

不服を申し立てることができる。ただし、不服申し立ては当該データ提供申請に対して1回とする。

4) 運営委員会は不服申し立てについて迅速に審査し、その結果を申請者に通知する。

(5) データ提供に関する契約

- 1) 申請者と事務局はデータ提供に関する契約を結ぶ。
- 2) 契約書は事務局が提示する。

(6) データ等の提供

- 1) データ本体の提供は電子媒体で事務局に送付する。データ形式はテキストファイル（CSV ファイル形式が望ましい）とする。やむを得ない場合はエクセル、SPSS および SAS のファイル形式を認める。
- 2) データに付随する資料は電子媒体で事務局に送付する。その際、データ形式はマイクロソフトワード、エクセルもしくは PDF ファイル形式とする。

(7) データ提供の中止

- 1) 申請者はデータ提供の中止を運営委員会に文書で求めることができる。
- 2) 運営委員会はデータ提供の中止について審査し、結果を申請者に通知する。
- 3) 運営委員会がデータ提供の中止が認めた場合は、速やかにデータおよび付随する情報等を申請者に返還もしくは廃棄する。

(8) データの修正等

- 1) 申請者はデータに間違い等があることが判明した場合は速やかに事務局にその旨を届け出る。
- 2) 申請者はデータ等を修正し、修正したデータおよび付随する資料等を事務局に電子媒体で送付する。

3-1. データ利用の手順

- 1) データ・アーカイブのホームページで利用可能データを確認する。
- 2) データ利用申請書を事務局に提出する。
- 3) 事務局は利用申請書を当該データ提供者に通知する。
- 4) データ提供者は共同研究に可否を事務局に連絡する。
- 5) 事務局は共同研究の可否をデータ利用者に通知する。
- 6) 共同研究が可となった場合、事務局はデータ利用者と契約を結ぶ。
- 7) データ利用者は事務局に誓約書を提出する。
- 8) 事務局はデータ利用者にデータおよび付随する資料を提供する。
- 9) データ利用者は事務局に1年に1回、データ利用状況および論文等の成果を報告する。

3-2. データ利用マニュアル

(1) 登録されているデータの検索

- 1) データ利用申請者はホームページから登録されているデータを検索し、研究概要、利用の条件等を確認する。
- 2) 研究概要、利用条件など研究に関する質問は、データ提供者にする。事務局は利用手続きに関する質問以外は受け付けない。

(2) データ利用申請書の提出

- 1) データ利用申請者はホームページからWEB申請を行う。
- 2) 利用申請書の書き方は別紙を参照する。

(3) 共同研究の場合のデータ利用の意向確認（注：これもWEB上で可能か）

- 1) 事務局はデータ提供者に当該データ利用について共同研究の意向を確認する。
- 2) 事務局は共同研究の可否をデータ利用申請者に通知する。
- 3) 共同研究が不可の場合はその理由を事務局に提示し、事務局はそれをデータ利用申請者に通知する。

(4) データ提供者の意向が不要な場合

- 1) データ利用にあたって、データ提供者の意向が不要な場合は、その利用について運営委員会で審査する。
- 2) 事務局はデータ利用に関する審査の結果をデータ利用申請者に通知する。データ利用が不可の場合はその理由を審査結果に記す。

(5) データが利用できない場合の不服申し立て

- 1) データ提供者の意向により共同研究が不可となった場合に、データ利用申請者は不服申

し立てをすることができない。

- 2) 運営委員会のデータ利用に関する審査結果に対して、データ利用申請者は30日以内に事務局に文書で不服申し立てをすることができる。ただし、不服申し立ては当該データ利用申請に対して1回とする。
- 3) 運営委員会は不服申し立てに対して、迅速に審査し、事務局はその結果をデータ利用申請者に通知する。

(6) 共同研究の協議とデータ利用の許可

- 1) データ利用者はデータ提供者と共同研究に関する協議をする。
- 2) 共同研究が可能となり、データ提供者がデータの利用を許可した場合は、データ利用者は事務局にデータの授受を求める。
- 3) 事務局はデータ利用申請者からのデータの授受に関して、データ提供者に確認を取る。
- 4) 事務局はデータ利用の許可が確認でき次第、速やかに、利用契約、誓約書などのデータの授受に関する手続きを行う。

(8) データ利用に関する契約

- 1) データ利用申請者は事務局とデータ利用に関する契約を結ぶ。
- 2) 契約書は事務局が提示する。

(9) データ利用に関する誓約書

- 1) データ利用申請者はデータ利用に関する誓約書を事務局に提出する。
- 2) 誓約書は事務局が提示する。

(10) データ等の授受

- 1) 事務局はデータ本体および付随する資料を電子媒体で提供する。

(11) データ利用の成果の報告

- 1) データ利用申請者は事務局に1年に1回、データ利用状況および論文等の成果を報告する。
- 2) 利用状況はホームページ等で公表する。

(12) データ利用にあたっての事故等の発生について

- 1) データ利用申請者はデータの紛失等の契約に反する事態が生じた際には速やかに事務局に連絡することとする。
- 2) 事務局は事故等の発生について、詳細な報告書をデータ利用申請者に提出させるとともに、運営委員会で被害拡大防止策及びデータ利用申請者の処分を協議する。

- 3) 事務局はデータ利用申請者に被害拡大防止策等および処分に関する運営委員会の決定事項を通知する。
- 4) データ利用申請者は事務局からの通知を受けて、速やかに通知事項に対応し、対応した事項を事務局に報告する。

4-1. 教育目的での利用に関するマニュアル

(1) 教育目的の利用の原則

- 1) データ利用申請は共同研究等のデータ利用申請に準ずる。
- 2) 教育目的の利用にあたっては、教育目的を明確にし、当該講義・セミナー・研修会内での利用のみとし、当該データおよび解析結果の外部への持ち出しを固く禁ずる。
- 3) 万が一、利用規定に反した場合の責任の一切は、データ利用申請者にある。

(2) 教育目的で提供するデータ本体の加工に関して

- 1) 教育目的で利用できるデータは、オリジナルのデータの件数を減じるとか変数を減じるなどの加工を施し、万が一、データ利用者が利用規定に違反した場合でも、最小限の被害にとどまるようにしておくことが望ましい。

(3) データ利用申請書、契約書、誓約書

- 1) データ利用申請書は共同研究に準じるが、データの利用者については「受講生〇人」などとし、氏名等は不要とする。
- 2) 契約書および誓約書は共同研究に準じる。

(4) データの廃棄

- 1) 教育目的のデータの利用・保存期間は最長1年間とし、速やかに廃棄し、データ利用申請者はその旨を事務局に報告する。

資料

米国でのデータベース研究活用: 事例と留意点

瀬戸口 聡子 (せとぐち そうこ)

准教授

デューク臨床研究所、デューク大学医学部
東京大学医学大学院薬剤疫学講座

Outline

- 米国のデータベース
 - データベースの紹介
 - 活用例
 - アクセスの問題点
- 米国でのデータベース活用推進の試み
 - BioLINCC by NHLBI
- 米国でのデータベース標準化の試み
 - OMOP
 - Mini Sentinel
 - AsPEN and SCAN
 - PCORI NDRN and PPRN

Computerization of Health-care Related Records

- Records for billing purposes were standardized/computerized
 - 1970's in pharmacoepidemiology and health service research
- Computerization of health information (e.g., as a form of EHR) is rapidly developing and being used

データベース

- 電子カルテデータ
- クレームデータ(=日本のレセプトデータ)
- 患者レジストリデータベース
- コホート研究データベース

2/16/2014

Setoguchi S

4

クレームデータベース (日本で言うレセプトデータ)

- 北米での例
 - Medicaid
 - Medicare
 - Commercial insurance companies
 - United Health
 - Blue Cross Blue Shield
 - Canadian Provincial claims data
 - Ontario
 - Quebec
 - British Columbia
 - Saskatchewan

2/16/2014

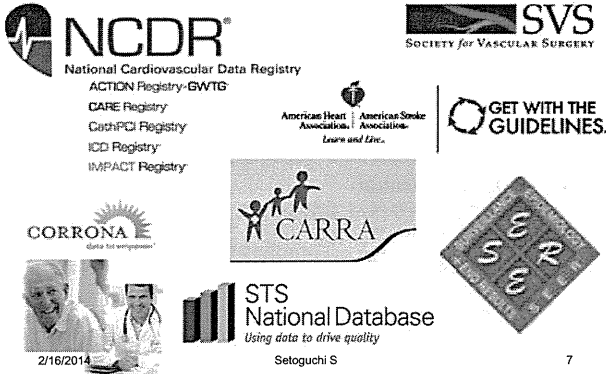
Setoguchi S

5

クレームデータの例

- Medicaid
- Medicare + State Pharmacy Assistance Programs or Part D
- Commercial insurance companies
 - United Health
 - Blue Cross Blue Shield
- MarketScan
- Premier Database

レジストリの例



2/16/2014

Setoguchi S

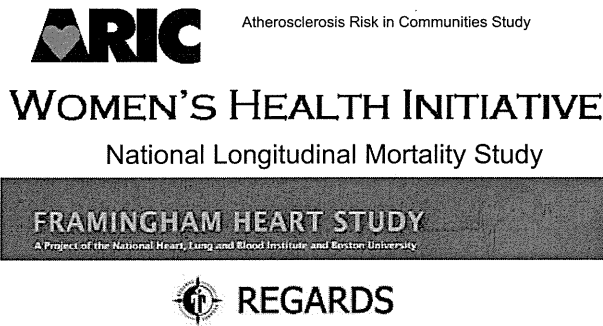
7

Electronic Health Record (EHR) Database

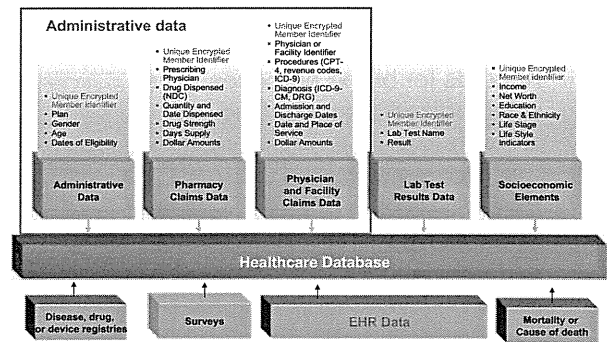
Examples

- Single provider
 - DEDUCE (Duke)
 - RPRD (Brigham and Women's Hospital)
- Multiple providers
 - Geisinger Clinic Electronic Health Records- 41 Clinics covering ~3 million patients
 - Kaiser Permanente

Cohort Studies



Better Data



データの種類	例	特徴
大規模コホート	<ul style="list-style-type: none"> • Nurses Health Study • Framingham • REGARDS 	前向きで項目が統一されたデータ。カルテやクレームデータにない項目も収集可能。データ収集が継続的でなく断片的(1-2年に一回)、詳しい薬剤の情報が少ない。クレームデータと比較すると比較的小規模
疾患レジストリ	<ul style="list-style-type: none"> • State Cancer Registries • SEER • NCDR • CORRONA • CARRA 	疾患臨床情報が豊富。薬剤使用の情報はあまり充実していないため、単独では安全性研究がしにくい。他のデータリンクし補助的に利用
クロスセクショナルサーベイ	<ul style="list-style-type: none"> • NHANES 	情報が豊富(臨床データや生活習慣データ)。クロスセクショナルであり因果関係の追及に限界がある

データの種類	例	特徴
クレームデータベース (レセプトデータ)	<ul style="list-style-type: none"> • Medicaid claims data • Medicare claims data • Claims Data from Large Private Insurers (BCBS/United Health) • MarketScan Data 	大規模、ポピュレーションベース、薬剤使用の詳細かつ連続的データ、低費用。詳細臨床データはない、診断名に誤分類
電子カルテ	<ul style="list-style-type: none"> • Geisinger Clinic Electronic Health Records • Kaiser Permanente 	詳細な臨床情報(ラボやあらゆる検査のデータも豊富)、処方情報のみで、薬局での受け取ったかがどうかはわからない、ポピュレーションベースでないためフォローが不完全、アウトカムのデータが不完全、データ項目が統一されていない

Examples of Studies Using Large Databases

Disease Burden

Trends in Out-of-Hospital Deaths Due to Coronary Heart Disease in Sweden (1991 to 2006)

Kerstin Dudas, PhD; Georg Lappas, BSc; Simon Stewart, PhD; Annika Rosengren, MD, PhD

Background—Case fatality associated with a first coronary event is often underestimated when only those who survive to reach a hospital are considered. Few studies have examined long-term trends in case fatality associated with a major coronary event that occurs out of the hospital.

Methods and Results—Record linkage documented all case subjects 35 to 84 years of age in Sweden during 1991 to 2006 with a first major coronary event (out-of-hospital coronary death or hospitalization for acute myocardial infarction). Of the 384 597 cases identified, 111 319 (28.9%) died out of the hospital, and another 36 552 (9.5%) died in the hospital or within 28 days of hospitalization. Mortality due to CHD in the hospital (within 28 days) and out of the hospital per 100 000 population 35 to 84 years of age, 1991 to 2006.

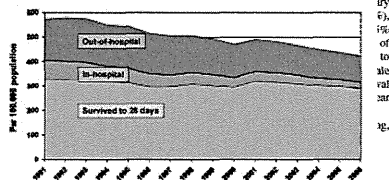
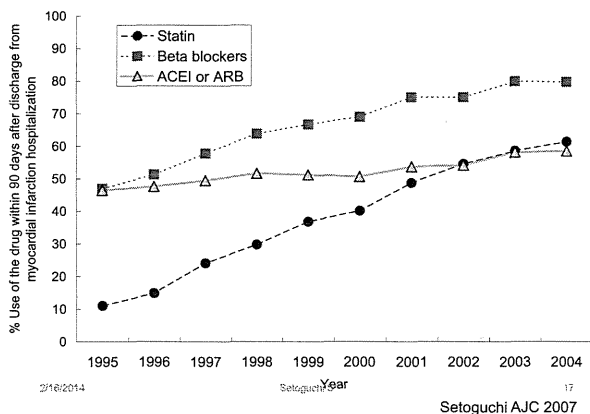


Figure 3. Mortality due to CHD in the hospital (within 28 days) and out of the hospital per 100 000 population 35 to 84 years of age, 1991 to 2006.

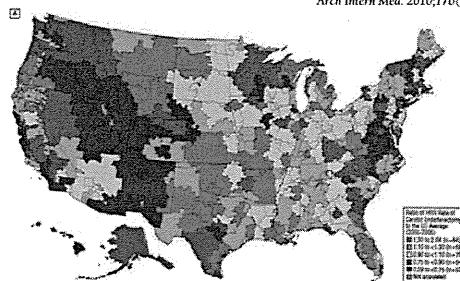
Health Services Research



Geographic Variation in Carotid Revascularization Among Medicare Beneficiaries, 2003-2006

Manesh R. Patel, MD; Melissa A. Greiner, MS; Lisa D. DiMartino, MPH; Kevin A. Schulman, MD; Pamela W. Duncan, PhD, PT; David B. Matchar, MD; Lesley H. Curtis, PhD

Arch Intern Med. 2010;170(14):1218-1225

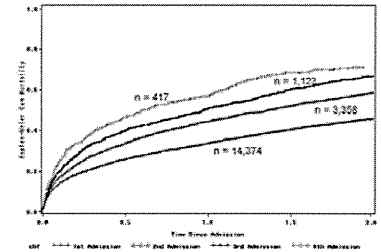


Outcomes Research

Congestive Heart Failure

Repeated hospitalizations predict mortality in the community population with heart failure

Soko Setoguchi, MD, DrPH,¹ Lynne Warner Stevenson, MD,² and Sebastian Schneeweiss, MD, ScD,¹ Boston, MA



Conclusions The number of HF hospitalizations is a strong predictor of mortality in community HF patients. This simple predictor of mortality in HF patients should help triage management and resources for HF and trigger patient planning for prognosis. (*Am Heart J* 2007;154:260-6.)

20

Journal of the American College of Cardiology
© 2008 by the American College of Cardiology Foundation
Published by Elsevier Inc.

Vol. 51, No. 13, 2008
ISSN 0735-1097/08/\$14.00
doi:10.1016/j.jacc.2007.10.062

CLINICAL RESEARCH

Acute Myocardial Infarction

Improvements in Long-Term Mortality After Myocardial Infarction and Increased Use of Cardiovascular Drugs After Discharge

A 10-Year Trend Analysis

Soko Setoguchi, MD, DrPH,¹ Robert J. Murray A. Mittleman, MD, DrPH,² Ra Boston, Massachusetts

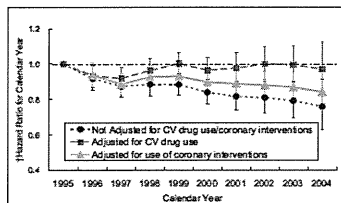
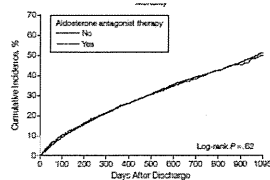


Figure 1 Hazard Ratio for Calendar Year Before and After 2001 Adjusting for CV Drug or Coronary Intervention

Comparative Effectiveness and Safety Studies

ORIGINAL CONTRIBUTION

Associations Between Aldosterone Antagonist Therapy and Risks of Mortality and Readmission Among Patients With Heart Failure and Reduced Ejection Fraction



No. at risk	4517	4380	4070	3780	3483	3088	2770	2388	2114	1859	1585
No	4517	4380	4070	3780	3483	3088	2770	2388	2114	1859	1585
Yes	1070	951	857	841	779	701	629	557	483	429	344

Adrian F. Hernandez, MD, MHS
Xiaojuan Mi, PhD
Bradley C. Hammill, MS
Stephen C. Hammill, MD
Paul A. Heidenreich, MD
Frederick A. Masoudi, MD, MSPH

JAMA. 2012;308(20):2097-2107

Statins Prevent Cancer?

Epidemiology

Statins and the Risk of Lung, Breast, and Colorectal Cancer in the Elderly

Soko Setoguchi, MD, DrPH; Robert J. Glynn, PhD, ScD; Jerry Avorn, MD; Helen Mogun, MS; Sebastian Schneeweiss, MD, ScD

Background—Although most randomized trials and meta-analyses suggest a slight or no increase in the risk of cancer in statin users, results from observational studies have been conflicting, and some have even suggested a large protective effect of statins on certain cancers. Long-term statin users tend to be healthier, less frail, and more adherent to therapy than nonusers, however. This could explain such apparent “protective” effects.

Methods and Results—We conducted the present cohort study by linking data from a large state drug benefit program with cancer registry data and Medicare healthcare utilization data. We identified all initiators of statins; initiators of glaucoma medications, another preventive drug, served as a comparison group. Outcomes included all registry-identified cases of colorectal, lung, and breast cancer. Multivariable Cox proportional models were used to adjust for confounding. Patient characteristics were similar in both groups, but statin initiators (n=24 439) were slightly younger and used some services more frequently than glaucoma drug initiators (n=27 84). The mean follow-up was 2.9 years, with the longest follow-up being 8.4 years. Incidence rates of colorectal, lung, and breast cancers in both groups were very similar to rates in the general population. Adjusted hazard ratios were 0.96 (95% CI, 0.70 to 1.31) for colorectal cancer, 1.11 (95% CI, 0.77 to 1.60) for lung cancer, and 0.99 (95% CI, 0.74 to 1.33) for breast cancer.

Conclusions—These data from a large population of typical older patients who began using statins indicate that it is unlikely that statins confer a clinically important decrease or increase in the risk of colorectal, lung, or breast cancer over the durations studied. (*Circulation*. 2007;115:27-33.)

Timely Assessment

Comparison of Cardiovascular Outcomes in Elderly Patients With Diabetes Who Initiated Rosiglitazone vs Pioglitazone Therapy

Wolfgang C. Winkelmayer, MD, ScD; Soho Setoguchi, MD, DrPH; Raisa Levin, MS; Daniel H. Solomon, MD, MPH

Background: Recent meta-analyses have raised the possibility that rosiglitazone therapy may increase the risk of ischemic cardiovascular events, whereas pioglitazone hydrochloride could not be linked to such a risk. We compared cardiovascular outcomes and mortality between patients initiating pioglitazone vs rosiglitazone therapy.

Methods: We assembled an inception cohort of Medicare beneficiaries older than 62 years with state-sponsored prescription drug benefits who had diabetes mellitus and initiated treatment with rosiglitazone or pioglitazone between January 1, 2006, and December 31, 2005. The study outcomes included all-cause mortality, myocardial infarction, stroke, and hospitalization for congestive heart failure.

Results: Of 28 961 patients selected, 50.3% initiated treatment with pioglitazone and 49.7% with rosiglitazone. Most baseline characteristics were similar between the groups. As preferred in drug safety research, we censored patients at crossover or at 60 days after discontinuation of

therapy with their study drug; during 28 960 person-years of follow-up, 1859 patients died. After adjustment for a large number of patient characteristics, Cox regression models revealed 15% greater mortality among patients who initiated therapy with rosiglitazone compared with pioglitazone (95% confidence interval, 5%-26%). Use of rosiglitazone was also associated with a 15% greater risk of congestive heart failure (95% confidence interval, 1%-26%). No differences between the 2 drugs were found in their rates of myocardial infarction or stroke.

Conclusions: Our findings from a large population-based cohort of US seniors are compatible with an increased risk of all-cause mortality and congestive heart failure in patients initiating therapy with rosiglitazone compared with similar patients initiating therapy with pioglitazone. Limitations of this study include residual confounding due to its nonrandomized nature.

Arch Intern Med. 2006;166(23):2360-2375

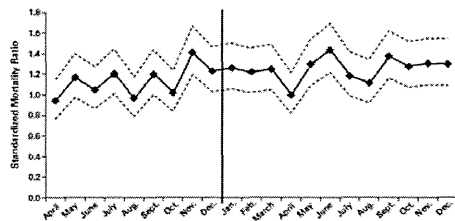
Policy Evaluation

The New England Journal of Medicine

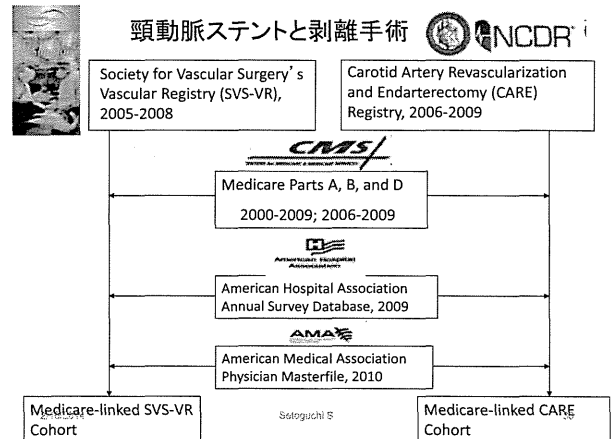
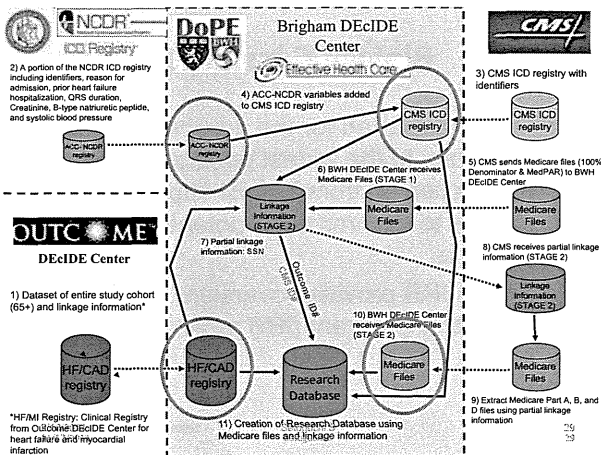
Special Article

OUTCOMES OF REFERENCE PRICING FOR ANGIOTENSIN-CONVERTING-ENZYME INHIBITORS

SEBASTIAN SCHNEWEISS, M.D., ALEXANDER M. WALKER, M.D., ROBERT J. GLYNN, PH.D., MALCOLM MACLURE, SC.D., COLIN DORMUTH, M.A., AND STEPHEN B. SOUMERAI, SC.D.



データへのアクセス



頸動脈狭窄の薬剤療法と手技の比較

BWH, 2002-2009 VL Imaging & EHR data

MGH, 2006-2009 VL Imaging & EHR data

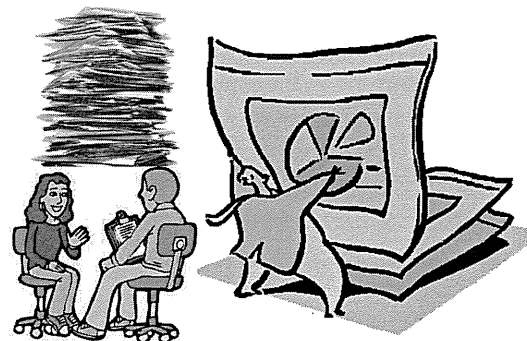
AHA Annual Survey Database, 2008

AMA Physician Masterfile, 2010

Medicare Inpt, Outpt, and Dispensing Data

2/16/2014

データ収集 vs. 既存のデータベース



2/16/2014 Setoguchi S 32

It is still a lot of work to access ALREADY EXISTING databases..

- 2年間
- 政治的な圧力、事務的手続き



2/16/2014 Setoguchi S 33

データベースの活用推進の例 NHLBI Bio LINCC

NHLBIのデータバイオスペシメンアーカイブシステム

- <https://biolincc.nhlbi.nih.gov/redirect/>
- Data preparation guideline
– <http://www.nhlbi.nih.gov/funding/setpreparation.htm>

BioLINCC by NHLBI

- Program started for the purpose of making available, to the widest possible audience, data collected from NHLBI supported studies to maximize the potential value of the data.
- NHLBI IRB provides oversight of the data repository program (IRB protocol 01-H-0031).

Sean Coardy NHLBI 2010

Timing for release of data through Data Rep

- **Clinical Trials:**
 - ✓ 3 years after the final visit of the participants to their clinical trial sites, or
 - ✓ 2 years after publication of primary outcome paper
- **Observational Epidemiology Studies:**
 - ✓ 3 years after the completion of each examination or follow-up cycle, or
 - ✓ 2 years after the baseline, follow-up, genetic, ancillary study, or other data set is finalized for within the study for analysis

Sean Coardy NHLBI 2010

Guidelines for submission of data to the repository

- **Data submitted by a study should:**
 1. Be prepared in a manner consistent with the informed consent.
 2. Obvious identifiers removed (i.e. HIPAA identifiers).
 3. Geography is generally removed.
 4. Dates are recoded as a time interval.
 5. Other data may be recoded to minimize the possibility of direct identification of a participant.

Sean Coardy NHLBI 2010

Procedures for obtaining data from the repository

1. Receive IRB approval from an expedited or full review for their project (IRB letters granting exemption are not sufficient).
2. Description of research project
3. Sign a Research Materials Distribution Agreement that:
 - ✓ Prohibits the transfer of data
 - ✓ Limits use of the data to three years
 - ✓ Specifies a required acknowledgment in manuscripts/abstracts
 - ✓ Requires notification of any new projects
4. Obtain institutional sign-off on the research materials distribution agreement.

Sean Coardy NHLBI 2010

Data Requests

- Data requests are subject only to a review for appropriateness.
- There is no review of scientific merit for data requests .

Sean Coardy NHLBI 2010

Current Status of the Data Repository (As of 2010)

- Data from more than 440,000 participants in 76 observational and clinical trial studies.
- **20 Observational studies:**

General population studies (Framingham, ARIC, CHS, CARDIA, MESA); Sarcoidosis (ACCESS); Sickle Cell (CSSCD); Iron Overload (HEIRS); Primary Pulmonary Hypertension (PPH); Human retrovirus infected blood donors (REDS-HTLV); Alpha₁ Antitrypsin deficiency patients (Alpha₁ registry); Cord blood transplantation (COBLT); Post-menopausal women (WHI-OS); Adolescents (Bogalusa, NGHS)

Sean Coardy NHLBI 2010

Current Status of the Data Repository (As of 2010)

- **56 Clinical Trials:**

Asthma Clinical Research Network Trials (BAGS, CIMA, SOCS/SLIC, MICE, DICE, BARGE, IMPACT, PRICE); Acute Respiratory Distress Network (ARMA/LARMA, ALVEOLI, FACTT, LaSRS); Community Interventions (REACT, PAD); Hypertension Treatment (TOHP I&II, DASH-Sodium, PREMIER); Primary Prevention (MRFIT, ACT); Secondary Prevention (BEST, BHAT, AMIS, CASS, MAGIC, SOLVD, TIMI II & III, PEACE, ESCAPE, AFFIRM); Diagnostic (PIOPED); Treatment (LHS, NETT, RTS, VATS); Primary/Secondary Prevention (HDFP, SHEP, WHI-CT)

Sean Coardy NHLBI 2010

Since 2000... (As of 2010)

- 498 Requests have been approved (381 investigators) for 818 study datasets
- 74 requests (15%) were from foreign institutions
- Over 224 Manuscripts have been published in over 68 Journals

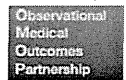
Sean Coardy NHLBI 2010

データベース標準化の試みの例

ご紹介する例

- OMOP (Observational Medical Outcomes Partnership)
- FDA Mini-Sentinel
- AsPEN (Asian Pharmacoepidemiology Network)
- PCORI NDRN PPRN

Observational Medical Outcomes Partnership OMOP



<http://omop.org/>

- 2008年に始まったUSの”public-private partnership”
- Claims database + HER databaseを使った medical productsのreal-world effects研究のためのmethodological research
- “currently funded by the Reagan-Udall Foundation(RUF)* for the FDA through contributions from AstraZeneca and Pharmaceutical Research Manufacturers of America (PhRMA)”
- RUF: FDAAAIに規程のある財団。FDA自体の新たなdrug safetyのための initiative (mini-sentinelとして実現)を支えることが想定されていた。

OMOP Databases

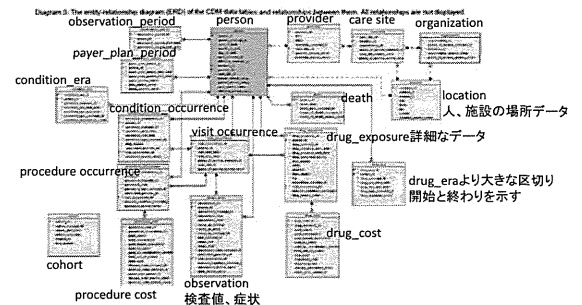
Table 1. Specific Databases Included in the Observational Medical Outcomes Partnership

	Database Name	Database Description	Provider No. (millions)
EHR	General Practice Healthcare (Olmstead, Wisconsin)	Defined from data entered from providers with state-wide Health Care Quality Improvement Program (HQIP) and Health Information Accuracy (HIA) program	11.2
Claims	TriNet Health-Analytic, Inc. (Ave. Adv. Michigan)	Medicaid (MI) Database represents 6 years of shared data with administrative claims from Medicaid, Outpatient, and Pharmacy services (representative of the population)	1.5
	MarketScan Medic and Med-Share Database	MarketScan Medic and Med-Share Database contains administrative claims data for Medicaid and Medicare from multiple states	11.1
	MarketScan Medicare Supplemental and Commercial of Revenue	MarketScan Medicare Supplemental and Commercial of Revenue Database represents administrative claims for Medicare and Medicaid beneficiaries (includes both employees, retired persons, post-65 Medicare beneficiaries, employer paid portion, and any and associated expenses)	4.4
Claims	MarketScan Commercial Claims and Encounter Database	MarketScan Commercial Claims and Encounter Database represents administrative claims and pharmacy claims from multiple states with patient-level identifiers	50
	Humana, Inc. (Miami, Florida)	Contains medical (inpatient, outpatient, and emergency room), pharmacy, and laboratory data (including test results) from Humana's administrative claims database of medical members	6.6
EHR	Partners HealthCare (Boston, Massachusetts)	Includes data from the Partners HealthCare Clinical Research and Informatics (PHRI) database as well as inpatient and outpatient (long-term) data from the Partners HealthCare (PHC) database	3
EHR	Regeneron Institute, Indiana Network for Patient Care (Indianapolis, Indiana)	Contains population-based, longitudinal, and observational data including inpatient, outpatient, pharmacy, medical, public health, laboratory, behavioral, reproductive, and other data from Regeneron, its affiliated hospitals, and other partners	2
Claims	CVS Health, LLC (Plymouth Meeting, Pennsylvania)	Contains health insurance (Medicaid and Marketplace) Administrative, Outpatient, and Inpatient patient level data from inpatient, outpatient, pharmacy, and other data and pharmacy services from all US states	90
EHR	Department of Veterans Affairs (Washington, DC)	MarketScan Database contains data for US veterans from prospectively collected and de-identified health records across facilities within the Veterans Affairs Health Care System as managed by the Veterans Affairs Center for Medication Safety	2

- 米国内の10データソース 1億3000万人をカバー
- Claims databaseとElectronic Health Record typeのデータベースを含む

Amer J Epidemiol 2013; 178:645-651

OMOP Common Data Model



- 四角の囲みが各テーブル種類とその要素を示す

FDAのミニセンチネル

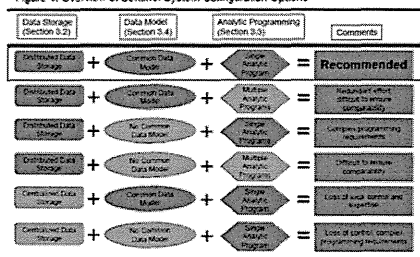
- http://www.mini-sentinel.org/data_activities/distributed_db_and_data/default.aspx

FDA Amendments Act FDAAA (2007)

- 1962年のKefauver-Harris修正法以来の大改正
 - 直接にはrofecoxibの心血管リスクに対するFDAの対応への批判などがきっかけ
 - FDAAAにおいて「FDAは“new safety monitoring system”を作り、2010.07までに2500万人、20120.7までに1億人のデータへのアクセスが可能にするように」とのゴールが明記され(達成され)た。
 - “Sentinel Initiative”と、その「パイロットスタディ」としてのmini-sentinel
 - 恒常的、系統的なマルチデータベース構築の原動力の一つ

Common Data Modelの提唱

Figure 1. Overview of Sentinel System Configuration Options



- Data storage: Distributed data storage
- Common Data Model
- Single Analytic Program

- Data Partnerは自らのデータを自ら管理する (distributed data storage)
- 各サイトで、データを共通フォーマットに変更 (Common data model)。
- 必要な時には、単一の解析プログラム (Single Analytic Program) を各サイトに送付し、返送された結果を集計。

Mini-Sentinel Common Data Model (MSCDM)

• Common Data Model v3.0(2013.9.16)

- The Mini-Sentinel Operations Center (MSOC) coordinates all Mini-Sentinel activities and the network of Mini-Sentinel Data Partners.
- MSCDM is a standard data structure that allows Data Partners to quickly execute distributed analytic programs against local data.
- The MSCDM currently includes 11 tables that represent information for the data elements needed for Mini-Sentinel activities.

http://mini-sentinel.org/data_activities/distributed_db_and_data/details.aspx?ID=105

Mini-Sentinel Distributed Database – 18 Data partners

- **Over 130 million individuals**
 - 382 million person-years of observation time
 - 41 million individuals currently enrolled
 - 34 million individuals have over 3 years of data
- 3.7 billion dispensings
- 4.1 billion unique encounters, including 46 million acute inpatient stays
- 24.3 million members with at least 1 laboratory result

Asian Pharmacoepidemiology Network (AsPEN)



AsPEN Special Interest Group

- Approved as one of SIGs within International Society of Pharmacoepidemiology (ISPE), April 2012
- Mission
 - To develop and advance multi-national database research in Pharmacoepidemiology in the Asia/Pacific region
- Website: aspennet.asia

SCAN Project To Develop Infrastructure in AsPEN

- Surveillance of Health Care in Asian Network (SCAN) Project: Infrastructure Development and Analysis with the Asian Pharmacoepidemiology Network
- Coordinating center: DCRI (PI: Setoguchi)
- Sites: Japan, Taiwan, HongKong, China, and Korea
- Collaborative projects with international sites and industry investigators
- To gain insights of health status and health service use in East Asian countries

Data Harmonization Effort by PCORI

- PCORI committed \$68 million in April to the development of the NCRN through the creation or expansion of up to eight health system-based Clinical Data Research Networks (CDRNs) and up to 18 Patient-Powered Research Networks (PPRNs) organized by groups of patients interested in participating in clinical research
- CD RN
 - <http://www.pcori.org/funding-opportunities/funding-announcements/clinical-data-research-networks-cdrn-improving-infrastructure-for-conducting-patient-centered-outcomes-research-pcor/>
- PPRN
 - <http://www.pcori.org/funding-opportunities/funding-announcements/patient-powered-research-networks-pprn-improving-infrastructure-for-conducting-patient-centered-outcomes-research-pcor/>
- To prepare for NCRN and PPRN PCORI/IOM sponsored a workshop inviting national experts in database network to discuss models and issues for data harmonization efforts.
 - <http://www.iom.edu/Activities/Quality/VSR/2013-OCT-31.aspx>

Data Harmonization for Patient-Centered Clinical Research

When: October 31, 2013 - November 1, 2013 (8:30 AM Eastern)
Where: National Academy of Sciences Building (Room 120) • 2101 Constitution Avenue, NW, Washington, DC 20546

Topics: Biomedical and Health Research, Health Care Workforce, Quality and Patient Safety, Public Health, Health Services, Coverage, and Access

Activity: Roundtable on Value & Science-Driven Health Care

Board: IOM Executive Office

A workshop, sponsored by the Patient Centered Outcomes Research Institute (PCORI) and convened by the Institute of Medicine, will consider in detail the issues and activities required of the National Patient-Centered Clinical Research Network (NCRN) in planning and implementing their inter-network data harmonization efforts. Workshop participants will provide expert input into priority areas for work and planning, given the terms of the PCORI awards; reflect on the current state of the art approaches to addressing challenges; and highlight data harmonization issues in need of further development in order to achieve the goal of network interoperability within 18 months.

Workshop discussions will focus on:

1. Informing the work of PCORI's NCRN awardees and Coordinating Center in working toward inter-network data harmonization;
2. Reflecting on past efforts and current state of the art approaches to help identify barriers and challenges, share best practices, and identify options for progress in the area of data harmonization;
3. Identifying strategic priorities for accelerating data interoperability between PCORI NCRN awardees and existing national data sharing activities.

While this workshop is currently at capacity for in-person attendance, please contact Elizabeth Johnston for information on additional viewing options.

INSTITUTE OF MEDICINE Advising the nation • Improving health
OF THE NATIONAL ACADEMIES

ABOUT THE IOM REPORTS ACTIVITIES MEETINGS Explore by Topic

BROWSE HISTORY

Meeting



Data Harmonization for Patient-Centered Clinical Research

When: October 31, 2013 - November 1, 2013 (8:30 AM Eastern)
Where: National Academy of Sciences Building (Room 120) • 2101 Constitution Avenue, NW, Washington, DC 20546

Topics: Biomedical and Health Research, Health Care Workforce, Quality and Patient Safety, Public Health, Health Services, Coverage, and Access

Activity: Roundtable on Value & Science-Driven Health Care

Board: IOM Executive Office

Desired Features of Data Model

- Transparency
- Access/Sharing
- Scalability and flexibility

アーカイブ、標準化における留意点

- データを巡る政治的圧力と最小化と事務手続きの標準・単純化
- 個人情報保護と研究に必要な個人情報へのアクセスへのバランス
- コスト(フリー あるいは アフォーダブル)
- データの適正使用と研究の質を保つ為のサポート
- 共通化によるデータのマニピュレーションと情報のロスとデータ使用の簡易性とのバランス

統計行政における データ利用の動向

～調査票情報の二次的利用を中心として～

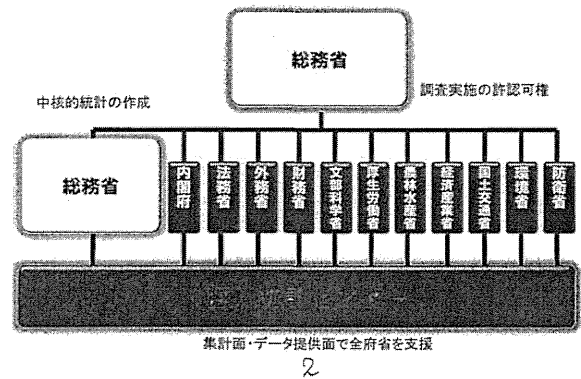
発表 上田 聖 (総務省統計局) 協力 総務省統計局
(独)統計センター

1. 総務省統計局、統計センターの紹介
2. 紀元前 匿名データ創世記
3. 改正統計法によるデータ利用
 - (1) データ利用の種類と匿名データ
 - (2) 擬似マイクロデータ
4. 次期統計行政の動向
5. 統計利用促進、ビックデータ

/

総務省統計局、 (独)統計センターの紹介

我が国の統計作成の体制(分散型)



紀元前(統計法改正前(2009年以前)) 匿名データ創世記

昔は
ひたすらデータを保護し
正確な統計を追及する時代

3

統計法

第十五条 何人も、指定統計を作成するために集められた調査票を、統計上の目的以外に使用してはならない。

2 前項の規定は、総務大臣の承認を得て使用の目的を公示したものについては、これを適用しない。

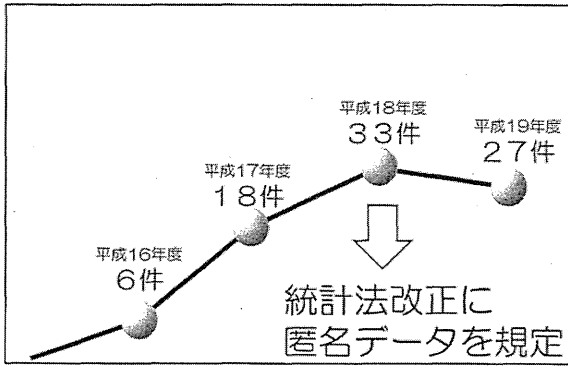
外国は
データ利用時代へ



我が国はデータ
後進国へ

4

試行的匿名データの提供実績



改正統計法における データ利用

5

2005年

「行政のための統計から
社会の情報基盤としての統計へ」

2007年

統計法全部改正

統計法における調査票情報（マイクロデータ）の利用規定

秘匿しない調査票情報の利用

- ・行政機関、地公体、独法
- ・上記と共同研究を行う者
- ・科研費を受けている者

秘匿処理した調査票情報の利用（匿名データの利用（手数料））

- ・学術研究を行う者
- ・高等教育を行う者

行政機関に特別集計を依頼（オーダーメイド集計（手数料））

- ・学術研究を行う者
- ・高等教育を行う者

6

匿名データ

匿名データ化された統計調査

【総務省】

- 全国消費実態調査（8万世帯の家計簿）
- 住宅・土地統計調査（350万世帯の住宅・居住）
- 就業構造基本調査（50万世帯の就業状況・転職）
- 社会生活基本調査（10万人の行動）
- 労働力調査（10万人の毎月の就業状況）

【厚生労働省】

- 国民生活基礎調査

7

実際に行われた研究の例

調査名	研究者	論文タイトル
全国消費実態調査	一橋大学経済研究所 高山 憲之	わが国世帯における消費税の負担水準
	三菱総合研究所 白石 浩介	給付つき税額控除による消費税負担の軽減
全国消費実態調査	慶應義塾大学大学院 渡辺 久里子	研究の内容等価尺度の推計と比較 —消費上の代価・家賃内訳・OECD国—
全国消費実態調査	北海学園大学 木村 和範	所得格差変動の年齢階級別要因分解

例えば

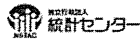
「医療費の詳細な支出」



「食料品の品目別の支出」

の関係などを8万世帯毎のマイクロデータで分析して健康に関する仮説の抽出などに使ってほしい。

8



統計センターについて | 個人情報 | 調査情報 | 情報公開 | 採用情報

- 業務案内
- 統計作成業務
- 公約統計基礎サービス
- 統計利用向けサービス及び公約統計サービス
- 公約統計のソフトウェア利用
- 匿名データの利用
- 労働力調査
- 住宅・土地統計調査
- 全国消費実態調査
- 経済生活基本調査
- 社会生活基本調査
- 匿名データの利用に関するFAQ
- アンケート調査の活用
- イベント開催
- アンケート利用
- アンケートベース教育
- 統計と匿名データの活用
- 学術研究推進等との連携
- 他府県における提供状況
- 資料案内
- 技術研究
- お問い合わせ

業務案内

匿名データの利用

提供する匿名データ

総務省
・ 経団連 (FEDERATION)
○ 全国消費実態調査 (最終更新日: 2006/4/7)

年次	調査の概要	データアウト及び付随データ	レコード数	【参考】項目名一覧
平成元(1989)年	【調査概要】 【調査対象】 【調査方法】	【普通世帯】 【単身世帯】 【二人以上世帯】	約4,420万レコード	CSV形式
平成6(1994)年	【調査概要】 【調査対象】 【調査方法】	【一般世帯】 【単身世帯】 【二人以上世帯】	約4,970万レコード	CSV形式
平成11(1999)年	【調査概要】 【調査対象】 【調査方法】	【一般世帯】 【単身世帯】 【二人以上世帯】	約4,970万レコード	CSV形式
平成16(2004)年	【調査概要】 【調査対象】 【調査方法】	【二人以上世帯】 【単身世帯】 【二人以上世帯】	約4,470万レコード	CSV形式

- 提供するファイルの構成区分は、調査年度別となります。
- 提供するファイルの構成区分は、1年毎に調査区分の別となります。
- 符号表は、「匿名統計処理マニュアル(統計法第10条)」平成14年3月31日付省令で改定された責任者(CID)識別番号表に基づいて作成されています。
- データ形式は「CSV」、文字コード(エンコード)は「JIS」です。
- 項目名一覧は、符号表から項目名のみを抜き出したものです。お使いになるプログラム等の仕様に合わせて適宜修正の上、ご利用ください。

※ 匿名データの提供は無料で行っております。詳細は「匿名データの活用」をご覧ください。

http://www.stat.go.jp/services/anonymity-zensho.html 9

2 匿名データの利用目的等

(1) 直接の利用目的の区分

■ 学術研究

① 学術研究の名称
学術研究の名称に関する研究

② 学術研究の必要性
1990年代のバブル崩壊後、我が国では、パート・アルバイトや派遣社員の増加に見られるように雇用形態の非正規化が進行している。特に、若年層においては非正規の比率が年々増加しており、これら若年層に対する雇用対策は喫緊の課題となっている。若年層の就業状況は、若年層がおかれた社会的経済的状況によって多様な変遷を遂げていることから、匿名データを用いて実証的に把握することが必要である。よって、本研究では、若年労働市場に焦点を当て、①に於ける学術研究を行う。

③ 学術研究の目的、利用する方法及び期待する統計データの種別
本研究では、若年層の雇用状況と就業形態の動向を明らかにし、若年層の就業に対する効果的な政策を立案するため、次のような実証分析を行う。
【分析1】 社会的経済的状況によって、若年層の雇用状況がどのように変化しているかを明らかにする。
【分析2】 性別、年齢、学歴といった個人の社会的属性や、就職活動期間や就業履歴等の就業に関する属性を説明変数とした回帰分析をのこることにより、若年層の就業状況に影響を及ぼす社会的経済的状況を明らかにする。

④ 学術研究の期待する統計データの提供期間
統計データの作成 平成21年10月～平成22年3月
統計データの提供 平成22年1月～平成22年3月

(2) その他の利用目的

①
②
③
④
⑤

※ (1)及び(3)に記載した利用目的のすべてが利用目的を記入する。

(3) 成果の公表方法
論文 投稿の方法: ○○大学○○研究所研究報告 予定時期 23年 3月)
報告書 公表の方法: ○○大学○○研究所研究報告 予定時期 22年 3月)
学会・研究会等で発表 (学会、研究会等の名称: 日本△△学会 予定時期 22年 6月)
2010年度統計調査学会連合大会 予定時期 22年 9月)
学会誌等に掲載 (学会誌等の名称: △△会報 予定時期 22年 9月)
その他 具体的な公表方法:)

※ 学術研究の目的等については、統計センターのホームページへ掲載を希望する場合は、具体的な公表の方法に「統計センターのホームページに掲載」と記入

3 匿名データの提供希望年月日 平成21年10月 1日

【記入例】 個人の出出、学術研究目的の出出の場合

様式第1-1号(個人が出出を行う場合) 匿名データの受領まで約1ヶ月間を見込んで、余裕を持って申し出ます。

匿名データの提供依頼申請書 (学術研究目的関係)
平成21年 8月 20日 (最終受領日: 平成 年 月 日)

独立行政法人 統計センター理事長 殿
新規申請時には使用しません。(記載内容に変更が生じた場合の記載事項変更申請書(様式10)の提出時に使用します。)

【提供依頼申請者】
(所属・職名) ○○大学○○学部 准教授
(氏名) 徳島 省太郎
(生年月日) 1970年 〇月 〇日
(自宅住所) 〒000-0000 東京都江東区〇〇1-2-3 統計ハイビル1018
TEL 03-0000-0000
e-mail 〇〇〇〇@〇〇〇〇.〇〇〇.jp

【連絡先所在地】
〒000-0000 東京都江東区〇〇1-2-3 ○○大学○○学部○○研究室
TEL 03-0000-0000
e-mail 〇〇〇〇@〇〇〇〇.ac.jp

【代理人】
(所属・職名) ○○大学○○学部 助教
(氏名) 徳島 省太郎
(生年月日) 1980年 〇月 〇日
(自宅住所) 〒000-0000 東京都江東区〇〇1-2-3
TEL 03-0000-0000
e-mail 〇〇〇〇@〇〇〇〇.〇〇〇.jp

【管理先所在地】
〒000-0000 東京都江東区〇〇1-2-3 ○○大学○○学部○○研究室
TEL 03-0000-0000
e-mail 〇〇〇〇@〇〇〇〇.ac.jp

統計法第10条の規定に基づき、下記のとおり申請をします。

1 匿名データの名称及び年次等並びにファイル数	総務省	(年次等)	(ファイル数)
○基本調査		平成4年、9年、14年	各2

利用者が複数の観点に分かれて利用する場合、すべての利用場を記載するとともに、それぞれの匿名データの保有責任者について、名前、自宅住所及び連絡先(住所、電話番号、eメールアドレス)の情報を本申請書の「11 その他必要な事項」に記載する必要があります。

4 匿名データの提供依頼申請書

1 提供する匿名データの名称及び年次等並びにファイル数

2 利用場所、保管場所及び管理方法

(利用場所、保管場所)
利用場所: ○○大学○○学部○○研究室 (第〇号館第〇号室)
保管場所: 同研究室内のキャビネット
利用場所: △△大学△△学部△△研究室 (第△号館第△号室)
保管場所: 同研究室内のキャビネット

(管理方法)
① 匿名データを利用する場所(匿名データファイルの保管を含む)は、距離可能な適切な場所に限定される。
② 匿名データは、限定された媒体に格納され、当該限定された媒体が可能な限りネットワーク等で保護される。
③ 匿名データの利用時に匿名データの利用場所が存在する者が制限される。又は、何らかの制限がなされる。
④ 匿名データの利用時の情報システムとの連携として、インターネット等の外部ネットワークに接続した状態としない。
⑤ 匿名データを使用する情報システムは、コンピュータウイルス対策、セキュリティ対策、物理的対策、アクセス制御対策、バックアップ対策等の不測事態対策が施されている。
⑥ 外部ネットワークに接続可能な電子計算機や利用者以外者が使用可能な電子計算機に匿名データ及び中間生成物を保管しない措置をとる。また、利用者以外者が匿名データ及び中間生成物を保管している電子計算機にアクセスできない措置をとる。
⑦ 提供される匿名データに加え、統計業務等によって生成される匿名データを含む中間生成物及び最終物についても、漏えい等事故を防止するために適切な管理が行われる。
⑧ その他()

3 提供期間 平成23年 3月 31日 まで

6 匿名データを取扱う者の氏名、所属、職名、利用期間

氏名	所属	職名等	利用場所
徳島 省太郎	〇〇大学○○学部	准教授	〇〇研究室
統計 徳太郎	〇〇大学○○学部	助教	〇〇研究室
〇〇 〇〇	〇〇大学○○学部	講師	〇〇研究室
△△ 利洋	△△大学△△学部	准教授	△△研究室
△△ △△	△△大学△△学部	大学院生	△△研究室

7 提供を受ける、又は今後提供を依頼する予定がある調査票情報及び利用の匿名データ
(別記提供を受けている調査票情報及び利用の匿名データ)
×××××調査票 (平成元年、6年、11年)

8 匿名データの提供の方法等

(1) 提供の方法(媒体) ※ 希望する提供方法を記載する。
 CD-R DVD-R

(2) 送付の希望の有無 ※ 希望する送付方法を記載する。
 直接の受取 郵送による送付