

臨床データ利用における個人非特定化に潜む危険性 - イニシャルは非特定化のために有効か？ -

佐藤美佳¹, 古川裕之^{*1,2}, 松嶋由紀子²,
横井祐子², 宮本謙一^{1,2}

金沢大学大学院自然科学研究科¹
金沢大学医学部附属病院臨床試験管理センター²

Risk of Patient Identification in Use of Clinical Data - Is the Use of Initials Effective for Preserving Identity ? -

Sato Mika¹, Hiroyuki Furukawa^{*1,2}, Yukiko Matsushima²,
Yuko Yokoi² and Kenichi Miyamoto^{1,2}

Kanazawa University Graduate School, Natural Science Research Course¹
Department of Clinical Trial Management Center, Kanazawa University Hospital²

{ Received March 20, 2006 }
{ Accepted December 26, 2006 }

In clinical trials, patient name, date of birth, and sex are items of medical information used to refer to patients. Generally speaking, patient names are indicated by their initials. However, there are doubts as to whether patients may remain completely unidentified when initials are used.

In this study, we investigated the identification levels of patients through the combination of the above three patient data items in post marketing surveillance data. We also investigated the necessity of items recorded in case registration forms for pharmaceutical companies and medical institutions.

We investigated combinations of "K.K." and "Jan. 2, 1947" which were assumed to relate to the largest populations in the database of patient medical information at Kanazawa University Hospital. The total number of patients in the database with combinations of the above was 358,251. The number of patients with the initials "K.K." was 7,892 (1/45), that with the initials "K.K." and "1947" was 173(1/2000), and that with the initials "K.K.", "1947" and "Jan" was 23, while there were only 4 patients (1/8500) with the initials "K.K.", "1947", "Jan" and "2".

These results indicate that it would be possible to identify patients from the combination of their initials and their date of birth (DD/MM/YY). However, the size of the patient population in each medical institution's database would be an important factor since it would determine the extent to which this could be done. From the point of view of protecting personal information, it is thus necessary to consider whether it is appropriate to use patients' initials and dates of birth based on the size of patient population.

Key words identification level, patient initials, size of patient population, patients' private information

緒 言

医療機関が取り扱う情報は機密度やプライバシー度が高いため、その情報が漏れることで患者の社会的な不利益や差別に繋がる可能性を持っている¹⁾。2005年4月1日に『個人情報の保護に関する法律(以下、個人情報保護法とする)』^{1,2)}が施行された。このことにより、患者自身

の個人情報保護に対する意識が高まることが予想される。また、同法が国会で成立する際の付帯決議において、医療、金融・信用、情報通信の3分野は特に個人情報の適正な取り扱いを確保すべき分野と指定された。したがって、医療機関には個人情報保護に対する積極的な取り組みが求められている。

臨床試験を行う際には、医療機関内だけではなく製薬企業など他の施設とも患者の情報を共有しなければなら

* 石川県金沢市宝町13-1; 13-1, Takara-machi, Kanazawa-shi, Ishikawa, 920 8641 Japan

ない。そのため、患者情報を医療機関から外へ持ち出す際には、患者氏名をイニシャル化する等により外部からの容易な個人特定を防止する対策がとられている。

『製薬企業における個人情報の適正な取り扱いのためのガイドライン』³⁾では、「市販後調査(現在は、製造販売後調査)では患者情報は匿名化されたものを入手しているため個人情報には該当しない」との記載がある。しかし、多くの医療機関において一般的に実施されている使用成績調査の際、製薬企業より提供を求められる患者情報の『症例登録票』への記載項目はイニシャル化された情報であっても、複数を組み合わせることで個人が特定される可能性があり、イニシャル化のみによって患者の非特定化が完全に行われているかについては疑問である。

そこで、臨床試験の中でもデータ量が少なく理解しやすい使用成績調査『症例登録票』を一例として取り上げ、個人特定度について調査を行うこととした。

また、同時に現在の『症例登録票』への一般的な記載項目の必要性に関する質問を医師と製薬企業に対して行うことで、不必要な項目の有無について調査することとした。

ここで用いる「個人特定度」とは、金沢大学医学部附属病院(以下、当院と略す)の病院情報システム中の患者情報をデータベース化した「診療支援システム」から患者抽出を行うことで患者が絞り込まれる度合いのことである。すなわち、患者の特定されやすさである。

方 法

1. 患者の個人特定度の調査

2003年と2004年に当院に依頼された使用成績調査の『症例登録票』において必須記載項目となっていた「患者イニシャル」、「生年月日」、「性別」を組み合わせ、診療支援システム(1984年12月から2005年5月までの登録分)より抽出した。最も特定されにくい項目の組み合わせを設定し特定度を示すことで、他のどの組み合わせの設定を行っても特定度がより高くなり、危険性が増すことは理解できる。このことより、特定度の一つの指標として患者数が最も多いと予想される項目の組み合わせを選択することとした。

イニシャルの組み合わせは、単純な計算でも676通り存在する。しかし、イニシャル毎の患者存在比にはバラツキがあると考えられる。そこで、患者数が多いと予想されるイニシャルと少ないと予想されるイニシャルにおける特定度の違いを調査することで、イニシャルの重要性を検証することとした。このため患者数の少ないと予想されるイニシャルをもう一つの設定項目とした。

まず、診療支援システムに登録されている患者について、イニシャルの「姓」と「名」のそれぞれについて人

数分布を調査した。この結果から患者数の多いイニシャルと少ないイニシャルを選択した。生年についても同様に分布を調査し、患者数の多い生年月日を選択した。また、イニシャルと生年月日を組み合わせで抽出を行った。組み合わせは「生年」までの入力のもの、「生年月」までのもの、「生年月日」までのものの3つの場合で行った。さらに、これに「性別」を追加したものについても抽出を行い、それぞれの特定度を調査した。

2. 項目の必要性に関する調査

1) 製薬企業に対する調査

調査期間は2005年6月から7月とし、製薬企業(内資系企業5社、外資系企業6社)を対象に調査用紙(図1)を用いて、これらの記載項目の必要性(製薬企業側からみた必要性、法的規制による必要性⁴⁻⁶⁾)に関する調査を行った。

2) 医師に対する調査

調査期間は2005年6月から7月とし、当院臨床試験管理センターの事前ヒアリング協力医師20名を対象に調査用紙(図2)を用いて、『症例登録票』記載項目(患者イニシャル、性別、使用理由、カルテ番号、生年月日、投与開始日)の必要性に関する調査を行った。

結 果

1. 個人特定度の調査

当院の診療支援システムに登録された延べ患者数は358,251名であった。

当院患者のイニシャルの分布調査によると、「姓」については『K, M, T, S, N』の順に多く、この上位5つで全患者数の約60%を占めていた(図3)。「名」については『K, M, T, Y, S』の順に多く、この5つで約70%を占めていた(図4)。この分布結果から、イニシャルによ

項目名	企業別	項目別に集めたデータ
患者イニシャル		
性別		
生年月日		
カルテ番号		
使用理由		
投与開始日		

調査
①企業にとって本当に必要な項目であるか?
②それらの項目に法規制などはないか? の調査

図1. 製薬企業(内資系企業5社、外資系企業6社)を対象とした「症例登録票」記載項目の必要性に関する調査



図2. 当院臨床試験管理センターの事前ヒアリング協力医師20名を対象とした「症例登録票」記載項目の必要性に関する調査

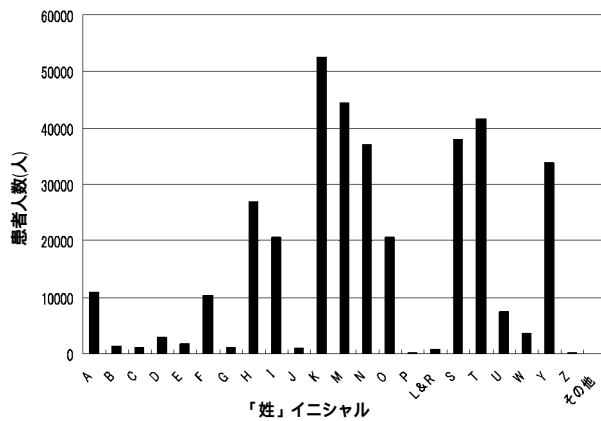


図3. 金沢大学医学部附属病院「診療支援システム」からの抽出による患者「姓」イニシャルの分布

り患者数は大きく異なることが示された。ここで、50,000人以上の患者にイニシャルとして使用されているアルファベットは「姓」で1つ(K),「名」で2つ(K, M)であった(図5)。多くのアルファベットがイニシャルとして使用される頻度が少なく、「姓」で約20%,「名」で約30%は患者数が1,000未満のところにと分類された(図6)。

生年による患者数の分布調査からは生年により患者数が異なることが示され、小児や高齢者では患者数が相対的に少ないために、特定度が高いことがわかった(図7)。また、分布調査結果より1947年1月2日生まれの患者が最も多かった。

この2つの調査より、K・Kを多いイニシャル、Z・Zを少ないイニシャルとして(外国人の氏名を含む)、抽出項目を『K・K, 1947年1月2日, 性別』・『Z・Z, 1947年1月2日, 性別』と設定した。

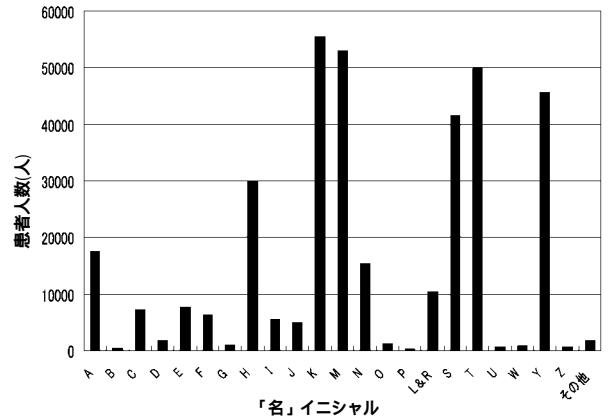


図4. 金沢大学医学部附属病院「診療支援システム」からの抽出による患者「名」イニシャルの分布

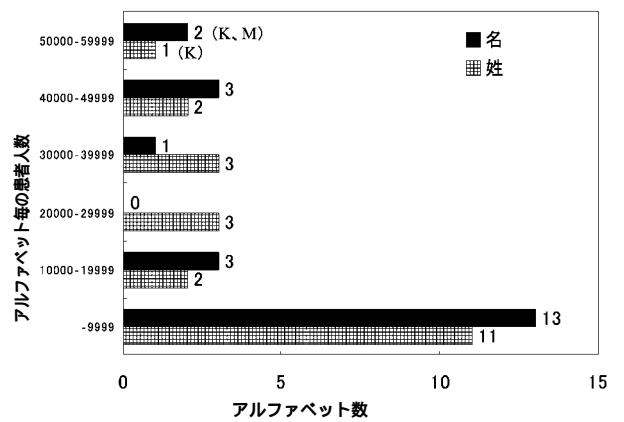


図5. 金沢大学医学部附属病院「診療支援システム」からの抽出による患者数段階別イニシャルの分布(1)

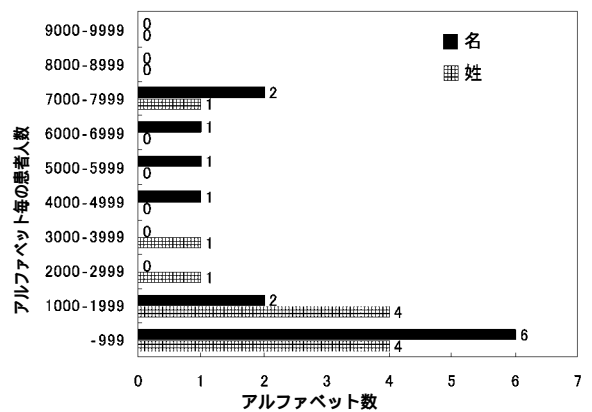


図6. 金沢大学医学部附属病院「診療支援システム」からの抽出による患者数段階別イニシャルの分布(2)

診療支援システムから抽出を行ったところ、登録された延べ患者数は358,251名であった。イニシャルK・Kについてはイニシャル入力のみで7,892名に絞られ、年入力までで173名、月入力までで23名、日まで入力すると3名となった。少ないと予想されるイニシャルZ・Z

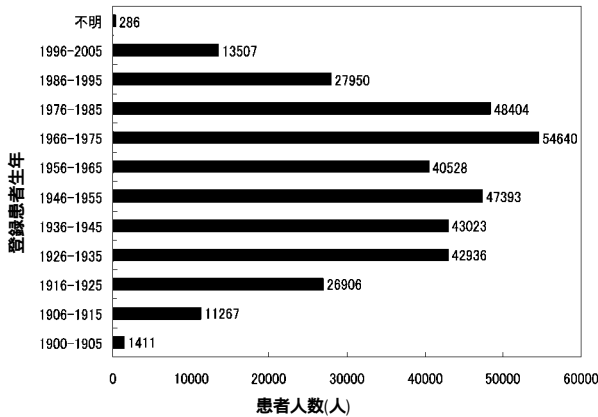


図7. 金沢大学医学部附属病院「診療支援システム」からの抽出による出生年別患者人数の分布

については、イニシャル入力のみでも4名に絞られ、年入力まですると0名となった(表1)。

患者氏名をイニシャル化(匿名化)したもので、生年月日と組み合わせるだけで、個人が特定される可能性があることが示された。

K・KとZ・Zでは患者数に約2,000倍もの差があり、このことから患者数はイニシャルによって大きく異なることがわかった。

生月と生日の入力により抽出した患者数は単純な確率計算で求めた値と近いのに対しイニシャルには多いものと少ないものでバラツキが大きいので、今回の調査からイニシャルの入力により絞られる患者数は確率計算では算出できないことがわかった。したがって、イニシャルにより決定される患者数は個人特定度において重要である。

イニシャルのみ、イニシャルと生年、またイニシャルと生年月の各組み合わせに性別を追加して抽出を行ったところ、患者数はほぼ半数ずつに分かれた(表2)。このことより、性別を追加するだけでは確率計算どおりであり、特定度が高くないことが示された。

2. 項目の必要性に関する調査

1) 製薬企業に対する調査

何も条件を付けなかった場合、製薬企業はほとんどの

表1. 総患者登録数358,251名の「診療支援システム」からの抽出人数

項目	K・K	Z・Z
イニシャルのみ	7892人	4人
イニシャル + 生年	173人	0人
イニシャル + 生年月	23人	0人
イニシャル + 生年月日	3人	0人

表2. 総患者登録数358,251名の「診療支援システム」からの抽出人数(男女別)

	K・K		Z・Z	
	男	女	男	女
イニシャルのみ	4421人	3471人	2人	2人
イニシャル + 生年	86人	87人	0人	0人
イニシャル + 生年月	8人	15人	0人	0人
イニシャル + 生年月日	0人	3人	0人	0人

項目を必要としていた。しかし、製薬企業側でも個人情報保護法の施行を受け、医療機関側から個人情報保護の措置に関する要望があった場合は「カルテ番号・患者イニシャル・生年月日」の記載については、他の方法での対応も可能とした回答が多く認められた(表3)。一方で、使用成績調査における安全性・信頼性の確保など調査の精度を上げるためには、現段階では対応できないという回答もみられた(図8)。この中には、医療機関側での番号管理表等の保管が徹底されていないこと(医師の異動などともなう番号管理表の紛失等)、有害事象発現時などに患者の紐付けが不可能となり、その後の迅速な対応が難しくなるためという理由がみられた。

製薬企業が項目を必要とする理由には、「登録票と調査票の照合をより正確に行うため、ダブルエントリーを防止するため」が多数の項目において認められた。性別と生年月日、投与開始日、使用理由に関しては、「薬剤と安全性・有効性の分析に重要なため」とする回答が多く認められた。患者イニシャルと性別、生年月日、投与開始日、使用理由では「再審査申請資料への記載が求められているため」とする回答がみられた。ただし、再審査申請資料の症例一覧表では、患者略名という記載欄になっており、患者イニシャルは必須記載事項というわけではない⁷⁾。また、生年月日については、年齢でもよいとされていた。患者イニシャルと性別、生年月日は「副作用感染症報告書への記載が必要なため」との回答があった。副作用感染症報告に該当する患者が使用成績調査の対象患者であるという紐付けは必要である。しか

表3. 医療機関からの個人情報保護措置に関する要望への対応方法

項目	代わりとして適切と思われるもの
カルテ番号	・施設内で患者を特定できるような識別番号(例:A-1等)を作成する。(但し、医療機関内での識別番号管理表の保管も必須とする。) ・カルテ番号記載欄は強制的にまで、施設側でカルテ番号と識別番号のどちらかを選択し記載する。
イニシャル	・カルテ番号同様、識別番号等を利用する。
生年月日	・生年もしくは生年月までの記載とする ⁸⁾ 。 ・年齢記載とする。 ※1: ±1歳の誤差が生じるので、その時点での正確な年齢を記載する(試験によっては除外記録は該当する場合や小児・高齢者の場合等の有害事象発現年齢を正確に知るため)

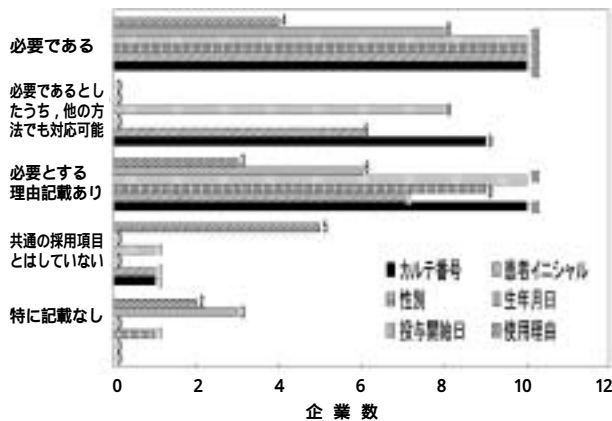


図8. 企業(内資系企業5社, 外資系企業6社)が「症例登録票」記載項目中で必要とした項目

し、副作用感染症報告書への記載項目も法律で規定されているのは「患者略名」・「年齢」であり「患者イニシャル」・「生年月日」までの記載は求められていなかった⁸⁾。また、副作用感染症報告は全医療機関で行うのに対し、使用成績調査は定められた機関で行うものであるため使用成績調査での情報を必ず使用しなければならないという規定はない(表4)。

個人情報とは、個人情報保護法において「生存する個人に関する情報であつて、当該情報に含まれる氏名、生年月日その他の記述等により特定の個人を識別することができるもの(他の情報と容易に照合することができ、それにより特定の個人を識別することができることとなるものを含む。)」と定義されている。個人情報を保護するためには、外部へ出す情報を必要最小限にとどめるということが一つの方法として挙げられる⁹⁾。

一方、使用成績調査の精度を向上させるということは、患者の背景をきちんと把握しその試験に適合している患者を登録し、正確なデータを得ること、また、患者の安全性を確保することである。その確認をするために

表4. 企業(内資系企業5社, 外資系企業6社)が「症例登録票」記載項目を必要とする理由

カルテ番号	登録票と調査票を照合するため ダブルエントリー防止のため 再調査や当局の審査時に該当患者を特定するため
患者イニシャル	カルテ番号と同じ 副作用・感染症報告書記載事項であるため 再調査申請資料記載事項であるため
性別	登録票と調査票を照合するため ダブルエントリー防止のため 再審査申請資料記載事項であるため 副作用・感染症報告書記載事項であるため 薬剤と有効性・安全性の分析に重要なため
生年月日	性別での理由と同じ 有害事象であった場合に発現時年齢の算出等に用いるため「日」まで必要 ダブルエントリー防止のため
投与開始日	登録票と調査票を照合するため 再審査申請資料記載事項であるため 薬剤と有効性・安全性の分析に重要なため 連続性を確認するため
使用理由	ダブルエントリー防止のため 再審査申請資料記載事項であるため 薬剤と有効性・安全性の分析に重要なため

はできる限り情報は多くあるほうがデータの信頼性は高くなる。

このように、個人情報保護と使用成績調査における精度の向上は対極にあるため、製薬企業の考え方によって個人情報保護への対応にも差が認められた。

2) 医師に対する調査

医師は、カルテ番号(89.5%)と患者イニシャル(63.2%)を患者特定のために必要としていた。性別・生年月日は、患者特定と安全性確認のために必要とする回答が、ほぼ半数ずつであった(図9)。

医師にとっては、患者を間違えなく特定することが重要であり、その方法としてはカルテ番号が最も有用であることがわかった。

考 察

今回の調査により、最も患者数が多いと予想されるイニシャルK・Kのイニシャルのみの入力では358,251名から7,892名(約50分の1)に絞られるだけで患者が特定されることはないが、少ないと予想されるイニシャルZ・Zについては4名(約10万分の1)まで絞られることから、患者数は個人特定における重要な因子であることがわかった。当院のような患者数の多い医療機関であっても、イニシャルと生年の組み合わせにより患者数が大きく絞られることから、規模の小さな医療機関ではイニシャルと生年の組み合わせでも患者が特定されてしまう可能性が示唆された。

具体的に規模の小さな医療機関の一例として、患者母数が当院より1桁小さい施設を想定し、当院と同様の絞り込み率を使用して患者数を算出した。イニシャルK・Kのイニシャル入力のみで、35,000名から700名まで絞られ、生年まで入力すると15名(約45分の1)、生年月日まで入力すると1名(約8分の1)となった。

製薬企業は入手した情報だけでは患者を容易に特定す

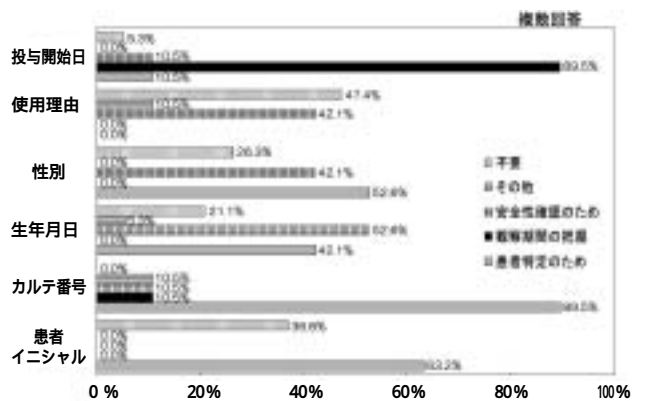


図9. 当院臨床試験管理センター事前ヒアリング協力医師20名の「症例登録票」記載項目中で必要とする項目

ることは不可能であるため、個人情報には該当しないとされている。確かに、その情報だけで個人を特定することは不可能である。

しかし近年、緊急事態を装ったなりすまし行為等による被害がみられる。情報が外部流出した場合には医療機関においても同様のことが発生し、その情報により個人が特定される可能性は否定できない。また、「医療・介護関係事業者における個人情報の適切な取り扱いのためのガイドライン」には、個人情報の漏洩が発生した場合における二次被害防止等についての記載がある¹⁰⁾。このため二次被害防止のための対策を立てることも重要であるといえる。

医師と製薬企業への調査では、使用成績調査対象薬の「使用理由」についてはそれぞれ50%程度が必要とするだけにとどまったが、それ以外の項目に関してはさらに高い割合で必要とされていた。よって、医師と製薬企業の間で必要とする項目の違いは認められなかった。

しかし、製薬企業は独自で「使用成績調査」登録患者を特定する必要はない。一方で、有害事象発現時の対応やその後の追跡調査のための製薬企業からの問い合わせに対応するために医療機関では「使用成績調査」登録患者を特定する必要がある。

そのため医師と製薬企業の間で必要とする項目の違いが認められなくても医療機関としては個人情報保護の観点から、特にカルテ番号は製薬企業には提出したくない項目である。

学会発表時の質疑応答から識別番号に関する実際の医療機関の取り組みとして、医師が自身の患者の一覧を作成し管理している、臨床試験管理センター等で責任医師一覧表を作成し管理している、調査に関しては法律上、同意取得の規定はないが、あらかじめ識別番号を割り当てた同意書による同意取得を行うことにより、同意書の管理と同時に識別番号を管理している、という方法が取られていることがわかった。

治験等の実施数が多く、情報の管理に慣れている施設では識別番号の管理についても同様に実施可能であると考えられるが、治験等の実施経験がない施設では十分に行われない可能性があるという製薬企業からの意見がみうけられた。

したがって、医療機関側は個人情報保護に関して、その規模など状況に応じて製薬企業と話し合った上で個別的な対応をとる必要がある。具体例としては、製薬企業に『症例登録票』を提出する際にカルテ番号やイニシャルをマスキングして、「施設略号」と「番号」(K 001)のように別に識別子をつくる等の対応が挙げられる。また、医療機関側でも「症例報告書」の保管方法の検討を行うことにより、製薬企業が懸念している有害事象が発生した際の素早い対処を可能とすることで使用成績調

査対象薬の使用に関する安全性を保つ必要がある。

医薬品は疾病治療のため患者にとっては必要不可欠なものである。しかし、全例調査において、個人情報保護への対応が極端すぎる場合は、市販後調査の全例調査対象の医薬品採用を控えることとなる。例えば製薬企業から患者IDの記載の要求を医療機関側が拒否したことによる患者の使用成績調査への組み入れの断念等が挙げられる。その結果、その医薬品を本当に必要としている患者の治療を受ける機会が奪われることとなり、利益を損なう可能性がある。逆に、医薬品が手に入り治療を受けられなかったとしても、個人情報が守られず社会的な不利益や差別を受けることがあれば、患者にとっては不幸なことである。

医療機関においては、個人情報保護の必要性和治療上の利益のバランスを考慮して、対応を行う必要がある。

謝辞 今回、調査にご協力いただいた企業の皆様、当院の医師の皆様、学会等でご意見をお寄せいただいた皆様に心より感謝いたします。

引用文献

- 1) 橋本佳子, 山崎大作, 井上俊明; 医療機関のための個人情報保護法対応マニュアル” 初版, 日経メディカル編, 日経 BP 社, 東京, 2005, pp.10 17.
- 2) 橋本佳子, 山崎大作, 井上俊明; 医療機関のための個人情報保護法対応マニュアル” 初版, 日経メディカル編, 日経 BP 社, 東京, 2005, pp.120 131.
- 3) “製薬企業における個人情報の適正な取り扱いのためのガイドラインについて”, 日本製薬団体連合会, 東京, 2005, p.3.
- 4) “医薬品の製造販売後の調査および試験の実施の基準に関する省令” 厚生労働省令 第171号, 2004年12月20日.
- 5) “医薬品の製造販売後の調査および試験の実施の基準に関する省令の施行について”, 薬食発 第1220008号, 2004年12月20日.
- 6) “医療用医薬品の市販直後調査等の実施方法に関するガイドラインについて”, 医薬安 第166号 医薬審第1810号, 2000年12月27日.
- 7) “新医療用医薬品の再審査申請に際し添付すべき資料について”, 薬食審査発 第1027004号, 2005年10月27日.
- 8) “薬事法施行規則の一部を改正する省令等の施行について”, 薬食発 第0317006号, 2005年3月17日.
- 9) 橋本佳子, 山崎大作, 井上俊明; 医療機関のための個人情報保護法対応マニュアル” 初版, 日経メディカル編, 日経 BP 社, 東京, 2005, p.118.
- 10) 橋本佳子, 山崎大作, 井上俊明; 医療機関のための個人情報保護法対応マニュアル” 初版, 日経メディカル編, 日経 BP 社, 東京, 2005, pp.147 151.