

1) 主任研究者

竹内正弘

北里大学薬学部臨床薬学研究センター臨床統計部門教授

連絡先：〒108-8641 東京都港区白金5丁目9番1号

TEL: 03-5791-6322

FAX: 03-3444-2546

E-mail:takeuchim@pharm.kitasato-u.ac.jp

2) 臨床研究実施施設

北里大学病院

国立国際医療センター

国立循環器病センター

国立病院機構東京医療センター

徳島大学病院

広島大学病院

3) 分担研究者

北里大学病院 矢後和夫

国立国際医療センター 吉野信次

国立循環器センター 森下秀樹

国立病院機構東京医療センター 青木 誠

徳島大学病院 水口和生

徳島大学病院 森口博基

広島大学病院 木平健治

4) 研究協力者

北里大学病院 木村利美

国立国際医療センター 斉藤真一郎

国立循環器病センター 高田充隆 小竹 武

国立病院機構東京医療センター 軍司剛宏

徳島大学病院 森川富昭

広島大学病院 新井茂昭

京都薬科大学 柴川雅彦

国立病院機構京都医療センター 北岡有喜

医薬品医療機器総合機構安全部 小島千枝

北里大学 高橋史朗 Mindy Fang

北里研究所 青谷恵利子

5) 研究オブザーバー

日本製薬工業協会

米国研究製薬工業協会

6) 医学専門家

医薬品医療機器総合機構 顧問 内藤周幸

7) テクニカルアドバイザー

日立製作所 村上憲之 小島淳一

8) 臨床研究事務局

南由紀子、小高康世

北里大学薬学部臨床薬学研究センター臨床統計部門

連絡先：〒108-8641 東京都港区白金5丁目9番1号

TEL: 03-5791-6322

FAX: 03-3444-2546

E-mail: kodakay@pharm.kitasato-u.ac.jp; minamiy@pharm.kitasato-u.ac.jp

1.1. 参考文献

1. Kim BS, Choi HY, Ryu DR, Yoo TH, Park HC, Kang SW, Choi KH, Ha SK, Han DS, Lee HY: Clinical characteristics of dialysis related sclerosing encapsulating peritonitis: multi-center experience in Korea. *Yonsei Med J.* 2005 Feb 28; 46(1):104-111.
2. Tulunay FC, Ergun H, Gulmez SE, Ozbenli T, Ozmenoglu M, Boz C, Erdemoglu AK, Varlikbas A, Goksan B, Inan L: The efficacy and safety of dipyrone (Novalgin) tablets in the treatment of acute migraine attacks: a double-blind, cross-over, randomized, placebo-controlled, multi-center study. *Funct Neurol.* 2004 Jul-Sep; 19(3):197-202.
3. Park SR, Oh DY, Kim DW, Kim TY, Heo DS, Bang YJ, Kim NK, Kang WK, Kim HT, Im SA, Suh JH, Kim HK, Kim HK: A multi-center, late phase II clinical trial of Genexol (paclitaxel) and cisplatin for patients with advanced gastric cancer. *Oncol Rep.* 2004 Nov; 12(5):1059-64.
4. Zhang SW, Wang C, Yin CH, Wang H, Wang BE: [Multi-center clinical study on the diagnostic criteria for multiple organ dysfunction syndrome with illness severity score system] *Zhongguo Wei Zhong Bing Ji Jiu Yi Xue.* 2004 Jun; 16(6):328-332....Chinese.
5. Sadick NS, Weiss R, Kilmer S, Bitter P: Photorejuvenation with intense pulsed light: results of a multi-center study. *J Drugs Dermatol.* 2004 Jan-Feb; 3(1):41-49.
6. Hening WA, Allen RP, Thanner S, Washburn T, Heckler D, Walters AS, Earley CJ: The Johns Hopkins telephone diagnostic interview for the restless legs syndrome: preliminary investigation for validation in a multi-center patient and control population. *Sleep Med.* 2003 Mar; 4(2):137-141.
7. Prummel MF, Bakker A, Wiersinga WM, Baldeschi L, Mourits MP, Kendall-Taylor P, Perros P, Neoh C, Dickinson AJ, Lazarus JH, Lane CM, Heufelder AE, Kahaly GJ, Pitz S, Orgiazzi J, Hullo A, Pinchera A, Marcocci C, Sartini MS, Rocchi R, Nardi M, Krassas GE, Halkias A: Multi-center study on the characteristics and treatment strategies of patients with Graves' orbitopathy: the first European Group on Graves' Orbitopathy experience. *Eur J Endocrinol.* 2003 May; 148(5):491-495.

8. Cabaniss DL, Schein JW, Rosen P, Roose SP: Candidate progression in analytic institutes: a multi-center study. *Int J Psychoanal.* 2003 Feb; 84(Pt 1):77-94. Erratum in: *Int J Psychoanal.* 2003 Apr; 84(Pt 2):451.
9. Larson VD, Williams DW, Henderson WG, Luethke LE, Beck LB, Noffsinger D, Bratt GW, Dobie RA, Fausti SA, Haskell GB, Rappaport BZ, Shanks JE, Wilson RH: A multi-center, double blind clinical trial comparing benefit from three commonly used hearing aid circuits. *Ear Hear.* 2002 Aug; 23(4):269-276.
10. Leenhouts GH, Kylstra WA, Everaerd W, Hahn DE, Schultz WC, van de Wiel HB, Heintz AP: Sexual outcomes following treatment for early-stage gynecological cancer: a prospective and cross-sectional multi-center study. *J Psychosom Obstet Gynaecol.* 2002 Jun; 23(2):123-132.
11. Applebaum-Shapiro SE, Finch R, Pfutzer RH, Hepp LA, Gates L, Amann S, Martin S, Ulrich CD 2nd, Whitcomb DC: Hereditary pancreatitis in North America: the Pittsburgh-Midwest Multi-Center Pancreatic Study Group Study. *Pancreatology.* 2001; 1(5):439-443.
12. Moine P, Timsit JF, De Lassence A, Troche G, Fosse JP, Alberti C, Cohen Y; OUTCOMEREA study group: Mortality associated with late-onset pneumonia in the intensive care unit: results of a multi-center cohort study. *Intensive Care Med.* 2002 Feb; 28(2):154-163. Epub 2002 Jan 16.
13. Covitz PA, Hartel F, Schaefer C, De Coronado S, Fragoso G, Sahni H, Gustafson S, Buetow KH: caCORE: a common infrastructure for cancer informatics. *Bioinformatics.* 2003 Dec. 12; 19(18):2404-2412.
14. Poulymenopoulou M, Vassilacopoulos G: An electronic patient record implementation using clinical document architecture. *Stud Health Technol Inform.* 2004; 103:50-57.
15. Boaz D, Shahar Y: A framework for distributed mediation of temporal-abstraction queries to clinical databases. *Artif Intell Med.* 2005 May; 34(1):3-24.
16. Blaser MJ, Saito D: Trends in reported adenocarcinomas of the oesophagus and gastric cardia in Japan. *Eur J Gastroenterol Hepatol.* 2002 Feb; 14(2):107-113.
17. Forman D: Counting cancers at the junction - a problem of routine statistics. *Eur J Gastroenterol Hepatol.* 2002 Feb; 14(2):99-101.
18. Strobl J, Enzer I, Bagust A, Haycox A, Smyth R, Ashby D, Walley T: Using disease registries for pharmacoepidemiological research: a case study of data from a cystic fibrosis registry. *Pharmacoepidemiol Drug Saf.* 2003 Sep; 12(6):467-473.
19. Ferretti S, Federico M, Contiero P, Stracci F: Comparability issues of survival data. *Epidemiol Prev.* 2001; 25(3 Suppl):37-41....English....Italian.
20. Brandt CA, Sun K, Charpentier P, Nadkarni PM: Integration of Web-based and PC-based clinical research databases. *Methods Inf Med.* 2004; 43(3):287-295.
21. Tatsumi N, Tsuda I, Kondoh H: [Standardization in hematology and the use of the data in clinical laboratories] *Rinsho Byori.* 1993 May; 41(5):592-595....Japanese.
22. Berman JJ: Nomenclature-based data retrieval without prior annotation: facilitating biomedical data integration with fast doublet matching. *In Silico Biol.* 2005 Apr 3; 5(3):0029 [Epub ahead of print]
23. Removal of Vioxx Shifts Drug Landscape, *New York Times*, 2004 Oct. 1
24. MERCK AND VIOXX: THE COMPANY; A Blow to Efforts To Close In on Rivals, *New York Times*, 2004 Oct. 1
25. Merck Announces Voluntary Worldwide Withdrawal of VIOXX, *Vioxx.com*, 2004 Sep. 30
26. First Vioxx trial begins in Texas 'Merck-y ethics' cited in opening statement, *The Boston Globe*, 2005 Jul.

27. The COX-2 Story and Cancer, *OncoLink*, 2005 Jan. 2
28. Morris NC: Parametric Empirical Bayes Inference: Theory and Applications. *J American Statistical Association*. 1983 Mar; 78(381):47-55
29. Efron B: Empirical Bayes Methods for Combining Likelihoods. *J American Statistical Association*. 1996 Jun; 91(434):538-550
30. Szarfman A et. al.: Use of Screening Algorithms and Computer Systems to Efficiently Signal Higher-Than Expected Combinations of Drugs and Events in the US FDA's Spontaneous Reports Database. *Drug Safety*. 2002; 25(6):381-392
31. Hauben M et. al.: Quantitative Methods in Pharmacovigilance Focus on Signal Detection. *Drug Safety*. 2003; 26(3):159-186
32. Purcell P, Barty S: Statistical Techniques for Signal Generation The Australian Experience. *Drug Safety*. 2002; 25(6):415-421
33. Liang KY, Zeger SL: Longitudinal data analysis using generalized linear models. *Biometrika* 1986; 73:13-22. 12
34. Laird, N. and Ware, J: Random-effects models for longitudinal data. *Biometrics* 38, 1982; 963 -974

1 2. 資料

- 資料 1. 「疫学研究に関する倫理指針」(文部科学省、厚生労働省、平成14年6月17日 [平成16年12月28日全部改正] [平成17年6月29日一部改正])
- 資料 2. 「臨床研究に関する倫理指針」(厚生労働省、医政発1228002号、平成16年12月28日)
- 資料 3. 「個人情報の保護に関する法律」(平成15年5月30日法律第57号、平成17年4月1日全面施行)

以上

訂正履歴（2005年7月13日以降）

訂正箇所	訂正日	訂正前	訂正後	理由	提案者
表紙 および 4 7-7-3 10	2005/8/26		北里大学薬学部臨床統計部門 北里大学薬学部臨床薬学研究センター臨床統計部門	正式名称採用	北里大学
全体	2005/8/26	各施設	6施設	4における定義採用	北里大学
目次	2005/8/10	6. 研究対象 7. 研究方法 8. 統計解析 9. 倫理的配慮およびプライバシーの保護	6. 研究の対象 7. 研究の方法 8. 統計学的解析 9. 倫理的配慮および個人情報の保護		北里研究所
1-2 および 5	2005/8/10	非比較、記述的、後ろ向きコホート研究 (Non-comparative, Descriptive, Retrospective Cohort Study)	後ろ向きコホート研究 (Retrospective Cohort Study)	後ろ向きデータ収集と分析を伴うため、「後ろ向きコホート研究」に変更。	徳島大学病院
2	2005/8/21	Pro Active な安全監視体制へ変換しようとしている。	Pro Active な安全監視体制への変換が望まれている。		医薬品医療機器総合機構
2	2005/8/21	“因果関係がありそうな医薬品と副作用の組み合わせ（シグナル）”を検出できることが重要である。	“因果関係がありそうな医薬品と副作用の組み合わせ（シグナル）”を検出できれば安全対策上有用である。		医薬品医療機器総合機構
4	2005/8/10		(追記) また、予め選択したモデル薬剤 (HMG-CoA リダクターゼ) に関して複数施設より収集・統合された電子化医療情報より、当該薬剤の有効性と安全性について評価が可能であるか否かを検討する。	疫学的研究についての記載を追記	徳島大学病院
7-2	2005/7/13	施設毎に、その施設の委員会の承認を得て	施設毎に、該当する委員会（倫理委員会等）の承認を得て		国立循環器病センター
7-3-2	2005/8/10	電子化情報から薬剤の有効性と安全性の評価を比較的考察可能という理由から HMG-CoA リダクターゼ阻害剤を選択する。	検査値などの既存の電子化情報を用いると薬剤の有効性と安全性を評価しやすいという理由から HMG-CoA リダクターゼ阻害剤を選択した。		徳島大学病院
7-4	2005/8/10	スタチン系薬剤の患者	スタチン系薬剤の処方を受けた患者	記述の具体化	北里研究所
7-5	2005/8/10	2004年9月30日に投与した患者の情報を6ヶ月間収集する。	2004年9月30日までにスタチン系薬剤の処方を受けた患者の情報を6ヶ月間収集する。	記述の具体化	北里研究所

7-5	2005/8/10	第一段階として、各施設で電子化情報を、可能な形式で収集し、患者をキーとして連結する。	第一段階として、各施設内に分散した本研究に必要な電子化情報を同一患者のデータであることが識別できる患者情報(患者名あるいは患者コード等)で連結する。	表記を具体的にする必要から	徳島大学病院
7-5	2005/7/13	処方内容は、投与日と投与量を推定して投与情報とする	電子カルテは、投与日と投与量を投与情報とし、オーダーリングの場合処方日、処方量から投与日と投与量を推定して投与情報とする		国立循環器病センター
7-5	2005/7/26	第二段階として、施設内の患者情報を、事前に施設内で患者特定が行えないよう匿名化(患者番号の暗号化等)、あるいは患者へ連結不可能な処理を行い、主任研究者へ渡す。匿名化等された多施設の集積情報は、主任研究者により管理される。	第二段階として、施設内の患者情報を、患者特定が行えないよう事前に匿名化(患者番号の暗号化等)、あるいは患者へ連結不可能な処理を施設内で行い、主任研究者へ渡す。匿名化等された多施設の集積情報及び統合された情報は、主任研究者により管理される。また、患者を特定できないように匿名化等したコードとオリジナルのコード情報は、各施設の分担研究者により管理される。		国立国際医療センター
7-5	2005/8/10	第二段階として、施設内の患者情報を、事前に施設内で患者特定が行えないよう匿名化(患者番号の暗号化等)、あるいは患者へ連結不可能な処理を行い、主任研究者へ渡す。匿名化等された多施設の集積情報は、主任研究者により管理される。	「第二段階として、施設内の患者情報を、事前に施設内で患者特定が行えないよう匿名化する。但し、施設内においては患者への連結は可能な処理を行い、連結のための対応情報は、各施設の分担研究者が施設内で厳重に管理する。」・・・「主任研究者に渡された匿名化等された多施設の集積情報は、主任研究者が任命した管理者により厳重に保管される。その保管期間は、本研究が終了するまでとする。」	患者の匿名化には暗号化は必ずしも必要ではない。また、データ保管に関しては詳細な記述が必要。	徳島大学病院
7-5	2005/8/10	施設間での統合にはXLM形式を利用する。	施設間での統合には、新規に主任研究者側で定義したXML形式を利用する。	統合したデータ解析のため、新規に主任研究者側で定義。	徳島大学病院、北里研究所
7-5	2005/8/21	但し、施設内においては患者への連結は可能な処理を行い、連結のための対応情報は、各施設の分担研究者が施設内で厳重に管理する。	但し、施設内においては患者への連結は可能な処理を行い、連結のための対応情報は、各施設の分担研究者又は施設により指名された管理者が施設内で厳重に管理する。		医薬品医療機器総合機構
7-6	2005/8/10	4. 生年月	4. 年齢(生年月)	各施設で年齢計算の対応が可能な場合は年齢を、対応できない場合は生年月まで収集。	徳島大学病院、北里研究所

7-7-2	2005/8/10	(1) 提供される CSV 形式などのファイルを変換する際に、CSV などデータ型が存在しないファイル形式では変換時にデータ型やサイズなど設定する。	(1) 提供される CSV 形式などのファイルを変換する際に、CSV などデータ型が存在しないファイル形式では変換時にデータ型やサイズなどを設定する必要がある。そこで、CSV 形式のファイルに関する添付資料として提出された各カラムの属性情報を参照して設定を行う。	各施設からは、各カラムの属性情報を提供してもらい、XML 形式化は主任研究者側で行う。	徳島大学病院、北里研究所
9	2005/8/10		各機関の意見を調整して全面改訂(医薬品医療機器総合機構)		徳島大学病院、北里研究所、国立国際医療センター、医薬品医療機器総合機構
9	2005/8/21	本研究は、既存資料のみを用いる観察研究であることから、本研究の目的を含む研究の実施についての情報を各実施施設において公開することとする。また、本研究は、他の機関に匿名化された既存資料を提供するため、施設の倫理委員会の承認を得て、研究を行う機関の長の許可を受けることとする。	本研究は、当該施設で収集した既存資料を匿名化して他の機関に提供するため、疫学研究に関する倫理指針第4の11の規程により、施設の倫理委員会の承認を得て、研究を行う機関の長の許可を受けることとする。なお、既存資料を利用するため、疫学研究に関する倫理指針第3の7の規程により、インフォームドコンセントの取得は簡略化する。簡略化の方法は各施設の指針によることとする。		医薬品医療機器総合機構
9	2005/8/26		(9-1として追加) 9-1 北里大学薬学部の倫理委員会 北里大学薬学部の倫理委員会に諮り承認を得る予定である。		北里大学

(資料 3)

Analysis of Total Cholesterol Data Random Mixed Effect Model

Abstract

A retrospective cohort study was conducted to investigate whether clinical data, collected through daily medical practice at each medical institute, can be merged based on each patient in terms of a patient background, the results of medical tests (a blood test etc), and observed adverse events. Once these data can be merged on the basis of a patient, the data can be merged among participated medical institutes. Then statistical analyses can be conducted to investigate how efficacy and adverse events can be controlled at daily patient life. We focus on the total cholesterol level and investigate how this variable can be controlled by prescribed drug over a time and investigate what risk factors can influence the total cholesterol level by applying random mixed effect model.

Primary outcome variable: TCHO total cholesterol level (measured at multiple visits)

Covariates: GENDER, HEIGHT, WEIGHT, AGE (recorded at baseline)

Random effect: TIME (time of the visits)

GENDER and AGE turned out to be significant fixed effects for total cholesterol level, and time of visit has a significant longitudinal effect for total cholesterol level.

Summary Statistics

The raw data set contains 3087 subjects. 220 subjects are missing the primary outcome variable “total cholesterol level” thus excluded from the analysis. The analysis set used in this report contains 2867 subjects, each of them has at least one visit to have their cholesterol level measured. Most of the subjects had less than 30 visits during the entire study period (Figure 1), however the number of visit can be as many as 60 times (one subject).

Histogram of the Number of Visit

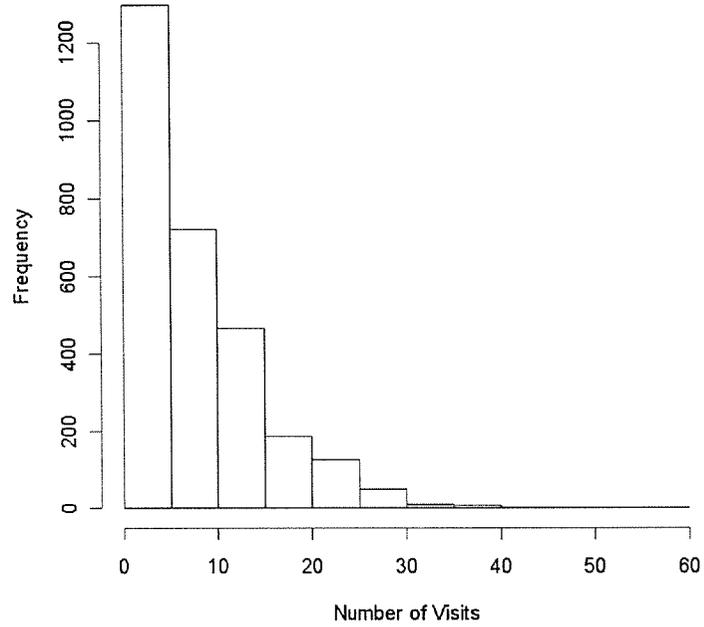


Figure 1

The duration of the study period varies for each subject as well. For the subjects who had only one visit at day 0 the duration of the study is considered to be 0 days. The longest follow-up duration is 866 days (Figure 2).

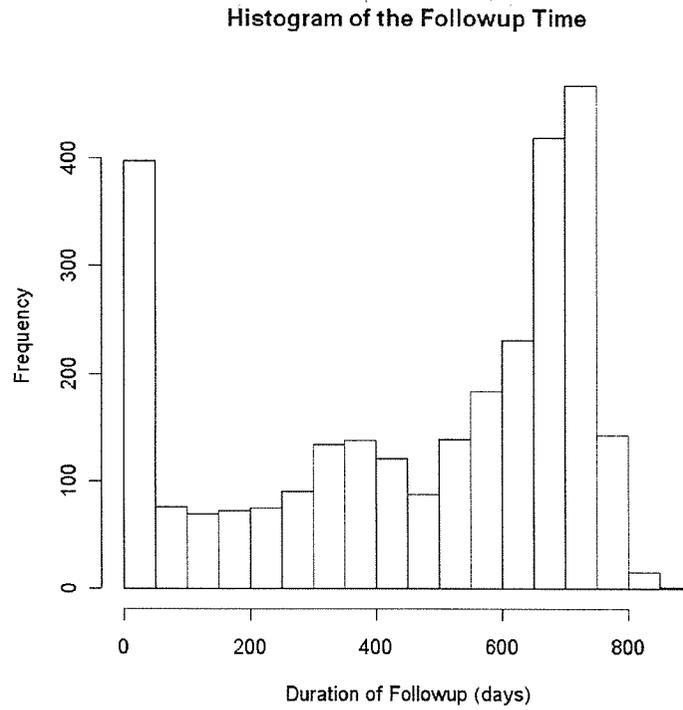


Figure 2

The primary outcome variable “total cholesterol level” is assumed to be normally distributed, which can be seen from both Figure 3 and Figure 4. Figure 3 shows the histogram plot of all the measured total cholesterol levels, allowing for multiple measurements for each subject; Figure 4 shows the histogram plot of the total cholesterol levels averaged within each subjects, allowing only one entry for each subject.

Histogram of the Total Cholesterol Level

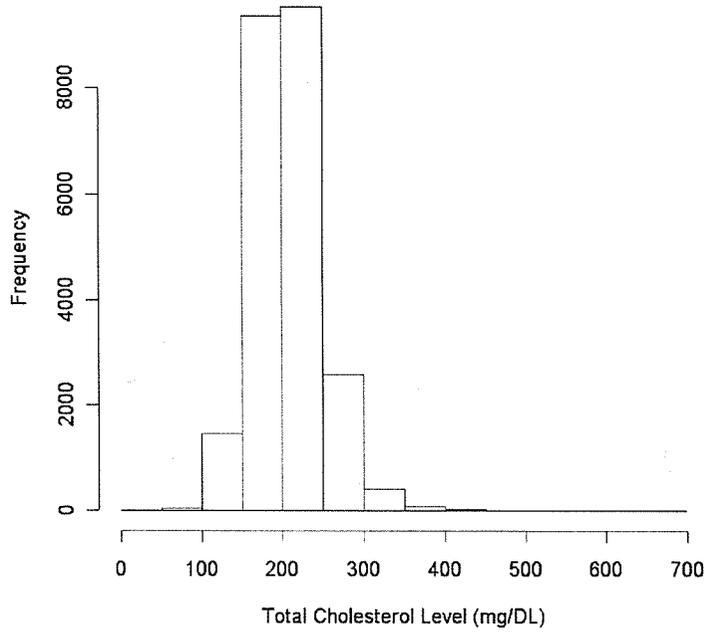


Figure 3

Histogram of the Total Cholesterol Level for Averaged by Subjec

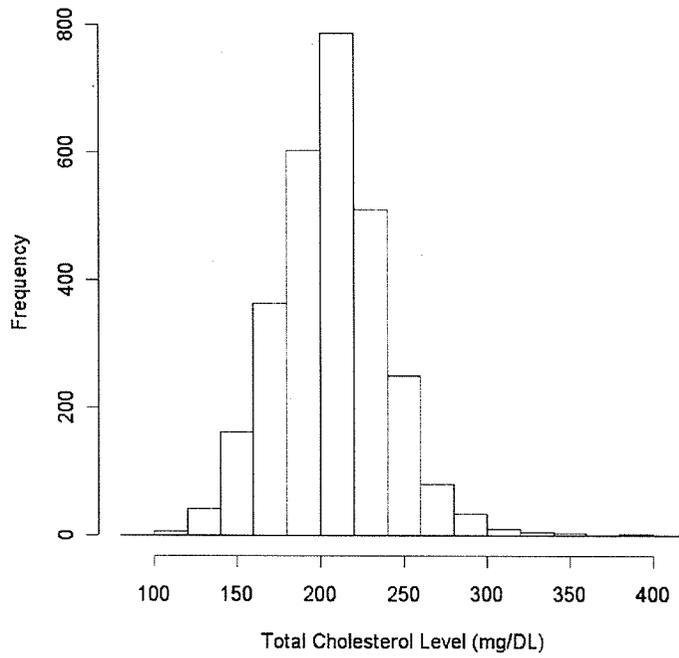


Figure 4

Scatter plots are made to visualize the relationship between the covariates and the outcome variable. Figure 5 shows the pair-wise scatter plots of the outcome variable TCHO with three of the continuous covariates HEIGHT, WEIGHT, and AGE. There is apparent positive correlation between HEIGHT and WEIGHT, negative correlation between AGE and HEIGHT, negative correlation between AGE and WEIGHT. Correlation between the covariates and the outcome variable is unclear from the scatter plot.

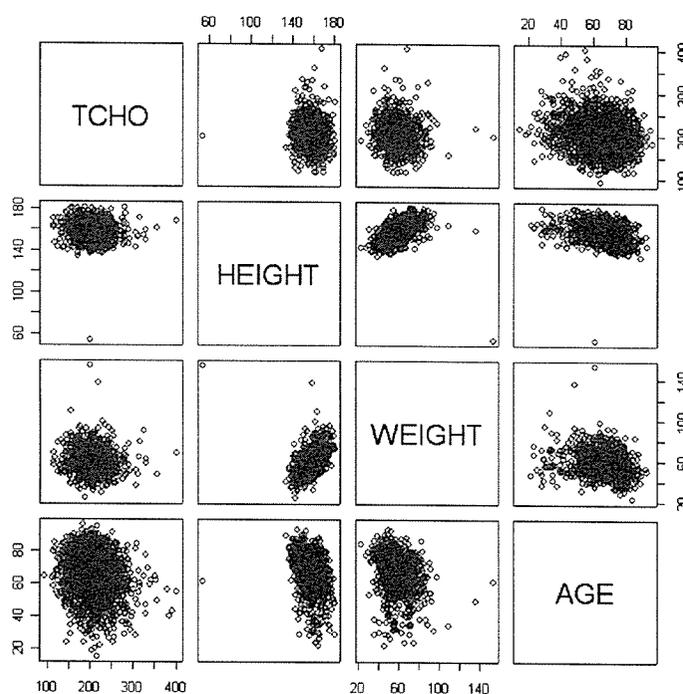


Figure 5

Linear Mixed Effect Model with Random Time Effect

As can be seen from Figure 1 and Figure 2, the study data is highly unbalanced over time. Each subject has various number of visit, and they can visit at different times as well. There is between-subject variation because the subjects are measured at different times, also there could be “longitudinal” or within-subject information that arises because subjects are measured repeatedly over time, yielding measurements of TCHO at different times.

The goal of the following analysis is to determine how total cholesterol level (TCHO) over time is related to the gender, height, weight, and age of the subject. Because the

data are inherently unbalanced, accounting for the covariance among the repeated observation on the same subject via a random effects structure is necessary. Here we allow the intercept and slope for TIME to varying randomly from one subject to another. Specifically, we consider the following model for TCHO:

$$E(TCHO_{ij} | b_i) = \beta_1 + \beta_2 GENDER_i + \beta_3 WEIGHT_i + \beta_4 HEIGHT_i + \beta_5 AGE_i + \beta_6 TIME_{ij} + b_{1i} + b_{2i} TIME_{ij}$$

where $TCHO_{ij}$ is the total cholesterol level for the i^{th} subject at the j^{th} visit, and $GENDER_i$, $WEIGHT_i$, $HEIGHT_i$, AGE_i are the initial or baseline information for the i^{th} subject. In this model, β_6 is the longitudinal effect of TIME.

The REML estimates of the fixed effects are displayed in Table 1. GENDER, AGE and TIME turn out to have significant effects on the total cholesterol level, while HEIGHT and WEIGHT are not significant. Remove HEIGHT and WEIGHT from the model and refit the mixed effect model, and the REML estimates of the fixed effects are displayed in Table 2.

$$E(TCHO_{ij} | b_i) = \beta_1 + \beta_2 GENDER_i + \beta_5 AGE_i + \beta_6 TIME_{ij} + b_{1i} + b_{2i} TIME_{ij}$$

Specifically, female subjects have higher total cholesterol levels than male subjects of the same age, and the difference is about 15.54 mg/DL on average. Meanwhile every 10 years increase in age results in a decrease in total cholesterol of 5.7 mg/DL on average. TIME has a significant longitudinal effect on the total cholesterol level as well, the level decreases by 1.67 mg/DL on average every 100 days.

Table 1

Solution for Fixed Effects

Effect	GENDER	Estimate	St Err	DF	t Value	Pr > t
Intercept		243.80	31.4748	857	7.75	<.0001
GENDER	F	15.7359	3.2574	7762	4.83	<.0001
GENDER	M	0
WEIGHT		-0.1996	0.1040	7762	-1.92	0.0551

HEIGHT	0.07488	0.1736	7762	0.43	0.6662
AGE	-0.6464	0.1043	7762	-6.20	<.0001
TIME	-0.02298	0.002890	799	-7.95	<.0001

Table 2

Solution for Fixed Effects

Effect	GENDER	Estimate	St Err	DF	t Value	Pr > t
Intercept		239.57	3.3576	2862	71.35	<.0001
GENDER	F	15.5398	1.1905	18E3	13.05	<.0001
GENDER	M	0
AGE		-0.5734	0.05057	18E3	-11.34	<.0001
TIME		-0.01669	0.001406	2553	-11.87	<.0001

The estimates of the variance and covariance of the random effects from the reduced model are displayed in Table 3. When these estimates are compared to their standard errors, there is evidence to support their retention in the model.

Table 3

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	St Err	Z	Pr>Z
Var(b_{1i})	ID	1032.58	34.8457	29.6	<.0001
Cov(b_{1i}, b_{2i})	ID	-0.7816	0.05618	-13.9	<.0001
Var(b_{2i})	ID	0.002358	0.000138	17.1	<.0001
Residual		724.08	7.4816	96.7	<.0001

The estimated variance and covariance matrix for the first subject (ID 9000001, 8 visits) is shown in Table 4. The result indicates that there is strong positive correlation among measurements of total cholesterol level, and the correlation declines by a modest amount over the visit times but does not decay to 0.

Table 4

Estimated V Matrix for ID 9000001

Row	Col1	Col2	Col3	Col4	Col5	Col6	Col7	Col8
1	1756.66	988.81	901.27	879.38	841.08	748.07	698.82	627.70
2	988.81	1676.51	879.68	861.49	829.67	752.37	711.45	652.34
3	901.27	879.68	1560.59	825.72	806.84	760.97	736.69	701.60
4	879.38	861.49	825.72	1540.86	801.13	763.12	743.00	713.94
5	841.08	829.67	806.84	801.13	1515.22	766.89	754.05	735.50
6	748.07	752.37	760.97	763.12	766.89	1500.11	780.87	787.86
7	698.82	711.45	736.69	743.00	754.05	780.87	1519.15	815.58
8	627.70	652.34	701.62	713.94	735.50	787.86	815.58	1579.7

Conclusion and Discussion

A linear mixed effect model is deployed to fit the total cholesterol level data with time of the visit as random effect. Gender and age turn out to be significant fixed effect factors to have an impact on the total cholesterol level, in that female subjects have higher total cholesterol levels than male subjects of the same age, and the difference is about 15.54 mg/DL on average. Meanwhile every 10 years increase in age results in a decrease in total cholesterol of 5.7 mg/DL on average. Time of visit turns out to have a significant longitudinal effect on the total cholesterol level as well, the level decreases by 1.67 mg/DL on average every 100 days.

The analysis used all the available data. Subjects with even a single observation are included in the analysis.

References

- ✧ Fitzmaurice G, Laird N, Ware J. Applied Longitudinal Analysis. 2004; Wiley Series in Probability and Statistics.
- ✧ Brown H, Prescott R. Applied Mixed Models in Medicine. 1999; Wiley Series.
- ✧ Laird N, Ware J. Random-effects Models for Longitudinal Data. Biometrics 1982; 38:963.

- ◇ Cnaan A, Laird N, Slasor P. Using the General Linear Mixed Model to Analyze Unbalanced Repeated Measures and Longitudinal Data. *Statistics in Medicine* 1997; 16:2349.

Correspondence

Mindy Fang

Research Associate

Division of Biostatistics

School of Pharmaceutical Sciences

Kitasato University

5-9-1 Shirokane Minato-ku Tokyo #108-8641 Japan

phone: +81-3-5791-6322 fax: +81-3-3444-2546

(資料4-1)

厚生労働科学研究費補助金（医薬品等医療技術リスク評価研究事業）
「医薬品等の市販後における有効性、安全性の評価方法に関する研究」
研究協力報告書

『HMG-CoA リダクターゼ阻害剤の使用成績調査をモデルに取り上げ、電子化された患者情報を複数の医療機関から収集する方法に関する研究』の
データ統合作業実施における技術的課題と検討に関する報告

研究協力者 村上憲之 株式会社日立製作所医薬システム統括部 主任技師/上級コンサルタント

研究要旨：

本研究は、初年度病院薬剤師会を通して実施した日本の各医療機関における電子化情報の実態調査をもとに、次年度は医療機関における医薬品の適正な使用実態を把握するために、簡便かつ迅速にデータを収集するためのITツールを開発することを検証するために倫理的な側面を十分に議論してプロトコル（実施計画書）を作成し、それと並行してシステム化の方式検討のため机上検討を行った。

最終年度は、臨床試験や市販後調査のデータマネージメントシステムを構築してきた経験を活かし、初年度からの研究の集大成として、机上検討してきた理想的なTO BEモデルを実際に念頭にプロトタイプを作成し、幾つかの技術的検証項目（①データ統合、②メッセージ交換、③データ交換、④タイムスライス、⑤重複投与、⑥欠側値（Missing Data））を設定して評価分析を行った。

最近、医薬品開発に要する投資が拡大する一方、その成功率は年々低下して来ている。これは承認され、患者が利用できる新しい医薬品が少なくなって来ていることを意味している。Critical Path Initiativeとして取上げられ、様々な取り組みが行われている。海外データの活用（Bridging Study）や医薬品の世界同時開発（Simultaneous Global Study）に代表される取り組みは、良い薬を早く患者へ提供することを使命に進められている。さらに、医薬品開発段階において、実際に十分な日本人における検証データが少ないことを補うために市販後に医療機関における医薬品の使用実態を把握して、早期に医薬品の安全性情報のProfileを確立することが重要である。医薬品開発と安全性監視は「車の両輪」のごとく取り組んで行かなければならない。

本研究が目的とする全国の各医療機関で使用されている医薬品の処方箋情報・検査情報・患者情報・診療情報を特定の薬効群や患者群で大規模に収集可能な基盤（情報収集拠点医療ネットワーク）を構築することが出来れば、医薬品の安全性・有効性情報の詳細な分析・評価が可能になり、医薬品の安全性監視（Pharmacovigilance）にとって非常に有効な手立てとなる。このことは、積極的な科学評価分析に基づく「予測・予防型」への安全対策業務へ転換を図り、「攻めの安全対策」を目指す上で大いに期待できると考えられる。また、製薬企業の視点においても、患者の安全確保や優れた医薬品のライフサイクルマネージメントにおいて、大いに期待できる研究であると考えられる。

A. 研究目的

医薬品が承認されて上市されると使用される症例数が急激に拡大し、時間の経過とともに変化する。「実際の医療現場」で使用されることで治験の際では想定できなかったような様々なケースが発生し、予期しない副作用や相互作用が発生することがある。実際、医療機関では、患者プロフィール（背景因子）と合併症や併用薬の組み合わせからくる様々な症状を医師が診断を行い薬剤の処方が行われている。

従来から実施されている企業による副作用自発報告は、副作用が発生した際に報告する消極的な調査（Passive Surveillance）が主体であった。しかし、今後は何かが起こってから対応から起こる前の対応へ行動指針が大きくPro Activeな安全監視体制へ変換しようとしている。

従って、積極的な調査（Active Surveillance）を実施するために拠点医療機関（Sentinel Sites）をネットワーク化し、特定薬効群や特定疾患に絞って、医療機関における医薬品の使用実態を的確

に把握することで、医薬品の副作用を早期に発見し、その未然防止策を講じるため、“因果関係がありそうな医薬品と副作用の組み合わせ（シグナル）”を検出できることが重要である。

「実際の医療現場」である医療機関が情報の発生源であり、全国各地の医療機関から収集して来るとなると、その数は膨大である。また、各医療機関の医療情報の電子化状況は、様々であり、標準化は全く進んでいないのが現状である。また、収集してきたデータを統合して分析・評価を行うためには、医療機関側で発生した Event Data を調査の目的に沿って収集するために医師が Protocol の内容を理解して抽出した CRF Data として統合する必要がある。さらに分析・評価するために MedDRA などの辞書のコードに Coding された ANALYSIS Data が最終的に必要になる。医療機関における医薬品の適正な使用実態を把握するためには、このような Data の特性の違いを良く理解しておく必要がある。

今回、Data の特性の違いを十分認識した上で、初年度実施した各医療機関の電子化情報を調査分析した結果、患者情報・処方情報・検査情報・診療情報の電子化情報に関しては、医師の判断を必要とする診療情報（有害事象など）以外は、レセプトコンピュータやオーダリングシステムの導入により、電子化が進んでいることが判明した。しかし、データ構造やコード体系など各医療機関独自の形式が一般的で予想された通り、標準化は未整備であることも判明した。

次年度は、各医療機関側に新たなシステム構築を行わず既に電子化情報をダウンロードしてデータの提供のみで実現できる方法をデザインし、速やかに安全性情報を収集・解析し適正使用情報を得ることが可能かどうかを机上検証した。本研究を目的に電子化された情報ではないが、既に電子化された情報を大量に収集することで、収集した情報に含まれる不確定要素（欠測値など）の中から、意味あるシグナル検出ができるかが重要である。さらに、モデル薬剤を設定して、分担研究者所属の医療機関の電子化状況を前提に実証実験を実施するための Protocol（実施計画書）を作成し、倫理面での検討が十分行われた。

最終年度は、実証実験を実施するために作成した Protocol（実施計画書）である『HMG-CoA リダクターゼ阻害剤の使用成績調査をモデルに取上げ、電子化された患者情報を複数の医療機関から収集・集積して解析する方法に関する研究』に従って、各医療機関からデータを収集するため

の IT ツールのプロトタイプを開発して技術的検証を実施して、将来の安全性監視システムの実現に向けての展望を提言する。

B. 研究方法

臨床試験や市販後調査のデータマネジメントシステム（HITCANDIS/DM：日立製作所の製品）構築の経験を活かし、医療機関における医薬品の適正な使用実態を把握することを狙いとした特定の薬効群や患者群で大規模にデータを収集可能な基盤（情報収集拠点医療ネットワーク）を構築するためのシステム化の方式検討をベースにプロトタイプにて、検証すべき技術的課題を整理し評価を行った。

まず、今までの研究内容を整理し、最終年度の研究に取り掛かった。

1. 安全性監視ネットワークと大規模治験ネットワークの違い

臨床試験や市販後調査のデータマネジメントにおいて、最近ではインターネットを介して医療機関から直接データ入力を実施する EDC（Electronic Data Capture）が普及し、ネットワーク化が進んでいる。しかし、ブラウザに表示された CRF に対して、入力する必要がある。

一方、医薬品の医療機関における医薬品の適正な使用実態を把握するための安全性監視ネットワークは、医療機関で既に電子化された情報からデータ抽出して収集方法を目指している。表 1 に違いを整理する。

表 1：安全性監視ネットワーク vs 大規模治験ネットワーク

	安全性監視ネットワーク	大規模治験ネットワーク
試験の種類	観察試験	介入試験
試験の目的	Signal Detection	仮説の検証
試験の対象	全患者	サンプル患者
収集方法	医療情報システムからのデータ抽出	CRF 記入
Data Size	膨大	小さい

2. 治療と臨床研究

本研究は、医療・治療の現場である医療機関から標準医療・治療を創出する臨床研究の分野へ電子化された情報を集積して活用することに通じる。

治療と臨床研究の双方において電子化されたデータは、患者情報であり、情報の相互利用は

容易であるように思われる。しかし、電子化された情報を管理しているシステムは、その目的の違いにより、相容れない状況を作り出していることが現状である。図1にその違いを整理した。

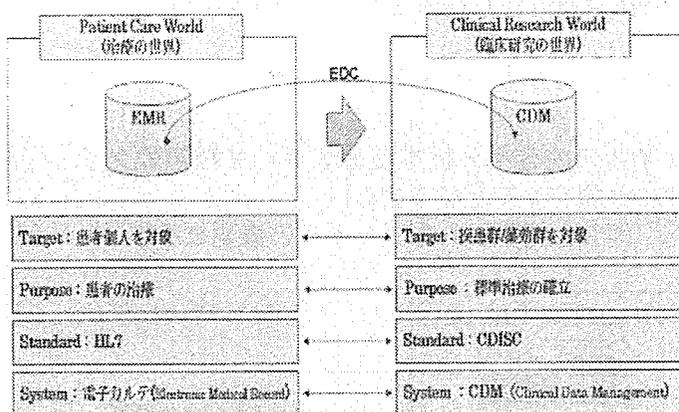


図1：治療と臨床研究

3. アンケートとヒアリング結果の纏め

本研究班への参加医療機関へのヒアリングや日本病院薬剤師会を通じたアンケート結果を整理することにより、データ抽出およびデータ統合を実施する上で考慮しなければならない項目を表2に列記した。

表2：各医療機関の電子化状況の実態

No.	電子化状況の実態
1	電子カルテの普及率は低い。
2	医療機関のシステム化状況は、レセプトコンピュータやオーダーリングシステムが中心。
3	データの関連情報が電子化されていない。(薬剤と疾患の関係) →医師が見れば、情報の関連はわかる。
4	医師の判断の伴うデータは、電子化されていない。
5	施設間のデータ共有が意識されていない。
6	欠側値が多い。

4. 実証実験の概要

【要求仕様】

各医療機関から抽出された医療情報を1つに統合し、薬剤に関する安全性と有効性を判断するシステムとしての将来運用を考察する。

【モデル薬剤】

スタチン系薬剤として、成分名で規定して下記薬剤を対象とする。

- ① アトルバスチンカルシウム水和物
- ② シンバスタチン
- ③ ピタバスタチンカルシウム
- ④ フルバスタチンナトリウム
- ⑤ プラバスタチンナトリウム

モデル薬剤の設定理由は、各医療機関の電子化状況の実態から、検査値など既存の電子化情報を用いることで薬剤の安全性と有効性を評価し易いからである。

【実施モデル医療機関】

北里大学病院/国立国際医療センター/国立循環器病センター/国立病院機構 東京医療センター/徳島大学病院/広島大学病院 の6医療機関である。

【各医療機関からの医療情報の統合方法】

図2に示すように6つの実施医療機関よりダウンロードされたCSV形式のデータからXML形式のデータに変換することでデータの統合を行う。

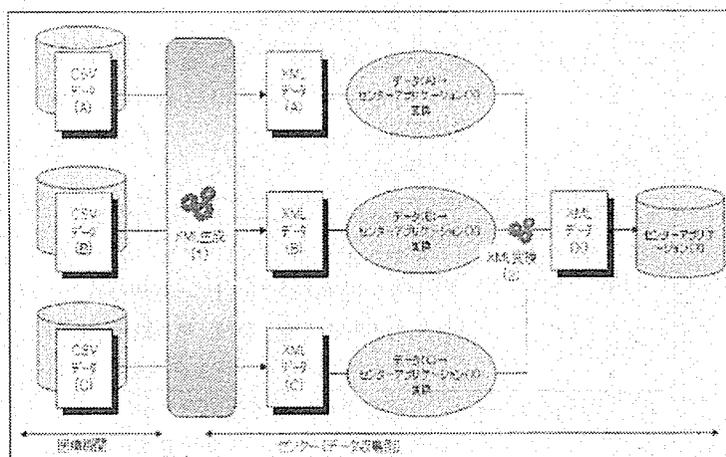


図2：医療情報の統合方法

- ① 各医療機関において、収集対象となるデータ項目を医療情報システムからダウンロードする。その際、匿名化対象の情報に関しては、データの提供前に各医療機関にて匿名化等を実施する。データを特定できるようにデータ提供前に施設を識別できるコードと患者のシーケンシャルNo.で構成されるKEY情報を付加する。
- ② 提供されるCSV形式などのファイルをXML形式に変換する際に、CSVなどデータ型が存在しないファイル形式では変換時にデータ型やサイズなどを設定する必要がある。そこで、CSV形式のファイルに関する添付資料として提出された各カラムの属性情報を参照して設定を行う。設定した内容をスキーマとして登

録することで、同様のデータフォーマットを取り込む際に再利用することができる。

- ③ 提供された医療機関からの XML データ毎にデータ構造変換スクリプトを指定することで異なるデータ構造の XML データから特定のデータ構造の XML データに変換する。

5. プロトタイプ作成における技術的検証項目

机上検証においては理想的な TO BE モデルを追及したデザインを実施したが、実証実験においては技術的検証項目を設定してプロトタイプにて評価を実施する AS IS のアプローチを試みた。表 3 に見極めるべき技術的課題を整理した。

表 3：技術的検証項目

No	技術的検証項目	内容
1	データ統合	医療機関の電子化データを同一患者の連続のデータとして連結できていることを検証する。
2	メッセージ交換	各医療機関でダウンロードされた電子化データを収集する側のデータ構造に変換することができるかを検証する。
3	データ交換	提供された電子化データと収集する電子化データに統合するために、属性・桁数の違いの対応、単位変換、コード変換、部分参照、項目結合によって解決できるかどうかを検証する。
4	タイムスライス	統合された電子化データをプロトコルのタイムウィンドウで患者のデータを相対的な時間軸関係に揃えることができるかどうかを検証する。
5	重複投与	処方量と実際の服薬量との違いを処方から自動的にどのように推定できるかを検証する。
6	欠測値 (Missing Data)	様々なケースの欠測値 (Missing Data) を自動的にどのように取り扱えば良いかを検証する。
7	全般的	不確定要素の多いデータから統計的に意味のあるシグナルが検出できるかどうかを検証する。

C. 研究結果

各医療機関から提供された CSV ファイルとファイルのフォーマットを記述したファイル仕様書をもとに各医療機関側の XML データを生成し、予め定義しておいた本研究側の統合データベ

ース側の構造 (Schema) を生成し、各医療機関側の XML データと統合データベース側の構造 (Schema) との Mapping 作業を行いデータ変換することでデータ統合を実現した。

1. 各医療機関から提供されたデータの状況

【ファイルフォーマット】

各医療機関から提供されたダウンロードファイルの内容を分析すると大きく 2 種類のタイプに分類することができる。

図 3 に提供されたファイルの概要を示す。

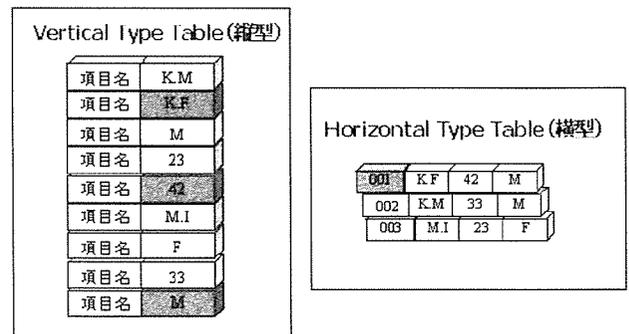


図 3：提供されたファイルの分類

患者ごとに複数の項目が横一列に並んでいるタイプのファイルで Horizontal Type Table (横型) 形式と 1 項目ごとに縦一列に積み上げたタイプのファイルで Vertical Type Table (縦型) 形式に大別できる。

【データ量】

各医療機関から提供されたデータ量を表 4 に整理する。

表 4：データ量 (単位：人)

患者数	北里大 学病院	徳島大 学病院	広島大 学病院	東京医 療セン タ	国立循 環器病 センタ	国立国 際医療 センタ
全体	5577	1310	2437	2407	-	3087
男性	2668	506	1008	1040	-	1362
女性	2909	804	1429	1429	-	1725
外来の み	2390	677	1232	-	-	1653
入院の み	1204	264	174	-	-	116
外来/ 入院	1320	369	915	-	-	1318