

## 今後の課題・および対応

- ✓ **模擬申出研究の開始**
  - ✓ 分析実働チームの立ち上げ
  - ✓ 東大との情報共有：解析結果についてのレビュー/テレビ会議
  - ✓ 有識者会議への適切な情報提供をどうするか。
- ✓ **利用者向けの教育・啓発**
  - ✓ 京大以外の利用希望者に対し、どういった建て付けでデータ利用を行って貰うこととするか、予め目処を付けておく必要がある。
  - ✓ 今後の連絡協議会は、研究成果を紹介する場にしてはどうか。第三者の参加はどう考えるか。

2016年3月16日  
第29回有識者会議  
(非公開) 報告を元に作成



## レセプト情報等オンサイトリサーチセンター (京都) における利用経験から

medical informatics

加藤源太 <sup>1)</sup>、岡本和也 <sup>2)</sup>、黒田知宏 <sup>2)</sup>

1) 京都大学医学部附属病院 診療報酬センター

2) 京都大学医学部附属病院 医療情報企画部



### 本日お話しする内容

- 1 : 運用体制の確立 (加藤)
- 2 : センター利用体験の紹介 (岡本)
- 3 : 改善が必要と思われる事項 (黒田)



## 1 : 運用体制の確立

### ■ 京都大学におけるオンサイトリサーチセンター開設 および試行的利用開始までの経緯

- 2014.10.10  
連携協力に関する基本協定書が締結される。
- 2015.02.19  
京都大学および厚生労働省、東京大学とで、連絡協議会開催。  
(以後、定期的に開催)
- 2015.03.25  
オンサイトリサーチセンター（京都） 検収実施。
- 2016.02.04  
レセプト情報等オンサイトリサーチセンター（京都）運用部の設置が、  
病院協議会（2月4日）に承認され、正規の組織として位置づけられる。
- 2016.02.17  
オンサイトリサーチセンター（京都）の試行的利用が開始された。

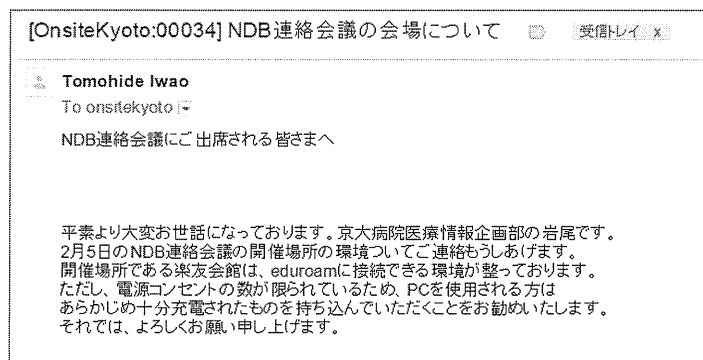


## 1 : 運用体制の確立

### ■ メールリスト「onsitekyoto」の開設

- メールリストを開設し、京大の利用者間での  
情報共有の場を立ち上げた。

(メールリスト運用の実例)





## 1 : 運用体制の確立

### ■ レセプト情報等オンサイトリサーチセンター運用部（京都）の設置

- センターの運用を担う組織として、正式に「レセプト情報等オンサイトリサーチセンター運用部（京都）」を設置し、病院医療情報企画部・医学研究科社会健康医学系専攻の四分野を中心にセンターの運用を行うこと明記するとともに、具体的なセンターの運用方針を決定する場として位置づけた。

京都大学医学部附属病院レセプト情報等オンサイトリサーチセンター（京都）運用部内規

（平成28年2月4日制定）

（設置）

第1条 この内規は、京都大学医学部附属病院規程（昭和41年通示第18号）第11条第2項の規定に基づき、レセプト情報等オンサイトリサーチセンター（京都）運用部（以下「運用部」という。）に関し必要な事項を定めるものとする。

第2条 運用部は、国立大学法人京都大学と厚生労働省との間における連携協力に関する基本協定書（以下「協定」という。）に基づき京都大学医学部附属病院内に設置されるレセプト情報等オンサイトリサーチセンター（京都）の運営に従事する。



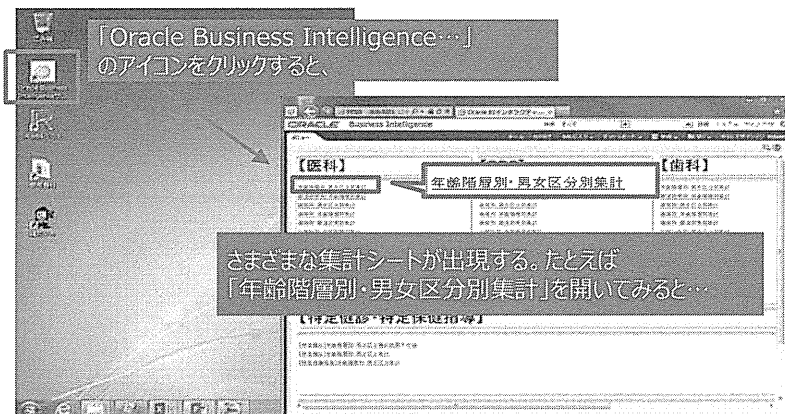
## 2 : センター利用体験の紹介（システム）

### ■ 利用方法は、大きく2通りに分けられる

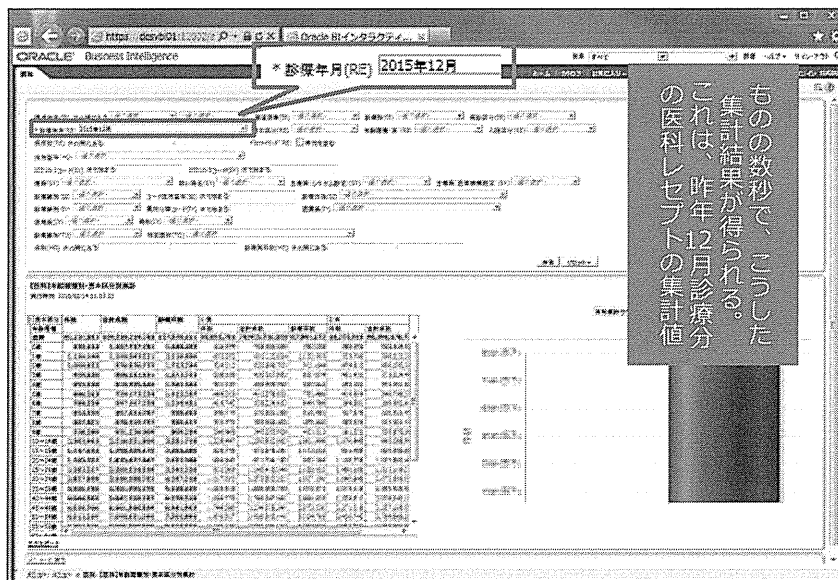
- BI ツールを用いたデータ抽出・分析（本日説明）
  - プルダウンメニューを用いた簡易集計ツールとSQL コードを記述してデータを抽出するツールの2種類の抽出・分析方法がある
- Oracle R Enterpriseを用いたデータ抽出・分析
  - 京都大学では利用実績がまだないため、今回は言及しない。

## 2 : センター利用体験の紹介 (システム)

- プルダウンメニューを用いた簡易集計ツール
- SQLの知識の無い研究者向け



## 2 : センター利用体験の紹介 (システム)



## 2 : センター利用体験の紹介 (システム)

たとえば、入院と入院外を区別して集計したいときはここから条件を指定

## 2 : センター利用体験の紹介 (システム)

薬剤などの場合は...

複数選択することも可能

## 2 : センター利用体験の紹介 (意見)

### ■ プルダウンメニューを用いた簡易集計ツール

【利点】速度は早い。

【欠点】実際に行われている処理が見えない。



(提案) 各種処理がどのようなロジックで行われているのかについての解説書の整備が必要ではないか。

## 2 : センター利用体験の紹介 (意見)

【欠点】集計単位がレセプトの行のみである。

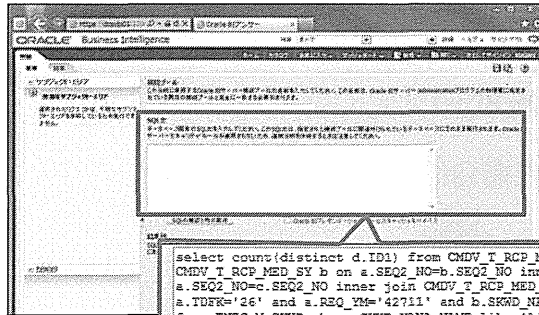
- 例えば、「2015年にワーファリンを処方された患者数」はわからない。
- 簡易集計ツールの集計結果では、一患者が月ごとにカウントされ、何倍もの値になる。



(提案) 被保険者証等記号・番号などから生成されるNDBのID1での集計ができるようにする。

## 2 : センター利用体験の紹介 (システム)

- SQLを記述してデータを抽出するツール
  - SQLの知識の有る研究者向け



SQLコマンドを書いて  
データ抽出できる  
ツールもあります。

```
select count(distinct d.ID1) from CMDV_T_RCP_MED_IR a inner join  
CMDV_T_RCP_MED_SY b on a.SEQ2_NO=b.SEQ2_NO inner join CMDV_T_RCP_MED_IY c on  
a.SEQ2_NO=c.SEQ2_NO inner join CMDV_T_RCP_MED_RE d on a.SEQ2_NO=d.SEQ2_NO where  
a.TDFK='26' and a.REQ_YM='42711' and b.SKWD_NAME_CD in (select SKWD_NAME_CD  
from TMS_M_SKWD where SKWD_KANA_NAME like '%ハイガン%') and c.MEDICINE_CD in  
(select MEDICINE_CD from TMS_M_MEDICINE where MSN_KANJ like '%アバチン%' or  
(MSN_KANJ like '%シスプラチン%' or (MSN_KANJ like '%タキソール%'))
```

「医科レセプトで」、「京都府の医療機関を」、「平成27年11月に受診されて」、「『ハイガン』という呼称を含む傷病名を有し」、「アバチン」 or 「シスプラチン」 or 「タキソール」を使用している患者ID数を抽出するコマンドになります。これを確認すると、

## 2 : センター利用体験の紹介 (システム)

The screenshot shows the Oracle Business Intelligence web interface. The top navigation bar includes 'Oracle Business Intelligence'. The main content area is divided into a 'メッセージ' (Message) section on the left and a '結果' (Results) section on the right. The Message section contains a warning icon and the text: '無効なサブジェクト-エリア' (Invalid subject area). Below this, it says: '選択されたリクエストは、不明なサブジェクト-エリアを参照しているため実行できません。' (The selected request cannot be executed because it references an unknown subject area). A box below the message says: '4-5分程度で数字が出ました。' (The number appeared within 4-5 minutes). The Results section shows a table with one row containing the value '110'.



## 2 : センター利用体験の紹介 (意見)

### ■ SQL を記述してデータを抽出するツール

【利点】速度はまあまあ。

- 処理が重そうな SQL でも数分程度。

【欠点】SQL のエラー情報が乏しい。

- SQL の構文を間違った場合に、修正に時間がかかる。



(提案) Native の SQL ツール  
(Oracle SQL\*Plus や Oracle SQL Developer)  
を使えるようにする。

## 2 : センター利用体験の紹介 (意見)

- SQLコードの記載を間違えるとエラーメッセージが出るが、その情報が非常に乏しく、「どこでどう間違えたか?」は利用者各自が目視で確認するしかない。

```
SQL文
データベース固有のSQL文を入力してください。このSQL文は、指定された接続プールに関連付けられているデータベースにそのまま発行されます。Oracle BI
サードパーティツールが適用されないため、直接分析を作成するときは注意してください。

select count(distinct d.ID1) from CMDV_T_RCP_MED_IR a inner join
CMDV_T_RCP_MED_SY b on a.SEQ2_NO=b.SEQ2_NO inner join CMDV_T_RCP_MED_IY c on
a.SEQ2_NO=c.SEQ2_NO inner join CMDV_T_RCP_MED_RE d on a.SEQ2_NO=d.SEQ2_NO where
a.TDFK='2E' and a.REQ_YM='2711' and b.SKWD_NAME_CD in ((select SKWD_NAME_CD
from THDS_M_SKWD where SKWD_KANA_NAME like '%ハイワン*') and c.MEDICINE_CD in
(select MEDICINE_CD from THDS_M_MEDICINE where (MSH_KANJ like '%アバズチン*') or
(MSH_KANJ like '%スプラチン*') or (MSH_KANJ like '%タキソール*')));

SQLの確認と列の取得
Oracle BIプレゼンテーションサービスキャッシュをバイパス

ODBCドライバによってエラー (SQLEjectDirect) が発生しました。
状態: HYDBA コード: 10056, [Microsoft] [SQL STATE HYDBA] [OracleError: 10056] - 致命的エラーが発生しました。 [OracleError: 43112] OBI5秒未満で
行の数が不足しています。 [OracleError: 43093] EXECUTE PLAN SQL文の処理中にエラーが発生しました。このエラーの詳細は、システム管理者にログの取組を
依頼してください。 (BY000)
このエラーの詳細は、システム管理者にログの取組を依頼してください。
```

- ちなみに、正解はこちらになります。違いがわかりますか？

```
SQL文
データベース固有のSQL文を入力してください。このSQL文は、指定された接続プールに関連付けられているデータベース
サードパーティツールが適用されないため、直接分析を作成するときは注意してください。

select count(distinct d.ID1) from CMDV_T_RCP_MED_IR a inner join
CMDV_T_RCP_MED_SY b on a.SEQ2_NO=b.SEQ2_NO inner join CMDV_T_RCP_MED_IY c on
a.SEQ2_NO=c.SEQ2_NO inner join CMDV_T_RCP_MED_RE d on a.SEQ2_NO=d.SEQ2_NO where
a.TDFK='2E' and a.REQ_YM='2711' and b.SKWD_NAME_CD in (select SKWD_NAME_CD
from THDS_M_SKWD where SKWD_KANA_NAME like '%ハイワン*') and c.MEDICINE_CD in
(select MEDICINE_CD from THDS_M_MEDICINE where (MSH_KANJ like '%アバズチン*') or
(MSH_KANJ like '%スプラチン*') or (MSH_KANJ like '%タキソール*')));
```

## 2 : センター利用体験の紹介 (意見)

不正解

SQL文

データベース固有のSQL文を入力してください。このSQL文は、指定された接続プールに関連付けられているデータベースサーバーセキュリティルールが適用されないため、直接分析を作成するときは注意してください。

```
select count(distinct d.ID1) from CMDV_T_RCP_MED_IR a inner join
CMDV_T_RCP_MED_SY b on a.SEQ2_NO=b.SEQ2_NO inner join CMDV_T_RCP_MED_IY c on
a.SEQ2_NO=c.SEQ2_NO inner join CMDV_T_RCP_MED_RE d on a.SEQ2_NO=d.SEQ2_NO where
a.TDFK='26' and a.REQ_YM='42711' and b.SKWD_NAME_CD in ((select SKWD_NAME_CD
from TNDS_M_SKWD where SKWD_KANA_NAME like '%ハイガン%') and c.MEDICINE_CD in
(select MEDICINE_CD from TNDS_M_MEDICINE where (MSN_KANJ like '%アバステン%') or
(MSN_KANJ like '%シスプラチン%') or (MSN_KANJ like '%タキソール%'))
```

- 余分な括弧がひとつ入っただけで、作業が先に進まなくなる。
- エラーメッセージが出るたびに現在のオンサイトリサーチセンターではスクリプトを目視しなければならない。一方、近年のソフトではエラー修正に際しての支援機能が充実してきており、こうした作業は必ずしも発生しない。

正解

SQL文

データベース固有のSQL文を入力してください。このSQL文は、指定された接続プールに関連付けられているデータベースサーバーセキュリティルールが適用されないため、直接分析を作成するときは注意してください。

```
select count(distinct d.ID1) from CMDV_T_RCP_MED_IR a inner join
CMDV_T_RCP_MED_SY b on a.SEQ2_NO=b.SEQ2_NO inner join CMDV_T_RCP_MED_IY c on
a.SEQ2_NO=c.SEQ2_NO inner join CMDV_T_RCP_MED_RE d on a.SEQ2_NO=d.SEQ2_NO where
a.TDFK='26' and a.REQ_YM='42711' and b.SKWD_NAME_CD in ((select SKWD_NAME_CD
from TNDS_M_SKWD where SKWD_KANA_NAME like '%ハイガン%') and c.MEDICINE_CD in
(select MEDICINE_CD from TNDS_M_MEDICINE where (MSN_KANJ like '%アバステン%') or
(MSN_KANJ like '%シスプラチン%') or (MSN_KANJ like '%タキソール%'))
```

## 2 : センター利用体験の紹介 (意見)

### ■ 各種ツールを用いて得られた集計結果

【欠点】得られた集計結果の解釈が難しい

○○科には※※病の人っていないのかなあ？

△△科で試してみたらどう？

う〜ん、やっぱりないみたい。

ところで、診療科情報って信頼できたかな？

.....

	※※病該当数
○○科	0
△△科	0
××科	0
○■科	0
...	

「利用者各自の努力」で解決するしかないのか？

## 2 : センター利用体験の紹介

### ■ 当初想定のパフォーマンステストを、早急に実施したいところだが…

- 出てきた値が正しいかどうかわからない
- 出力値の検証に、複数人の研究者がとりかかり、数時間かけて「正しそうかどうか」を見極め、また出力値の検証にとりかかり…といった作業をつづけているところ

そのため、パフォーマンステスト以前に、現時点での課題の洗い出し・共有が重要だと感じている。

## 3 : 改善が必要と思われる事項

### ■ 実際の利用を踏まえた提言・意見から

- 各種処理がどのようなロジックで行われているのかについての解説書の整備が必要ではないか。  
→データ利用環境の整備
- 被保険者証等記号・番号などから生成されるNDBのID1での集計ができるようにする。  
→データ構造の見直し
- NativeのSQLツールを使えるようにする。  
→データ利用環境の整備
- 「利用者各自の努力」で解決するしかないのか？  
→利用者のリテラシー向上



### 3 : 改善が必要と思われる事項

#### ■ 教育プログラムの整備

- 1: レセプトのデータ構造についての解説
- 2: BIツールの動きについての解説（事例を含む）

を整備する。



### 3 : 改善が必要と思われる事項

#### ■ 情報環境の見直し

- 1: 利用者へのストレスが少なくなるような各種ソフトの整備  
（エラーメッセージの情報量が多いSQLソフトや  
エディタの導入）
- 2: データを再編して格納できるデータベースソフトの提供
  - 2-1: 各種検索に耐えうるデータマートの整備

### 3 : 改善が必要と思われる事項

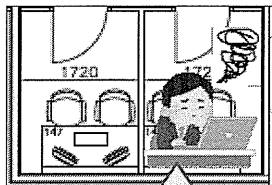
#### ■ 公開演習環境の提供

- 1: 教育マテリアルの提供
- 2: SQL/BIツールテスト実行環境の提供
  - 2-1: ダミーデータを整備
  - 2-2: エラーチェックなどを事前に実行させる

### 3 : 改善が必要と思われる事項

#### ■ 現在のオンサイト利用環境

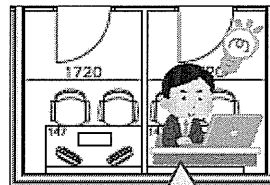
“NDBデータってどんな感じだろう…”  
“事前に出来ることも限られるし、  
行ってみないと分からないなあ…”



“「診療科名」が使えないことがあるって  
知らなかったよ…”  
“ネットが使えないから検索も出来ないし…”  
“ところで、この操作って合ってるのかなあ…”  
“この作業に4日もかかってるよ…”

#### ■ 公開演習環境の提供によって…

“NDBデータの操作は、単に構造を  
知っているだけでは難しいなあ…”  
“ダミーデータで作業して初めて  
データの特徴が見えてくるなあ…”



“事前の想定通りに集計結果が出てくる！”  
“これだけ多くの事例を分析できると、  
同じ結果でも重みが違うなあ…”

厚生労働科学研究費補助金 戦略型研究  
分担研究報告書

NDB 研究に資する、バリデーションスタディのシステマティックレビューに向けて  
—主にクレームデータのバリデーションスタディを対象としたパイロットスタディー—

研究分担者 加藤 源太 京都大学医学部附属病院診療報酬センター 准教授

研究要旨

レセプト情報・特定健診等情報データベース（NDB：National Database of Health Insurance Claims and Specific Health Checkups of Japan）は、国民皆保険の日本において、非常に高い悉皆性を誇るデータであり、政策応用や各種研究等への利活用が期待されている。一方、傷病名情報をはじめとしてレセプトに記載される情報の精度についてバリデーションスタディを行うことの重要性が指摘されているなど、情報の使用に際しては留意が必要である。このため、今後の NDB 研究に資する目的で、主に「レセプト」すなわちクレームデータを対象としたバリデーションスタディのシステマティックレビューを開始した。2015 年 10 月 27 日時点で、PubMed を使用して文献検索を行ったところ、1,164 編の論文が抽出された。これら論文のレビュー過程では、バリデーションスタディの対象となっているデータは、単に "claims data" と記されているものだけでなく、"administrative claims data(base)"、"administrative and claims data(base)" などと記されていることが少なくないことが確認された。これは、国民皆保険下で診療報酬請求の明細書が「レセプト」という特定の単語で表記され、仕様も統一されている日本以外の国のなかには、「診療報酬請求」という概念が必ずしも一義的に定まっているとは限られない可能性があることが示唆された。「クレームデータのバリデーションスタディ」という現象を同定することそれ自体の難しさが明らかにされるとともに、保健医療データの二次利用においては、各国間の医療制度の相違も踏まえ、データの質に対する意識を研究者間で高めていくことの重要性も確認された。

A. 研究目的

レセプト情報・特定健診等情報データベース（NDB：National Database of Health Insurance Claims and Specific Health Checkups of Japan）は、2008 年度以降の特定健診・特定保健指導情報、および 2009 年度診療分以降の電子化されたレセプト情報を格納するヘルスケア領域のビッグデータである。国民皆保険の日本において、非常に高い悉皆性を誇るデータであり、政策応用や各種研

究等への利活用が期待されている<sup>1</sup>。

とはいえ、レセプト情報にはヘルスケアに関連するすべての情報が含まれているわけではなく、例えば、検査値やがんのステージ分類などといった情報は含まれていない。また、レセプトに記載される「保険病名」の精度についても、信頼性の高い傷病名とそれほど高くない傷病名がある可能性も指摘されており<sup>2</sup>、信頼性の高い情報を入手しようと考えれば、記載されている情報を適切

に抽出、分析することが求められる。特に、レセプトが保険医療機関によって保険者に診療報酬が請求される際の情報であることを踏まえれば、行った診療行為や投与した医薬品に関する実施情報と比較して、傷病名情報には「疑い」といった修飾語がかなり含まれていたりしており<sup>3</sup>、情報の使用に際しては留意が必要である。

データの二次利用において、当該データに含まれる情報を信頼性の高い別の情報源と突合して調査することはバリデーションスタディと呼ばれており、医薬品医療機器総合機構による「医療情報のデータベース等を用いた医薬品の安全性評価における薬剤疫学研究の実施に関するガイドライン」では、特に請求書データ、すなわちレセプト情報において傷病名コードのみでアウトカムを定義する場合には、対象になりうる、と記されている<sup>4</sup>。また、データの対象者に連結不可能匿名化等の処理が施されているがゆえに他のデータベースを用いたバリデーションが実施できない場合もあり得るため、そうした場合には他のデータベースで同様の定義に関して実施したバリデーションスタディの結果を参考にすることも提案されている<sup>4</sup>。

一方で、データベースのバリデーションを行う際に、同一データベースの他の情報（例：心筋梗塞事例をバリデーションするため、トロポニン検査、線溶性薬剤による治療、冠動脈治療、あるいは3日以上入院を伴う事例のみを選択する、など<sup>5,6</sup>）を用いてバリデーションを行う場合も、内部バリデーション（internal validation）として位置づけ、他のデータベースの情報を用いて行うバリデーションを外部バリデーション

（external validation）と定義づけて、両者を区別することもある<sup>6</sup>。

ただ、バリデーションスタディに対するシステマティックレビューは、例えばイギリスのプライマリケアデータベース（the General Practice Research Database; GPRD）を使用した研究に対するもの<sup>6</sup>などにみられるものの、クレームデータに対するこれまでのバリデーションスタディがどのように行われているのか、その概要は必ずしも十分には明らかにされてはこなかった。どういった疾患がバリデーションの対象になることが多いのか、どういったデータあるいはレコード項目がバリデーションの対象となり、どういったデータあるいはレコード項目がバリデーションの評価基準として使用されることが多いのか、どういった指標によってアウトカムが妥当と判定されるのか、内部バリデーションはどのくらいの頻度で行われるのか、等々、バリデーションスタディの動向などについて、必ずしも研究者間で認識が共有されているとは言えない状況にある。こうした背景を踏まえ、本研究ではクレームデータに対するバリデーションスタディが実際にどのように行われているのかについてシステマティックレビューを行い、現状に対する評価を行うこととする。ただし、当研究は報告書執筆時点で該当文献の抽出および一部文献のアブストラクトのレビューといった、パイロットスタディの段階にとどまっていることから、本稿では現時点での研究進捗を報告するとともに、次年度以降の研究に向けた方針の概要を紹介する。これにより、バリデーションスタディの概要を明らかにするとともに、NDBを用いた研究において、得られたデー

タに対しどのようにして適切なバリデーションを行っていくのかについての方針を確立させることを、本研究の目的としている。

## B. 研究方法

2015年10月27日時点で、PubMedを使用して、`valid*[tiab] AND (receipt*[tiab] OR insurance[tiab] OR medicare[tiab] OR medicaid[tiab] OR claim*[tiab] OR administr*[tiab]) AND (database[tiab] OR electronic health records[mh]) AND English[la] NOT letter[pt]` の条件で文献検索を行った。もとより対象として想定するのはクレームデータであるが、論文抽出の段階では、該当論文や対象データベースによってはクレームデータとその他の医療データとの区分けが必ずしも容易につけられないと想定されたため、「`electronic health record`」「`administr`」等の条件も付与することとした。ここで得られた論文を対象論文とするとともに、米国食品医薬品局(FDA)によるミニセンチネル計画の一環として行われたバリデーションスタディ論文なども参照しながら、検討を行った。

なお、当研究は既刊論文を対象とした研究であり、不利益や危険性、あるいは人権擁護上、特段の配慮が必要となるような対象者が存在しないと考えられるため、倫理面での問題が発生しないものと判断された。

## C. 結果

PubMedによる検索で得られた文献数は合計1,164編であった。先に述べたFDAによるバリデーションスタディ論文なども参考にしながら現在システムティックレビューを実施しているところであり、本稿では、

レビューの過程で得られた現時点でのいくつかの知見について紹介を行うこととする。ただし、本稿は抽出した論文に対して網羅的な評価を行ったうえでの報告ではないため、以下の知見は今後のシステムティックレビューに向けたパイロットスタディという位置付けのものであることをここに記しておく。

### 【対象とするデータに関する問題】

NDBデータは、レセプト情報と同一の情報が含まれているわけではなく、一部加工が施されている。とはいえ、その処理は患者個人の特定に繋がるような一部情報(患者氏名、生年月日の「日」、保険医療機関の所在地及び名称、カルテ番号等、被保険者証等記号・番号など)の削除、および同一患者を特定するためのハッシュIDの付与に留まっており、その他のレセプトデータは、NDBデータとして格納されている<sup>8</sup>。このため、NDBデータにはレセプト情報のほとんどが含まれているとともに、その他の情報は加工後のハッシュID等を除き、含まれていない。

一方、海外の文献においては、データベースについての表記は“claims data”となっているものばかりではなく、“administrative and claims data(base)”、“administrative claims data(base)”などと表記されている場合が少なくなかった。“administrative data”は、請求業務に直接関係しないデータであり、医療機関等が患者管理に用いるデータであることから、「医事データ」、「病院データ」、「病院管理データ」などといった呼称が該当するのだろうが、適切な日本語の同定が困難である。



#### 【データ項目に関する問題】

保険制度の相違等の理由から、他国のクレイムデータに含まれていて、日本のレセプトに含まれていないデータ項目が一定数存在する。一例をあげれば、米国の例となるが、ワクチン接種情報は米国のクレイムデータに含まれているものの、日本のクレイムデータには含まれていない。例えば、米国におけるワクチン接種情報についてのバリデーションスタディが行われている場合<sup>7</sup>、クレイムデータのバリデーションスタディであることは間違いないものの、日本のレセプトデータの項目には含まれていないため、NDB 研究に直接資する研究として位置づけて知見を活用することは難しいと思われる。

#### 【日本のデータベースを扱っている論文】

合計で 1,164 編の論文が抽出されたが、そのうち、日本の administrative data あるいは claims data を対象としている論文は 5 編であったが、これらはすべて DPC データが対象となっていた。また、これとは別に日本の特定の医療機関から集めたデータを用いた論文は 2 編、民間の保険者のデータを使用した論文は 1 編であった。

#### D. 考察

米国を例にとると、クレイムデータのバリデーションスタディで対象となっているデータの確保については、米国政府の Department of Health and Human Services のエージェンシーである Center for Medicare and Medicaid Services (CMS)が行っているメディケア・メディケイドのデータ提供、いわ

ば公的機関が関与するデータ提供<sup>9</sup>の枠組みに依拠する場合と、例えば MarketScan や PharMetrics といった民間のヘルスケア産業が提供するデータを使用する場合<sup>10</sup>とが認められる。ただ、いずれの場合も、提供に際してデータに対しては研究者のニーズに応じて何らかの加工が加えられているものとみられる。

これは日本においても同様である。厚生労働省から NDB データが申出される際には、研究者の要望にあわせて必要項目を抽出した上で提供される特別抽出や、一ヶ月分のデータから一定の割合でランダムに事例を抽出し、希少性の高い情報を匿名化する等の処理を施したサンプリングデータセットなど、データ提供にあたっては研究者のニーズを踏まえたデータ加工が行われている<sup>11</sup>。同様に、民間業者が研究支援の目的で第三者に提供しているレセプトデータにおいても、患者毎に個別の ID を付与してその ID を軸にデータベース化されているなど、研究者が使用しやすい体裁に加工されており<sup>12</sup>、「レセプト」そのままの情報を利用者が分析する、ということではないのは同じである。

しかし、英語でいうところの“claim data”とは、医療費請求のデータに限らず「請求データ」として認識されている一方で、日本では国民皆保険制度が定着しており、医療保険の請求に用いられる請求書に対しては広く「レセプト」という特定の呼称が定着している。また、現時点で NDB は他データと照合を行うことは認められておらず<sup>13</sup>、利用者向けに提供される NDB データは、「レセプト」が含む情報を超える情報を含んでいない。

このように、請求データが他の領域でも使用される一般的な“claims data”という呼称で認識されている米国と、診療報酬の請求という特定の現象を全国的に統一された仕様に基いて請求する書類に対し、やはり特定の「レセプト」という呼称で認識されている日本とでは、データが指し示す数値の意味や果たしている役割、データに対する研究者の認識、さらにはデータの位置づけ自体も、全く同一であると見做すことはできない。それは、こうしたデータを産み出している医療制度それ自体が各国ごとに大きく異なっており、データ自体もそうした違いに大きく影響されているからである。したがって、自然科学領域の研究のように世界中で同じ文脈のもとで認識されている現象を扱うことと、各国の医療制度の相違に基いた「クレイムデータ」に対する「バリデーションスタディ」という現象を扱うことは一概に同列に扱うことができない、ということには留意すべきである。

とはいえ、このことは「クレイムデータのバリデーションスタディ」、もっといえば「クレイムデータを扱った研究」が、各国の医療制度に応じた多様性という障壁を有しているため、国際比較や多国間研究を行うことが不可能である、ということの意味するわけではない。大事なことは、データを扱う研究者が、クレイムデータが内包する各国の医療制度に基いた多様性を十分に認識し、そうしたデータの性質を踏まえたうえでバリデーションスタディのレビュー、さらにはバリデーション全般に臨む必要がある、ということである。また、こうした多様性はクレイムデータ、すなわち「レセプト」に限定されるものではない。研究目的で

収集されたコホートデータなどとは異なり、医事業務や請求業務で収集されたデータを二次利用する場合には、各国の医療制度に基づくデータの特徴を十分に踏まえて臨むとともに、自らが取り扱うデータの特徴を十分に留意し、公表物においてその特徴を適切に言及したうえで成果の公表を行う必要があるといえるだろう。今後、保健医療データの二次利用の進展<sup>14</sup>が世界規模でみられることとなれば、得られた知見についての国際比較や、さらには多国間での共同研究の試みなども出現してくるとも思われる。そうした場合に、単にデータのレコード項目を機械的に比較するだけでは十分ではなく、折に触れ、各国の医療制度に踏み込んでデータの格納過程やデータの質について検証を行っていくという姿勢を研究者間で共有することが、得られる知見の精度を向上させるためには重要ではないだろうか。

## E. 結論

今後のNDB研究に資するべく、主にクレイムデータを対象としたバリデーションスタディのシステマティックレビューに向けて、研究を開始したところである。関連するキーワードからPUBMEDで検索を行ったところ、2015年10月27日時点で1,164編の論文が抽出された。これらの論文では、対象となるデータは“claims data”と記されている場合ばかりでなく、“administrative claims data(base)”, “administrative and claims data(base)”などと記載されていることがあり、「クレイムデータのバリデーションスタディ」という現象を同定することそれ自体の難しさが明らかになった。同時に、保健医療データの二次利用においては、各国間の

医療制度の相違も踏まえ、データの質に対する意識を研究者間で高めていくことの重要性も確認された。

#### F. 健康危険情報

なし

#### G. 研究発表

##### 1. 論文発表

1. 黒田知宏, 田村 寛, 加藤源太, 糸 直人, 荒巻英治, 岡本和也, 小林慎治:京都大学デザインスクール テキストシリーズ 1, デザイン学概論(共立出版株式会社、編集:石田 亨) 2016. (2016年4月10日初版第1刷発行)

##### 2. 学会発表等

1. Genta KATO, Shusuke HIRAGI, Tomohide IWAO, Kazuya OKAMOTO, Hisashi SAITO, Hiroshi TAMURA, Tomohiro KURODA. An introduction of the database of health insurance claims in Japan. In: The 1st Asian Researcher Symposium 2016 Asian Role in Sustainable World Development: 2016.04.25: West Java, Indonesia.
2. 本間健太郎, 石田洋平, 寺前利治, 中村正次, 加藤源太, 田村 寛。手術時の医薬品請求における手術部スタッフの業務軽減と請求漏れ防止対策の試み。平成 27 年度 大学病院情報マネジメント部門連絡会議 抄録集,pp212-215, ポスター, 2018.1.28,長崎市
3. 村野博信, 加藤源太, 田村 寛, 黒田知宏, 平家俊男。病院運営企画室における経営改善の取り組み～教員と事務職員協働で発信する経営改善～。平成

27 年度 大学病院情報マネジメント部門連絡会議 抄録集,pp216-217,ポスター, 2018.1.28,長崎市

4. Genta KATO, Use of IT Solutions in Japan. Regional Meeting on maximizing the Potential of Health Insurance Data Systems to Support Universal Health Coverage, WHO Regional Office for Western Pacific, 2015.12.01, Manila.
5. Genta KATO, Contents and Scope of Data Captured by Ministry in Japan. Regional Meeting on maximizing the Potential of Health Insurance Data Systems to Support Universal Health Coverage, WHO Regional Office for Western Pacific, 2015.12.01, Manila.
6. 加藤源太、レセプト情報・特定健診等情報データベース (NDB) の円滑な利活用を目指すにはーその課題と可能性についてー。日本医療情報学会関西支部、第 2 回講演会、発表日: 2015 年 11 月 21 日、常翔学園大阪センター
7. 岡本和也、森由希子、加藤源太、黒田知宏、胃がん患者の治療実態調査のためのレセプトデータ再構築。第 35 回医療情報学連合大会、2015 年 11 月 3 日、沖縄コンベンションセンター
8. 赤羽根直樹、加藤源太、レセプト情報等オンサイトリサーチセンターの今後の方向性について。第 35 回医療情報学連合大会、2015 年 11 月 2 日、沖縄コンベンションセンター
9. 加藤源太、米国におけるレセプトデータ利活用の支援についてーResDAC の取り組みからー。第 35 回医療情報学連合大会、2015 年 11 月 2 日、沖縄コンベ

- ンションセンター
10. 加藤源太、黒田知宏、レセプト情報等オンラインリサーチセンターの設置・運用に係るこれまでの経緯等について。第 35 回医療情報学連合大会、2015 年 11 月 2 日、沖縄コンベンションセンター
  11. 岡本和也、森由希子、加藤源太、黒田知宏、武藤学、胃癌症例を対象としたレセプト情報解析のためのデータマートの構築。日本癌治療学会 第 53 回学術集会、2015 年 10 月 30 日、国立京都国際会館
  12. 武藤学、森由希子、金井雅史、松本繁巳、佐久間朋寛、小柳智義、岡本和也、加藤源太、黒田知宏、森田智視、小杉眞司、多門啓子、種石慶、奥野恭史、高折晃史、京大病院がんセンターにおけるクリニカルバイオバンク、クリニカルシーケンス、ビッグデータ解析の取組。日本癌治療学会 第 53 回学術集会、2015 年 10 月 29 日、国立京都国際会館
  13. 加藤源太、三宅康史、小池薫、吉村健佑、平野景子、レセプトデータの救急医療領域への利用可能性について—熱中症診療ガイドラインにおけるレセプト情報等の活用事例報告—。日本救急医学会 第 43 回総会・学術集会、2015 年 10 月 21 日、東京国際フォーラム
  14. 加藤 源太、増井 俊彦、八木 真太郎、西井 美歩、田村 寛、海道 利実、岡島 英明、黒田 知宏、平家 俊男、上本 伸二。持続可能な移植医療をいかにして目指していくか—京大病院の移植事例における査定動向を手がかりに—。第 51 回日本移植学会総会 2015 年 10 月 3 日 (土) ホテル日航熊本
  15. 平木 秀輔、田村 寛、加藤 源太、後藤 励、岡本 和也、鈴間 潔、吉村 長久、黒田 知宏。レセプト情報・特定健診等情報 (NDB) サンプルングデータセットを用いた糖尿病網膜症などへのレーザー治療の実態調査。第 28 回日本レーザー医学会関西地方会 2015 年 7 月 25 日(土) TKP ガーデンシティ京都
  16. 森由希子、岡本和也、加藤源太、杉野剛史、首藤真義、黒田知宏、武藤学、レセプト情報を用いた本邦における胃がん治療の実態調査。日本臨床腫瘍学会 第 13 回学術集会、2015 年 7 月 17 日、ロイトン札幌
  17. 武藤学、采野優、西川佳孝、森由希子、金井雅史、松本繁巳、加藤源太、黒田知宏、森田智視、奥野恭史、京大病院がんセンターにおけるバイオバンクとバイオインフォマティクスプロジェクト。日本臨床腫瘍学会 第 13 回学術集会、2015 年 7 月 16 日、ロイトン札幌
  18. Genta Kato, An Introduction to the Universal Healthcare System in Japan、Medical Grand Round at Memorial Hospital of Rhode island. 2015.6.10. Memorial Hospital of Rhode island, Providence, USA.
  19. Genta Kato, Hiroshi Tamura, Rei Goto, Kazuya Okamoto, Kazuki Yoshida, Shusuke Hiragi, Tomohiro Kuroda. An Introduction of the Database of Health Insurance Claims and Health Checkups of Japan.In: Academyhealth 2015 Annual Research Meeting: 2015.06.13: Minneapolis, USA.