に格納し、以降の処理には使用しない。 ・親様式1と子様式1の両方を除外する。(過去に仮確定した子様式1も除外する。) 	

8	ERR110	<p>在院日数が1日未満である</p>	<p>入院日 > 退院日</p>	<p>FF1 FF1T</p>	<p>除外</p>	<p>・エラーデータは FF1 データを除外データ FF1 テーブルに、エラー情報と共に格納し、以降の処理には使用しない。 ・親様式1(診療情報番号=0)がエラーの場合、子様式1(診療情報番号≧1)のレコードも一緒に除外する。 ・子様式1のエラーの場合は、該当子様式1は除外するが、親様式1および他の子様式1は除外しない。 ・子様式1で退院日が00000000である場合は、許容する。親様式1はこのエラーとなる。</p>	
9	ERR120	<p>入院時年齢が0歳未満である</p>	<p>生年月日 > 入院日</p>	<p>FF1 FF1T</p>	<p>除外</p>	<p>・エラーデータは FF1 データを除外データ FF1 テーブルに、エラー情報と共に格納し、以降の処理には使用しない。 ・親様式1(診療情報番号=0)がエラーの場合、子様式1(診療情報番号≧1)のレコードも一緒に除外する。(過去に仮確定した子様式1も除外する。) ・子様式1のエラーの場合は、該当子様式1は除外するが、親様式1および他の子様式1は除外しない。</p>	

10	ERR130	年月日が誤っている	実在しない年月日 (13月1日、7月32日など)、および SQLserverの datetime型で扱えない日付(1753年1月1日以前)	ALL (FF1Tも)	除外	<ul style="list-style-type: none"> ・エラーデータはFF1データを除外データFF1テーブルに、エラー情報と共に格納し、以降の処理には使用しない。 ・対象は全ての日付情報とする ・あくまで実在しない年月日のチェックのみであり、手術日が入院日と退院日の間にあるか、などの相関チェックは行わない。 ・00000000は許容する。 ・親様式1(診療情報番号=0)がエラーの場合、子様式1(診療情報番号≧1)のレコードも一緒に除外する。(過去に仮確定した子様式1も除外する。) ・子様式1のエラーの場合は、該当子様式1は除外するが、親様式1および他の子様式1は除外しない。 ・様式1開始日、様式1終了日をチェック対象に追加。 ・前回退院年月日は日を00にする表記を共用しているため、本チェックの対象外とする。
11	ERR140	(欠番)				
12	ERR150	(欠番)				
13	ERR160	統括診療情報番号が異常である (3日以内再入院集約レコー	統括診療情報番号 が0以上の数字でない	FF1 FF1T	除外	<ul style="list-style-type: none"> ・他のエラーデータ除外より最初に判定する。 ・エラーデータはFF1データを除外データFF1テーブルに、エラー情報と共に格納し、以降の処理には使用しない。

14	ERR170	ド)	様式1対象期間が1日未満である	様式1開始日>様式1終了日	FF1 FF1T	除外	<ul style="list-style-type: none"> ・エラーデータは、FF1 データを除外データ FF1 テーブルに、エラー情報と共に格納し、以降の処理には使用しない。 ・親様式1(診療情報番号=0)がエラーの場合、子様式1(診療情報番号≧1)のレコードも一緒に除外する。 ・子様式1のエラーの場合は、該当子様式1は除外するが、親様式1および他の子様式1は除外しない。
15	ERR180		親様式1において様式1開始日、様式1終了日が入院日、退院日に一致していない	統計診療情報番号=0 かつ (様式1開始日が入院年月日と一致しない または 様式1終了日が退院年月日と一致しない)	FF1 FF1T	除外	<ul style="list-style-type: none"> ・エラーデータは、FF1 データを除外データ FF1 テーブルに、エラー情報と共に格納し、以降の処理には使用しない。 ・親様式1(診療情報番号=0)がエラーの場合、子様式1(診療情報番号≧1)のレコードも一緒に除外する。(過去に仮確定した子様式1も除外する。) ・子様式1のエラーの場合は、該当子様式1は除外するが、親様式1および他の子様式1は除外しない。

16	ERR190	<p>様式1開始日 が入院と退院 の範囲外であ る</p>	<p>様式1開始日<入 院年月日 または 様式1開始日 > 退院年月日</p>	<p>FF1 FF1T</p>	除外	<p>・エラーデータは FF1 データを除外データ FF1 テーブルに、エラー情報と共に格納し、以降の処理には使用しない。 ・親様式1(診療情報番号=0)がエラーの場合、子様式1(診療情報番号≧1)のレコードも一緒に除外する。 ・子様式1のエラーの場合は、該当子様式1は除外するが、親様式1および他の子様式1は除外しない。</p>	
17	ERR200	<p>様式1終了日 が入院と退院 の範囲外であ る</p>	<p>様式1終了日<入 院年月日 または 様式1終了日 > 退院年月日</p>	<p>FF1 FF1T</p>	除外	<p>・エラーデータは FF1 データを除外データ FF1 テーブルに、エラー情報と共に格納し、以降の処理には使用しない。 ・親様式1(診療情報番号=0)がエラーの場合、子様式1(診療情報番号≧1)のレコードも一緒に除外する。 ・子様式1のエラーの場合は、該当子様式1は除外するが、親様式1および他の子様式1は除外しない。</p>	

							<ul style="list-style-type: none"> ・他のエラーデータ除外より最初に判定する。 ・エラーデータは FF1 データを除外データ FF1 テーブルに、エラー情報と共に格納し、以降の処理には使用しない。 	
ERR210	回数管理番号が異常である	回数管理番号が0以上の数字でない	FF1T	除外	除外	<ul style="list-style-type: none"> ・エラーデータは FF1 データを除外データ FF1 テーブルに、エラー情報と共に格納し、以降の処理には使用しない。 20_FF1_キー重複チェックでチェック実施。 		
ERR220	同じ入院日の入院が複数存在する	回数管理番号が複数存在する。 最小の回数管理番号のみの様式1情報のみを採用し、他の回数管理番号の様式1情報はERR220として除外する。	FF1T	除外	除外			
18	ERR510	EファイルとFファイルが不整合である(Fファイルデータが存在する)	Eファイル、Fファイル	フラグ	フラグ	<ul style="list-style-type: none"> ・親様式1(診療情報番号=0)のみチェック対象。 ・子様式1(診療情報番号≧1)はチェック対象外。フラグは0を設定する。 	0:エラー無し 1:Fファイルデータ欠落有り	

			在しない。						
19	ERR520	EファイルとFファイルが不整合である(Eファイルデータ非存在である)	Fファイルの各レコードに対して、同じ月のEファイルに同じ「施設コード+データ識別番号+入院年月日+データー区分+順序番号」を持つレコードが存在しない。	Eファイル Fファイル	フラグ	・親様式1(診療情報番号=0)のみチェック対象。 ・子様式1(診療情報番号≧1)はチェック対象外。フラグは0を設定する。	0:エラー無し 1:Eファイルデータ欠落有り		
20	ERR530	入院期間外のEファイルレコードが存在する	Eファイルの実施日<FF1の入院日またはFF1の退院日<Eファイルの実施日	FF1Eファイル、統合Eファイル	フラグ	・親様式1(診療情報番号=0)のみチェック対象。 ・子様式1(診療情報番号≧1)はチェック対象外。フラグは0を設定する。	0:エラー無し 1:入院期間外Eファイルデータ有り		

21	ERR540	Eファイル、Fファイルの退院日がFF1と異なる	「施設コード+データー識別番号+入院年月日」がFF1とEファイル、Fファイルで同一であるが、退院日が異なる。	FF1,FF1 T,Eファイル、Fファイル EF 統合ファイル	フラグ	<ul style="list-style-type: none"> 親様式1(診療情報番号=0)のみチェック対象。 子様式1(診療情報番号≧1)はチェック対象外。フラグは0を設定する。 	<ul style="list-style-type: none"> 0:エラー無し 1:Eファイルの退院日がFF1と異なる 2:Fファイルの退院日がFF1と異なる 3:EファイルとFファイルの両方の退院日がFF1と異なる 4:EF 統合ファイルの退院日がFF1と異なる
22	ERR541	Dファイルの退院日がFF1と異なる	「施設コード+データー識別番号+入院年月日」がFF1とDファイルで同一であるが、退院日が異なる。	FF1,FF1 T,Dファイル	フラグ	<ul style="list-style-type: none"> 親様式1(診療情報番号=0)のみチェック対象。 子様式1(診療情報番号≧1)はチェック対象外。フラグは0を設定する。 	<ul style="list-style-type: none"> 0:エラー無し 1:Dファイルの退院日がFF1と異なる
22	ERR542	Hファイルの退院日がFF1と異なる	「施設コード+データー識別番号+入院年月日」がFF1とHファイルで同一であるが、退院日が異なる。	FF1,FF1 T,Hファイル	フラグ	<ul style="list-style-type: none"> 親様式1(診療情報番号=0)のみチェック対象。 子様式1(診療情報番号≧1)はチェック対象外。フラグは0を設定する。 	<ul style="list-style-type: none"> 0:エラー無し 1:Hファイルの退院日がFF1と異なる

23	ERR550	入院基本料または特定入院料を算定しない日がある	Fファイルでデータ区分90(入院基本料)または92(特定入院料)の点数がない日がある。	Eファイル	フラグ	<ul style="list-style-type: none"> ・親様式1(診療情報番号=0)のみチェック対象。 ・子様式1(診療情報番号≧1)はチェック対象外。フラグは0を設定する。 	<p>0:エラー無し</p> <p>1:入院基本料または特定入院料を算定した前に、入院基本料または特定入院料を算定しない日がある(ただし、2のケースを除く)</p> <p>2:入院基本料または特定入院料を算定した後に、退院前に入院基本料または特定入院料を算定しない日がある</p>
24	ERR610	医科レセプトのみでない	FF4で「1.医科レセプトのみ」以外	FF4	フラグ	<ul style="list-style-type: none"> ・FF4にレコードがない場合もフラグ設定する。 ・親様式1(診療情報番号=0)のみチェック対象。 ・子様式1(診療情報番号≧1)はチェック対象外。フラグは0を設定する。 	<p>0:エラー無し(医科レセプトのみ)</p> <p>2:歯科レセプトあり</p> <p>3:保険請求なし</p> <p>4:保険と他制度の併用</p> <p>5:その他</p> <p>9:FF4に対応レコード無し</p>

25	ERR710	24時間以内の死亡である	FF1で「24時間以内死亡の有無」が「24時間以内死亡の有り」または「救急患者として搬送され、入院前に処置室、手術室等で死亡有り」	FF1 FF1T	フラグ	0:エラー無し(24時間以内死亡無し) 1:24時間以内死亡有り 2:救急患者として搬送され、入院前に処置室、手術室等で死亡有り
26	ERR720	生後7日以内の死亡である	退院時転帰が死亡、かつ退院日が生年月日から7日以内(生年月日当日は1日とカウント)	FF1 FF1T	フラグ	0:エラー無し(生後7日以内の死亡無し) 1:生後7日以内の死亡有り
27	ERR730	治療対象である	FF1で「治療実施の有無」が「有り」	FF1 FF1T	フラグ	0:エラー無し(治療実施無し) 1:治療実施有り
28	ERR740	移植手術を実施している	Fファイナルに移植手術に該当するコードに該当するレセコードを持つ。移植手術のKコードは厚労省告	Fファイナル	フラグ	0:エラー無し(移植手術無し) 1:移植手術有り

30	ERR760	一般病棟外への移動がある	FF1で「一般病棟外への移動あり」がある(～H16)。「精神病棟への入院あり」または「その他の病棟への入院有り」がある(H17～)。	FF1 FF1T	フラグ	0:エラー無し 1:一般病棟外への移動あり
31	ERR770	年齢が120歳以上である	入院時年齢が120歳以上	FF1 FF1T	フラグ	0:エラー無し(入院時年齢120歳未満) 1:入院時年齢120歳以上
32	ERR780	手術が輸血のみである	FF1の手術1～5に輸血だけが存在しない(輸血管理料は含まないこと)	FF1 FF1T	フラグ	0:エラー無し(手術なし、または、輸血以外の手術あり) 1:手術が輸血のみ
33	ERR790	短期滞在手術基本料を算定している	Eファイルに短期滞在手術基本料に該当するレセ電コードがある。	Eファイル	フラグ	0:エラー無し(短期滞在手術基本料なし) 1:短期滞在手術基本料あり

ERR800 (旧 ERR140)	診療報酬制度 変更跨り (2010年度か らの調査通年 化により、条 件変更)	様式1開始日・終 了日が診療報酬年 度を跨っている。 (2016年度データ では、様式1開始 日が2016/4/1以 前である。)	FFIT	フラグ	H23 新規追加、H24 変更、H26 変更、H28 変更	0:エラー無し エラー有りの場合は、跨った診 療報酬年度を設定(2009年度→ 2010年度跨りなら2010を設定)
-----------------------------	--	---	------	-----	-------------------------------	---

34

D. 考察

本研究の結果から、提出されているデータについて、一定数のエラーが含まれていることがわかった。特にキー情報の重複や必要なデータが欠損している症例などは、分析に影響度が大きいので、それを確実に除去できるようになったことは成果である。

また、分析の方向性によって、使用の可否が決まる入院患者レコードがあるということが本研究の過程で判明した。たとえば、入院の途中で保険適応になる患者について、それを分析に含めるか否か、その場合の在院日数はどの範囲を指すべきなのか、などといった点は、研究の目的に応じて、それを研究者が容易に判断できる環境にあることは結果の妥当性を維持するためにも重要である。今回の研究で、留意が必要な入院レコードに対してそのフラグメント化ができたことは大きな成果であったといえる。今後の開発においてもさらなる留意コードが必要かについてその使い勝手とともに検

討していきたい。

現時点ではいまだにいくつかのテーブルについては一般的な研究者の持つ環境ではハンドルが難しいサイズのレコード数を持っている状況にある。今後、このデータセットを使って、研究者が共通して使える集計データを作成し、より容易に研究ができるデータ環境を構築していく必要がある。

E. 結論

本年度、平成 29 年度分までの分析用データセットの作成が完了し、運用することもできた。また、平成 30 年度以降のデータについても同様に処理が出来るシステムの稼働を行うことが出来た。

次年度以降、より容易に研究ができるデータ環境を構築していく研究を行っていきたい。