医療情報学 (2011.07) 30巻4号:241～250.

DPC導入の影響評価に係る調査データに関するセキュリティ考察

山上 浩志, 廣川 博之

資 料

DPC導入の影響評価に係る調査データに関するセキュリティ考察

山上 浩志*1 廣川 博之*1

2003年4月に始まったDPC（Diagnosis Procedure Combination）に基づく包括評価による入院医療費の定額支払い制度は，2010年7月1日時点で一般病院全体の18.0%に相当する1,391施設に適用されている．厚生労働省はこれら施設を対象にした「DPC導入の影響評価に係る調査」を通じて，DPC分類の精緻化，DPCごとの診療報酬点数の決定や病院係数の決定等を行うためのデータ収集を継続的に行っている．

本稿では，この調査システムの運用前提になっている医療機関が行うべき提出データの連結可能匿名化について，ある都道府県内のDPC対象病院を対象に実施したアンケート調査に基づき，31.4%の施設が何ら匿名化せずにデータを提出している実態を報告する．更に，現行調査システムにおけるセキュリティ上の脆弱さと医療機関側で想定すべきリスクを指摘し，今後の調査システムのあり方について提言を述べる．

■キーワード：DPC，連結可能匿名化，DPC調査，セキュリティ，プライバシー

Security and Privacy Concerns on DPC Surveys Conducted by the Ministry of Health, Labor and Welfare: Yamakami H*1, Hirokawa H*1

The medical payment system based on Diagnosis Procedure Combination (DPC), which is a Japanese prospective payment system, was introduced in April 2003. By July 2010, 1,391 hospitals, which cover 18.0% of all hospitals, implemented this DPC-based payment system. The Ministry of Health, Labor and Welfare periodically conducts an impact assessment survey (hereafter referred to as a DPC survey) at those hospital to update DPC classification categories and also determine, for instance, payment weight and hospital adjustment factors. To each DPC survey, the hospitals report patients' medical data, which should be pseudonymized by certain pseudonymization methods according to the ministry's guideline.

We surveyed DPC-based hospitals in a prefecture and conclude that 31.4% of hospitals have failed to pseudonymize personal data when reporting them to DPC surveys. We find security vulnerabilities and privacy concerns in the current DPC survey system. Moreover, we propose a way to make the survey system safer in the future.

*1旭川医科大学病院 経営企画部
〒078-8510 旭川市緑が丘東2条1丁目1番1号
受付日：2010年12月15日
採択日：2011年6月3日
E-mail：yamakami-ash@umin.ac.jp

*1Department of Medical Informatics, Asahikawa Medical University Hospital
2-1-1-1 Midorigaoka-higashi, Asahikawa 078-8510, Japan

Key words: Diagnosis Procedure Combination (DPC), Pseudonymization, DPC Survey, Security, Privacy

1. はじめに

2003 年 4 月に特定機能病院を中心とした 82 施設に導入された DPC（Diagnosis Procedure Combination；診断群分類）に基づく包括評価による入院医療費の定額支払い制度※1は，2010 年 7 月 1 日時点で一般病院全体（7,714 施設；2008 年医療施設調査）の 18.0％にあたる 1,391 施設に適用されている．

これら DPC 対象病院（診断群分類点数表によって算定している病院を指す）と DPC 準備病院（対象病院ではないが後述する DPC 関係調査に参加している病院を指す）は，厚生労働省（以下，厚労省という）が実施する「DPC 導入の影響評価に係る調査」（以下，厚労省 DPC 関係調査という）に適切に参加することが課せられている．厚労省はこの調査を「DPC については，わが国における急性期入院医療に初めて本格的に導入された包括評価制度であることから，中央社会保険医療協議会の付託を受け，DPC 評価分科会のもと本調査を実施し，制度導入の影響評価を行うとともに，診断群分類の継続的な見直しのための資料とするものである」と位置付けており[1]，この調査データに基づき DPC 分類の精緻化，DPC ごとの診療報酬点数の決定や病院係数の決定が行われる[2]．厚労省はまた，DPC データの活用によって，患者の入退院の動態として，例えば短期間で再入院する患者数や退院時転帰の「治癒」割合を把握できることや特定の診断群分類の患者の診療内容と他の医療機関の平均的診療内容の比較が行える等，医療機関における診療内容の評価が可能になることを説明する[3]と共に，一部の調査集計データを自らのウェブサイトで公表している[1]．

厚労省 DPC 関係調査により収集されるデータは患者別匿名化情報とされ，データを提出する医療機関側で連結可能匿名化処理を行う必要がある[4]．匿名化の方法には「必要な場合に個人を識別できるように，その人と新たに付された符号または番号の対応表を残す方法による匿名化（連結可能匿名化）」と「個人を識別できないように，その人と新たに付された符号または番号の対応表を残さない方法による匿名化（連結不可能匿名化）」とがあり[5]，前者においては「対応表」の機密水準が匿名性の強度を決めることから，臨床試験データやヒトゲノム・遺伝子解析データ等の医学研究では，「対応表」は当該研究に直接従事しない個人情報管理者によって管理され，且つ金庫内で厳重に保管されるような運用が行われている．

本稿では，厚労省 DPC 関係調査における連結可能匿名化の実態について，著者らが行った調査をもとに報告し，そこから見えた同調査システムの運用における問題点を指摘すると共に，今後のシステムのあり方について提言を述べる．

2. 厚労省 DPC 関係調査の概要[4]

1）提出データファイル

厚労省 DPC 関係調査において，医療機関側が用意するデータは様式 1, 3, 4 および D, E, F ファイルの 6 種類である※2．

- 様式 1：主傷病名，入院の目的，手術術式等の診療録情報
- 様式 3：病床数，入院基本料，算定状況等の

※1 2010 年 12 月 16 日に開催された中央社会保険医療協議会の診療報酬調査専門組織・DPC 評価分科会において，この制度を DPC/PDPS（Diagnosis Procedure Combination/Per-Diem Payment System）と呼称することが決定された．

※2 調査は 2003 年度に始まったが，提出するデータ種類には変遷があり，様式 5 や 6, G ファイルと呼ばれたデータファイルも一時期存在した．現在の 6 種類に固まったのは 2006 年度からで，2010 年度調査では E ファイルと F ファイルを統合した「EF ファイル」を提出するようになった．

施設調査票

• 様式4：医科保険診療以外のある症例調査票

• Dファイル：診断群分類点数表により算定した患者に係る診療報酬請求情報（包括レセプト情報）

• Eファイル：医科点数表に基づく出来高レセプト情報（診療明細情報）

• Fファイル：医科点数表に基づく出来高レセプト情報（行為明細情報）

このうち，様式3を除くデータファイルは患者単位に把握する情報であって，医療機関内で患者データを管理するために使われている「患者ID番号」と連結可能なように匿名化された「データ識別番号」（10桁の数字列）を用いて作成される必要がある．このデータ識別番号は全ファイルで整合が図られていなければならず，調査期間に相当する単一年度内に変更することは認められていない．

2）データ提出手順

医療機関側では各データファイルを作成後，厚労省から配布されたソフトウェアを用いてエラーチェックを行い，その時生成されるパスワードの付けられた1つの圧縮ファイルをMO媒体に格納し，郵便書留またはそれと同等の配達記録が残る宅配便でDPC調査事務局宛てに送付する．この時，MO媒体に貼付するラベルには施設名，施設コード（都道府県番号と医療機関コードからなる9桁数字列）とデータ当該月を記すことになっている．

データ提出は7月から翌年3月まで毎月行われ，提出したデータに対する疑義照会が医療機関にCD媒体で返送される場合がある．そこにはExcelやPDF等の形式でのファイルが含まれるが，ファイルを個別にもしくは全体を一塊に暗号化するといった処置は施されていない．

3. 方　法

都道府県A内のDPC対象病院71施設に対して，**表1**に示す質問項目からなるアンケート調査票をDPC関係調査担当者宛てに郵送し，その回収後には必要に応じ，回答内容に関する問い合わせを電話で行った．実際の調査票様式では，製品名称の引用や回答の例示等，補助的文言を添えた設問としていたが，表中では割愛している．また，調査実施にあたり，収集したデータは研究目的でのみ用いること，研究成果を発表する場合には回答内容を施設名称と関連づけて公表しないことを説明したうえで調査への協力を依頼した．

調査は2010年5月に実施し，51施設から回答を得た（回収率71.8%）．

4. 結　果

1）調査回答施設の分布（問1）

回答の得られた51施設について，DPC算定病床数は40～874（中央値292）床であり，様式1における提出データ件数は55～1,284（中央値450）件/月であった（**図1**）．

表2には病院類型別にみた回答施設の分布を示す．本調査は限定地域で行われたものであるが，DPC対象病院全体の分布と比較した場合，200床以上規模で良い近似がみられた．

2）様式1データの作成方法（問2）

様式1データの作成方法では，病院情報システム（オーダエントリや医事会計システム）に組み込まれた機能を用いていた施設が28（54.9%），様式1入力支援ソフトウェアを使う施設が19（37.3%）あり，残り4施設はFileMaker（ファイルメーカー）を用いて独自に作成したソフトウェアを使用していた（**表3**）．

3）データ識別番号の生成手法（問3，問4）

データ識別番号を生成するアルゴリズム（計算方法）は**表4**に示す［イ］から［ヘ］の6つに分類された．表中には，7桁の患者ID番号「1234567」からデータ識別番号を生成した例示と共に，各手法を採用していた施設数を「回答数(問3)」欄に示した．

［イ］桁揃えを採っていた施設が16（31.4%）と最も多く，次いで［ロ］加算が13（25.5%）あり，続く［ハ］方式不詳とはDPC関係調査担当者がデータ識別番号の生成手法をわからないと

表1　アンケート調査票

設問	質問内容	回答
問1	貴院での基礎的データをお答えください.	記述　DPC導入年度，DPC算定病床数，様式1での提出データ件数（平均 件/月）
問2	様式1をどのように作成していますか.	選択　{a. 病院情報システム（オーダや医事）に備わる機能を利用している　b. 様式1入力支援ソフトを利用している　c. 病院が独自に開発したソフトウェアを用いている}
問3	データ提出時の「データ識別番号」（10桁数字）はどのような計算方法（アルゴリズム）で生成していますか.	記述　（患者ID番号との対応が分かるように記載ください）
問4	問3でのアルゴリズムはどのようにして決められましたか.	選択　{a. メーカまたはソフトウェアが元々提供する機能をそのまま利用している　b. 病院が独自に考案したものを反映させている}
問5	貴院で採用するアルゴリズムに基づく情報匿名性について，「厚労省DPC関係調査」，「連結可能匿名化」の中で考えたときに，どう評価しますか.	選択　{a. 大いに良好である　b. ある程度良好である　c. 余り良好とは言えない　d. 全く良好でない　e. わからない}
問6	貴院ではASP*型DPCデータ分析サービスを利用していますか. *ASP：アプリケーションサービスプロバイダ	選択　{a. はい　b. いいえ}
問7	（問6ではいと答えた場合）サービス業者に送付するデータに含まれる「データ識別番号」は厚労省DPC関係調査向けと同一ですか.	選択　{a. はい　b. いいえ}

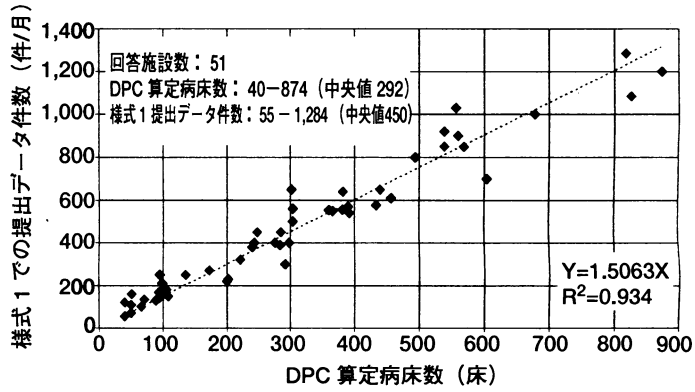

図1　調査回答施設におけるDPC実施規模

答えたもので12（23.5%）あった．残りの10（19.6%）施設では，[ニ] 桁立て，[ホ] 反転や [ヘ] 左シフトの手法がみられた．

各施設が採用する手法について，44施設（86.3%）がシステムメーカの標準的に提供するソフトウェア機能に従っており，そこに病院側の意図が反映されていたのは7施設（13.7%）に過ぎなかった．

4）提出データ匿名性の主観的評価（問5）

提出データの匿名性について，回答施設が自己

表 2　病院類型別にみた施設分布

病院類型 (DPC 算定病床数による)	100 床未満	100 床以上 200 床未満	200 床以上 300 床未満	300 床以上 400 床未満	400 床以上 500 床未満	500 床以上	計
調査回答施設	12 (23.5%)	4 (7.8%)	11 (21.6%)	9 (17.7%)	5 (9.8%)	10 (19.6%)	51
2010 年度 DPC 対象病院全体	154 (11.1%)	288 (20.7%)	284 (20.4%)	244 (17.5%)	146 (10.5%)	275 (19.8%)	1,391

表 3　様式 1（診療録情報）の作成方法

作成方法	回答数（表 1 での質問項目） (問 2)
病院情報システム（オーダや医事）に備わる機能を利用している	28
様式 1 入力支援ソフトウェアを利用している	19
病院が独自に開発したソフトウェアを用いている	4

表 4　データ識別番号の生成手法

手法			回答数（表 1 での質問項目）	
分類	説明	生成例	(問 3)	(問 6)
［イ］桁揃え	先頭を 0 埋め	1234567→0001234567	16	8
［ロ］加算	1 を加算	1234567→0001234568	13	8
［ハ］方式不詳	アルゴリズムの詳細不明	—	12	4
［ニ］桁立て	未使用桁に数字をセット	1234567→0101234567	5	3
［ホ］反転	数字を反転	1234567→0007654321	3	3
［ヘ］左シフト	10 の冪数を乗算	1234567→0123456700	2	1

評価した結果を**表 5** に示す．これを｛大いに/ある程度良好である｝，｛余り/全く良好でない｝，｛わからない｝の三区分に再集計した上で，上位 3 手法（［イ］桁揃え，［ロ］加算および［ハ］方式不詳）を採用する施設の回答について，カイ二乗検定を用いて分析したところ，帰無仮説：「回答割合に違いがない」に対して，［イ］－［ハ］：9.90（棄却；$p<0.01$），［ロ］－［ハ］：5.77（採択；$p<0.05$），［イ］－［ロ］：1.69（採択；$p<0.05$）となり，「桁揃え」と「方式不詳」の施設間では回答傾向に有意な違いを認めたが，「加算」と「方式不詳」，「桁揃え」と「加算」の施設間の回答傾向には違いを認めなかった．

5）DPC データ分析サービスの利用状況（問 6，問 7）

DPC データ分析のサービスを利用していた施設は 27（52.9%）あり，データ識別番号の生成手法別の内訳を**表 4** の「回答数（問 6）」欄に示した．いずれの施設も，サービス提供者側には厚労省 DPC 関係調査用と同一のデータを提出していた．

表 5　提出データの匿名性評価

	a. 大いに良好である	b. ある程度良好である	c. 余り良好でない	d. 全く良好でない	e. わからない
［イ］桁揃え	1	3	3	1	8
［ロ］加算	0	5	1	0	7
［ハ］方式不詳	2	8	0	0	2
［ニ］桁立て	0	1	3	1	0
［ホ］反転	0	3	0	0	0
［ヘ］左シフト	0	0	0	0	2

5. 考　察

1）連結可能匿名化の実態

データ識別番号を患者 ID 番号の先頭に 0 埋めしただけの「桁揃え」で生成していた施設は 31.4%に上り，これは何ら匿名化を行っていないに等しい．その他の施設でも「加算」,「桁立て」や「左シフト」といった単純な処理が行われており，データが大量にあった場合には生成手法を容易に類推できてしまう可能性が高い．「方式不詳」の施設にあっては，データ識別番号と患者 ID 番号との関係性を少なくとも一見して判らない程度の処理が施されていると思われた．

医療機関の多くがこうした手法については無関心であり，提出用データはシステムメーカから提供される機能に頼って作成されている状況が窺えた．データ識別番号の生成方式は，「加算」や「反転」が特定のメーカ製システムにだけみられるなど，システムメーカによって特徴があり，調査地区以外の地域でも同様な実態がある可能性を示唆していた．

提出用データは単一のメーカ製システムで生成されるとは限らず，実際，医事システムと DPC 入力支援システムとを併用しながら提出用データを準備している施設が小規模病院を中心に多くみられた．こうした施設においては，システムメーカに固有な手法を適用すると，一連の提出ファイル間でデータ識別番号を共通化するのが難しくなる．「桁揃え」処理に留めるシステムメーカの多い背景にはこうした事情が関係していると考えられた．また，「加算」は，患者 ID 番号の下 1 桁目にはチェックディジットが設定されることが一般的であるため，1 を足すことによって院内に実在しない患者 ID 番号に変換されることを利用した発想と思われた．

2）提出データ匿名化の必要性

厚労省 DPC 関係調査の実施要領[4]には，「患者別の情報については，匿名化を行った上で提出すること．ただし，提出データについてはデータの品質管理上疑義照会を行う必要があるため，連結可能匿名化（医療機関内において，匿名化情報と実データとの対応表を管理し，カルテなどの原資料が確認できる方法）を用いる必要がある」,「データ識別番号はカルテ番号である必要はない．院内で利用する患者 ID と連結可能な匿名化番号を使用することが望ましい」と記されている．しかしながら，この説明個所の解釈と具体的な対応は医療機関によって区々である．

個人情報の匿名化とは，「当該個人情報から，当該情報に含まれる氏名，生年月日，住所等，個人を識別する情報を取り除くことで，特定の個人を識別できないようにすること」[6]であって，個人識別における秘匿性とはなにも患者 ID 番号だけに依存したものではなく，様式 1 における基本属性情報でいえば，データ識別番号以外にも施設コード，性別，生年月日，患者住所地域の郵便番号等が該当する．例えば，人口密度が疎な地域においては性別，年齢や受診時期で，あるいは更

に少ない情報のみで患者が特定される場合があり，たとえ患者ID番号だけを匿名化（暗号化）したとしても，特定の個人を識別できてしまう可能性がある．

個人情報の保護に関する法律[7]の第23条1-四項によれば，「国の機関若しくは地方公共団体又はその委託を受けた者が法令の定める事務を遂行することに対して協力する必要がある場合であって，本人の同意を得ることにより当該事務の遂行に支障を及ぼすおそれがあるとき」には個人情報取扱事業者は本人の同意なしに，個人データを第三者へ提供できるとされている．つまり，法令に基づく厚労省DPC関係調査として第三者提供が可能と解釈でき，本来，匿名化することなしにもデータ提出できるのであるが，敢えて厚労省はデータ提出時に匿名化を義務づけていることになる．

こうしたデータ匿名化の視点とは別に，厚労省DPC関係調査システムには次のような伝送路上でのセキリティの脆弱性が存在している．DPC調査事務局宛てに提出されるMO媒体内の圧縮データファイルにはパスワードが掛けられて一見安全そうであるが，著者らの行った実験では，フリーソフトウェアを用いて僅か十数秒で圧縮ファイルのパスワードが解読できてしまった．また，その逆向きの経路では，疑義照会のために返送されるCD媒体もしくはデータファイルにはデータセキュリティ上の考慮が一切なされていない．

診療録等の記録された可搬媒体が搬送される際の個人情報保護について，厚労省ガイドライン[8]では，診療録等を記録した可搬媒体の遺失防止，診療録等を記録した可搬媒体と他の搬送物との混同の防止，搬送業者との守秘義務に関する契約，を実施すべき事項として挙げている．これは，医療機関が診療録等を外部に保存するユースケースを想定した内容であるが，この場合にはデータを運搬する業者との契約の内で，運搬する内容物，遺失の危険性を軽減する具体的な処置，守秘義務や責任分界点等を明確化しやすい．これに対し，厚労省DPC関係調査における可搬媒体の搬送手段である郵便書留では，万一事故が起きた場合の損害賠償額として，500万円を上限とした実損額と定められるだけである．個人情報漏洩インシデントにおける平均想定損害賠償額を一人あたり約5万円[9]とすると，今回調査施設における損害額は275万円～6,420万円（中央値2,250万円）ともなり，わが国の郵便システムがいかに信頼性に優れているとしても，それのみに頼るセキュリティは十分な担保にはなり得ない．

3）調査データの他目的利用

匿名化の不十分な調査データを厚労省へ提出する限りにおいては，個人情報の第三者提供には当たらないとしても，それを営利目的事業者等に渡す場合は，同一解釈とはならないことに注意を要する．

今回の調査から，52.9％の施設が何らかのDPCデータ分析サービスを利用しており，そうしたサービス提供者に厚労省DPC調査データをそのまま渡している事実が明らかになった．DPCデータ分析サービスの提供者にはデータ分析を生業とする会社の場合もあれば，グループ系病院のデータを集約し組織的に分析する形態もあろうが，サービス提供者側に医療機関名と一緒に集まる調査データに対し，［イ］桁揃え，［ロ］加算，［ニ］桁立て，［ホ］反転，［ヘ］左シフトの5通りの逆変換を試せば，そのうち1つは実患者ID番号に符合してしまう可能性がある．また，患者ID番号の同定可能性の有無に関わらず，仮に医療機関側が毎年同じアルゴリズムに基づいてデータ識別番号を生成し，且つ同じサービス提供者へ分析業務を委託していた場合には，患者の経年データがデータ分析サービス提供者側に蓄積されていくことになる．当然，サービス提供者側では善管注意義務のもと，収集したデータを適正に管理，運用するはずであるが，データ提出以後は医療機関がそこに介入することは不可能である．

DPCデータ分析サービス会社と交わす契約の一例をみると，医療機関（甲）は委託先業者（乙）に医療機関，学会，企業，患者等の個人にとって有益なコンサルティング，情報システムサービ

ス，研究，執筆，学会発表等を目的としたデータ利用を認め，甲は乙に個人を識別することができる情報を提出しない．そのためにデータに含まれる個人を識別することができる情報（氏名，住所，生年月日等）を削除または暗号化すること．万一，甲が第三者からクレーム等を請求された場合でも乙に責任が発生しないこと等が明記されている．

前述したとおり，厚労省 DPC 関係調査データにおける患者 ID 番号個所だけを暗号化したところで，データ全体が匿名化されたとはいえない．外部の DPC データ分析サービスを利用する以上は自院のデータを持ち出すことは避けられないが，各医療機関はそれと引き換えに大きなリスクを背負い込んでいることを自ら認識し，十分慎重に行動する必要がある．

このように，厚労省 DPC 関係調査データを他目的に転用する場合，その方法や倫理面等の問題が種々あるが，本稿ではここまでの指摘に留め，以降では厚労省へのデータ提出を前提に論を進める．

4）データ識別番号生成の例示

調査では，「桁揃え」を採用していた施設の担当者は自施設における生成データの匿名性を良好でないもしくはわからないと評価していた一方で，「方式不詳」を採っていた施設の担当者は良好と評価する傾向がみられた．この差異は，患者 ID 番号をもとに符号化された 10 桁数値について，一見してその成り立ちを推測できるか否かによる違いと思われた．

個人情報取扱事業者は取り扱う個人データの漏えい，滅失またはき損の防止その他の個人データの安全管理のために必要かつ適切な措置を講じなければならない（個人情報の保護に関する法律，第 20 条）が，提出するデータの匿名性について自ら安心できないようでは個人データの安全な管理に関して最低限の説明責任を果たせていないことになる．

臨床試験等の医学研究領域で行われている連結可能匿名化といえば，匿名化符号と原符号との間に対応表を持ち，その対応表を用いない限り再連結ができない手法を指すが，そのためには堅牢なシステム設計と同時に，組織内での安全な運用体制の確立も必要となる．1,400 もの病院が参加して全国規模で実施される DPC 関係調査において，そこまで厳格な対応を厚労省が求めているとは思えない．

「方式不詳」以外の施設で実装されていた「連結可能匿名化」手法は，患者 ID 番号に簡易な演算加工を施す暗号化を行いデータ識別番号としていた．患者 ID 番号集合を $X:\{X_1, X_2, \cdots, X_N\}$，データ識別番号集合を $Y:\{Y_1, Y_2, \cdots, Y_N\}$（X と Y はいずれも正の整数集合）とすると，X は 5〜8 桁，Y は 10 桁であるから，これらを 1 対 1 に変換するような演算を考えた時に，元より大きな自由度はない．例えば，X と Y との変換に簡単な一次変換式を用いた場合，隠された 2 定数 a, b を使い $Y_k=aX_k+b$ となり，二次変換式であれば，隠された 3 定数 a, b, c を使って $Y_k=aX_k^2+bX_k+c$ となる．この変換式（暗号学でいうアルゴリズム）と隠された定数組み（暗号学でいう鍵）とを知らない限り Y から X への逆変換ができない．このことが暗号化であって，より高次式の方が多くの定数が関与するため暗号性は強まるが，高次式では計算が複雑になり，計算上の桁溢れが起こりやすいことや変曲点の存在を意識しながら定数組みを選定する必要があること等の考慮が必要になってくる．

実は「桁揃え」，「加算」，「桁立て」，「左シフト」方式は，各々 $Y_k=X_k$, $Y_k=X_k+b$, $Y_k=X_k+b$, $Y_k=aX_k$ で表される 1 定数による一次変換式とみなせる．同じ一次変換でも 2 定数を用いた $Y_k=aX_k+b$ を用いると，例えば $(a, b)=(-2,9638527410)$ を選べば X＝1234567 から Y＝9636058276 へと変換でき，先行 4 方式に比べれば変換前後の数字列に関係を類推しづらくなっている．

2 定数を用いたところで，高々線形変換なことから，データが大量にあって，特に，患者 ID 番号の桁数が小さいような場合に定数組みを絞り込むことは不可能でないが，匿名化アルゴリズムの

優劣をここで論じたいのではない．具体例を挙げ，この程度の演算処理であれば医療機関の担当者がシステムメーカに頼らずとも，表計算ソフトウェア等を利用してデータ識別番号を編集できることを，特に「桁揃え」の施設に対して示そうとしたものである．なお，これが厚労省DPC関係調査での実施要領とどのように整合するかについてここでは議論しないでおく．

5）調査システムへの提言

厚労省DPC関係調査は今後もその規模を拡げながら継続されていくと思われるが，この調査システムのあり方について2つ提言したい．

第一に，すべての医療機関が安心して調査に参加できるよう，厚労省は実施要領で示すところの「連結可能匿名化」処理をするためのソフトウェアを主体的に配布すべきである．このソフトウェアの機能要件は，複数の提出ファイルに対し，内部のデータ識別番号フィールドの値を一斉に書き換えることができること，匿名化に用いる暗号化鍵を各医療機関が任意に決定できることである．匿名化処理を既に安全に実行している医療機関はこれを使用する必要はないから，配布ソフトウェアを利用するかしないかは各自の判断に委ねれば良い．但し，このような方法で患者ID番号を隠蔽したとしてもこれは本来のデータの匿名化とはいえず，そのデータが他の目的にそのまま利用されている好ましくない状況があることを厚労省は把握した上で，医療機関に対して啓発すべきである．

第二に，MOやCD媒体の書留による送付方式ではなく，技術進歩に見合った安全なプラットフォーム上で調査が効率良く行われるように調査システムを転換していくべきである．国立大学病院データベースセンター※3には旧国立系大学病院におけるDPCデータが集積されているが，そこでは，データファイルを専用ツールにより暗号化した後に，WEBブラウザを通してSSL通信でデータを提出する機能が用意されている．それに倣ったとしても，同程度のセキュリティ水準の下で調査データをオンライン提出するシステムを構築できるはずであり，この調査が開始された2003年当時には可搬媒体を郵送する形態でシステム設計せざるを得ない事情があったにせよ，診療報酬明細書（レセプト）のオンライン請求病院数が95.8%（医科病院計；2010年10月31日時点）[10]になった今，オンラインデータ提出システムに切り替えることについて，関係病院の理解と協力は十分に得られると考える．

6. 結　論

厚労省DPC関係調査において，医療機関がデータ提出時に行うとされている連結可能匿名化処理は実質的に行われていないか，行われていても匿名性の弱いまま運用されているケースの多いことが調査により明らかになった．また，こうした匿名性の十分でないデータがDPCデータ分析サービス提供者へそのまま渡されている実態も見受けられる．

各医療機関は現行システムに在るセキュリティの脆弱性を認識し，調査データの匿名化処理を適正に実行することはもちろん，同データを院外に持ち出す際のリスクアセスメントを行うことが肝要である．

厚労省DPC関係調査システムの今後のあり方として，医療機関が安心して調査データを提出できるように匿名化処理ソフトウェアが配布されること，データ提出形態を郵便システムに依存したシステムからオンライン型システムへと転換していくことを提言する．

謝辞

本稿のアンケート調査にご協力頂いた医療機関各位に感謝申し上げる．

※3大学病院の診療，臨床教育および臨床研究の質の向上に有用な情報を各国立大学病院から収集，解析，還元を行い管理運営の改善，充実に資することを目的として作られた国立大学附属病院長会議が運営する組織体で，2006年より運用が始まっている[11]．

参考文献

1) 厚生労働省．DPC導入の影響評価に関する調査．
http://www.mhlw.go.jp/bunya/iryouhoken/database/sinryo/dpc.html
[accessed 2010-12-12].

2) 酒巻哲夫．臨床データの収集と分析．新版 医療情報「医学・医療編」，篠原出版新社，2009: 435-8.

3) 厚生労働省保険局医療課．「平成22年度新規DPC対象病院説明会」資料．平成22年3月1日．

4) 厚生労働省保険局医療課．平成22年度「DPC導入の影響評価に係る調査」実施説明資料．平成22年7月5日．
http://www.mhlw.go.jp/topics/2010/06/dl/tp0610-1a.pdf
[accessed 2010-12-12].

5) 厚生労働省．臨床研究に関する倫理指針．平成15年7月30日，平成16年12月28日全部改正，平成20年7月31日全部改正．
http://www.mhlw.go.jp/general/seido/kousei/i-kenkyu/rinsyo/dl/shishin.pdf
[accessed 2010-12-12].

6) 厚生労働省．医療・介護関係事業者における個人情報の適切な取扱いのためのガイドライン．平成16年12月24日，平成18年4月21日改正，平成22年9月17日改正．
http://www.mhlw.go.jp/topics/bukyoku/seisaku/kojin/dl/170805-11a.pdf
[accessed 2011-04-15].

7) 個人情報の保護に関する法律（平成一五年五月三十日法律第五十七号）．最終改正：平成二十一年六月五日法律第四十九号．
http://www.caa.go.jp/seikatsu/kojin/houritsu/houritsu.pdf
[accessed 2010-12-12].

8) 厚生労働省．医療情報システムの安全管理に関するガイドライン第4.1版．平成22年2月．
http://www.mhlw.go.jp/shingi/2010/02/dl/s0202-4a.pdf
[accessed 2010-12-12].

9) NPO日本ネットワークセキュリティ協会セキュリティ被害調査ワーキンググループ．2009年情報セキュリティインシデントに関する調査報告書第1.1版．2010年7月1日，2010年9月2日改訂．
http://www.jnsa.org/result/incident/data/2009incident_survey_v1.1.pdf
[accessed 2010-12-12].

10) 社会保険診療報酬支払基金．レセプト電算処理システム普及状況．
http://www.ssk.or.jp/rezept/pdf/hukyu02.pdf
[accessed 2010-12-12].

11) 荘野典文，桶谷文紀，藤江　進，塩崎英司，櫛山博．国立大学附属病院長会議データベース管理委員会，データベースセンターの役割．大学病院情報マネジメント部門連絡会議プログラム・論文集；平成19年度：296-8.