

## レセプト情報・特定健診等情報データベース（NDB） ：日本の糖尿病診療の全体像

2017.08.18  
奈良県立医科大学公衆衛生学講座大学院医学研究科博士課程2年  
奈良県立医科大学糖尿病学講座

西岡 祐一

## 公衆衛生学サマーセミナーの目的

- 公衆衛生学講座で研究する者が、  
普段従事している研究内容を発表し、  
その成果を共有する事

## 本日の内容

- 自己紹介
- 研究の内容と現時点の成果
- 今後の展望

## 本日の内容

- 自己紹介
- 研究の内容と現時点の成果
- 今後の展望

## 自己紹介

- 公衆衛生学講座大学院博士課程2年  
研究/勉強内容
  - レセプト情報・特定健診等情報データベース（NDB）
  - 医学統計学研究センター「臨床試験特論コース」
  - Data Science Online Course（東京大学グローバル消費インテリジェンス寄付講座）修了
- 奈良県立医科大学糖尿病学講座医員  
仕事/研究内容
  - 糖尿病センター医師として日常診療
  - 2型糖尿病患者のDTR-QOL質問紙、EQ-5D質問紙
  - 糖尿病患者の治療満足度

## 本日の内容

- 自己紹介
- 研究の内容と現時点の成果
- 今後の展望

## 背景①



## 背景②

- 糖尿病患者は全世界で2014年までに4億人以上  
爆発的に増加し続けている
- 発展途上国でも爆発的に増加しており、もはや贅沢病ではない
- 糖尿病そのものや合併症によるQOL低下や経済的損失は計り知れない

### 背景③

- 日本には糖尿病の全国症例登録システムが存在しない
- 糖尿病患者数や治療内容、その地域差は正確に把握できていない
- これらを把握することは臨床的にも政策的にも極めて重要

9

### 目的

レセプト情報・特定健診等情報データベース（NDB）を使用し、日本の糖尿病診療の全体像を明らかにする

- 糖尿病の患者数や治療の組み合わせを病型ごと、地域ごとに把握
- 地域差を詳細に把握し、専門医の数、医療費、入院率や合併症との関連を示す

10

### 方法：糖尿病薬処方患者数を数える

対象：糖尿病名を有し糖尿病薬を処方された2013年度のすべてのレセプト  
※NDB全体の問題点として医療扶助である生活保護は除外されている（ことになっている）

- 複数のファイルやレコードに分置された同一人物のデータを紐付け（紐付けには当研究班で開発したID0を用いた）
- 上記の紐付け（=名寄せ）により糖尿病患者数を人数として把握

11

### 従来の日本の糖尿病患者の推計値

推計時期	調査名/調査機関	推計人数
2012年	国民健康・栄養調査	950万人
2014年	患者調査	316万6,000人
2015年	国際糖尿病連合（IDF）	720万人

12

### 結果

#### 【NDBを用いて集計した糖尿病薬処方患者数】

糖尿病薬のいずれかを2013年度に1回以上処方された患者数 **6,781,514人**

うち、糖尿病注射薬を2013年度に1回以上処方された患者数 **1,350,125人**

2013年度に内服薬と注射薬の両者1回以上処方された患者数 **867,288人**

13

### 内服薬の2013年度の保険診療での処方人数

薬剤の分類	処方患者数（人）
スルホニル尿素薬	3,483,968
速効型インスリン分泌促進薬	473,762
αグルコシダーゼ阻害薬	2,020,212
ビグアナイド系薬	2,331,309
チアゾリジン系薬	1,295,424
DPP-4阻害薬	4,933,123
SGLT-2阻害薬	未発売

※同じ分類の薬剤を2種類処方されていた場合や規格を変更した場合、2人と数えてしまっている

14

### 注射薬の2013年度の保険診療での処方人数

薬剤の種類	処方患者数（人）
超速効型インスリン	609,378
速効型インスリン	267,103
持効型溶解インスリン	742,020
中間型インスリン	56,953
混合型インスリン	437,052
GLP-1受容体作動薬	77,925

15

### 血糖自己測定器加算（1型糖尿病除く）

診療行為	患者数
血糖自己測定器加算（20回以上）（1型糖尿病の患者を除く）	286,260
血糖自己測定器加算（40回以上）（1型糖尿病の患者を除く）	244,959
血糖自己測定器加算（60回以上）（1型糖尿病の患者を除く）	566,202
血糖自己測定器加算（1型糖尿病の患者を除く）	1,097,421

16

## 自己血糖測定器加算（1型糖尿病）

診療行為	患者数
血糖自己測定器加算（20回以上）（1型糖尿病・小児低血糖症等）	26,862
血糖自己測定器加算（40回以上）（1型糖尿病・小児低血糖症等）	25,669
血糖自己測定器加算（60回以上）（1型糖尿病・小児低血糖症等）	51,598
血糖自己測定器加算（80回以上）（1型糖尿病・小児低血糖症等）	41,770
血糖自己測定器加算（100回以上）（1型糖尿病・小児低血糖症等）	17,768
血糖自己測定器加算（120回以上）（1型糖尿病・小児低血糖症等）	43,864
血糖自己測定器加算（1型糖尿病・小児低血糖症等）	207,531

17

## 考察①：総患者数

- 2013年度の糖尿病薬処方患者数は6,781,514人であり、
  - 生活保護の糖尿病患者
  - 受診していない糖尿病患者
  - 食事運動療法のための糖尿病患者
 を数えられていない点を鑑みれば、糖尿病患者は国際糖尿病連合（IDF）の720万人より多いと考えられる

18

## 考察②：糖尿病注射薬

- 対象のうち糖尿病注射薬を処方された人数は1,350,125人
  - 対象のうち自己血糖測定器加算を算定された人数は1,304,952人
- 加算と処方の有無という別々の条件から、精度高く糖尿病注射薬を処方された患者を検出できている

19

## 本日の内容

- 自己紹介
- 研究の内容と現時点の成果
- 今後の展望**

20

## まとめと展望

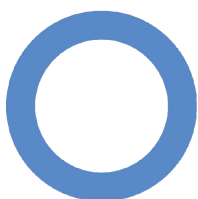
- 日本の糖尿病薬処方患者数は6,781,514人
- NDBを用いて、一定の精度で患者数を把握できる
- 今後は、3年分のデータを用いて糖尿病の患者数や治療の組み合わせをより詳細に分析し、日本の糖尿病診療の全体像を明らかにしていきたい

21

## 以下非表示スライド（参考資料）

22

## 背景



world diabetes day  
14 November

23

## 全保険診療患者 （生活保護除く）

HbA1c年間検査回数	患者数 (ID0)
1	8,666,823
2	3,256,851
3	1,698,118
4	1,313,821
5	975,661
6	1,007,160
7	630,054
8	456,705
9	416,596
10	423,332
11	512,987
12	679,732
13	86,427
14	26,829
15	18,459

糖尿病薬処方患者	HbA1c測定回数		患者数 (ID0)
		1	2
		2	509,178
		3	565,869
		4	625,324
		5	589,657
		6	685,909
		7	456,020
		8	346,668
		9	324,552
		10	335,482
		11	409,790
		12	539,505
		13	69,991
		14	21,733
		15	9,925

注射薬処方患者	測定回数		患者数 (ID0)
		1	2
		2	68,391
		3	71,455
		4	85,057
		5	96,218
		6	131,007
		7	103,463
		8	91,232
		9	91,829
		10	96,948
		11	117,017
		12	158,715
		13	25,605
		14	8,819

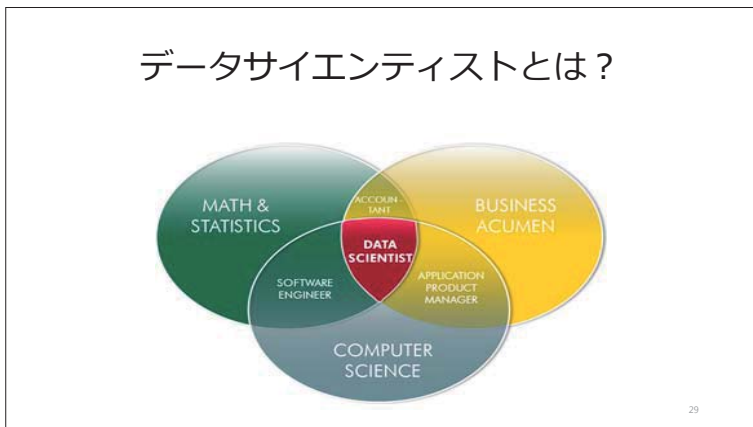
### 項目ごとの2013年度算定患者数

診療行為	患者数
HbA1c	20,329,397
IRI	608,775
CPR	570,895
抗インスリン抗体	36,482
抗GAD抗体	251,416
抗IA-2抗体	768
皮下連続式グルコース測定	7,770
アルブミン定性 (尿)	183,137
アルブミン定量 (尿)	1,038,743
在宅患者訪問栄養食事指導料 (同一建物居住者以外)	27 265

### Data Science Online Course

- 東京大学グローバル消費インテリジェンス寄付講座の育成カリキュラムとして東京大学学生向けに設置され、3年間で延べ300名以上が受講し、約80名の修了者を輩出したデータサイエンス講座のエッセンスを凝縮し、完全自習可能なオンラインコースとして公開
- 60名限定無料公開！

28



### コースの内容紹介

<練習問題 2>  
1から50までの自然数の和を計算するプログラムを書いて、最後の計算結果を表示させるプログラムを書いてください。

```

In [241]: #解答
#普通の方法
n = 0
for x in range(1, 51):
    n += x
print(n)

#sumを使う方法
print(sum(range(1, 51)))

#sum関数を使う方法
print(sum(x for x in range(1, 51)))

1275
1275
1275

```

30

### 医学統計学研究中心 平成 29 年度「臨床試験特論コース」

31

### シラバス

科目名	担当者講師	評価の方法
臨床試験特論	丹後俊郎	平常考査及びレポート

途中で20分間のCoffee Break (2017年2月21日現在)

番号	日時	教科内容	備考
1	4月6日 (木)	オリエンテーション、RCTの原理と統計モデル	2017年
		RCTとは、同時対照の必要性、無作為化の重要性、ヘルシンキ宣言、倫理的問題、エンドポイント、バイアス、等	
2	4月27日 (木)	無作為割り付けの方法/目標症例数の計算 (I)	
		完全無作為化法、重モデル、置換ブロック法、Efron's biased coin デザイン、層別無作為化法、最小化法、症例数の計算 (母平均の差、割り付け数の不均等の影響、漸近的正規近似に基づく公式)	

32

## 初回授業の内容

- 『統計手法なんてどうでも良い、  
臨床試験はデザインですべてが決まる』
- 医師が（医療者が）効果があると信じることの重要性  
信じていない医療行為に関する試験は絶対に行われない！
- 日本の現状～教育を受けていない、専門家もいない～
- 生物統計学者の役割～数理的に表現～

33

## 初回授業の内容

- 『統計手法なんてどうでも良い、  
臨床試験はデザインですべてが決まる』
- 医師が（医療者が）効果があると信じることの重要性  
信じていない医療行為に関する試験は絶対に行われない！
- 日本の現状～教育を受けていない、専門家もいない～
- 生物統計学者の役割～数理的に表現～

34

## 初回授業の内容

- 『統計手法なんてどうでも良い、  
臨床試験はデザインですべてが決まる』
- 医師が（医療者が）効果があると信じることの重要性  
信じていない医療行為に関する試験は絶対に行われない！
- 日本の現状～教育を受けていない、専門家もいない～
- 生物統計学者の役割～数理的に表現～

35

## 初回授業の内容

- 『統計手法なんてどうでも良い、  
臨床試験はデザインですべてが決まる』
- 医師が（医療者が）効果があると信じることの重要性  
信じていない医療行為に関する試験は絶対に行われない！
- 日本の現状～教育を受けていない、専門家もいない～
- 生物統計学者の役割～数理的に表現～

36

## 宿題の一部をさらっと！

自由度の  $\chi^2$  分布に従う  $n$  の観測値の  $k$  次の積率は

$$\int_0^{\infty} x^k \frac{1}{\Gamma(k/2)} x^{k/2-1} \exp(-x/2) dx$$

で計算できる。

$$\int_0^{\infty} x^k \frac{1}{\Gamma(k/2)} x^{k/2-1} \exp(-x/2) dx = \frac{2^{k/2}}{\Gamma(k/2)} \int_0^{\infty} x^{k/2-1} \exp(-x/2) dx$$

$$= \frac{2^{k/2}}{\Gamma(k/2)} \int_0^{\infty} x^{k/2-1} \exp(-x/2) dx$$

$$= \frac{2^{k/2}}{\Gamma(k/2)} \Gamma(k/2) = 2^{k/2}$$

$$= \frac{2^{k/2}}{\Gamma(k/2)} \Gamma(k/2) = 2^{k/2}$$

と計算できる。以上より、スコア検定統計量は

$$X_0^2 = n(\hat{\theta}_0)_{\alpha}^{-1}(\hat{\theta}_0)$$

$$X_0^2 = \frac{2n \log \frac{1}{\alpha}}{2} \left( \frac{1}{\alpha} \right)^{\frac{2n-2}{2}} \left( \frac{1}{\alpha} \right)^{\frac{2n-2}{2}} = \frac{2n \log \frac{1}{\alpha}}{2} \left( \frac{1}{\alpha} \right)^{2n-2}$$

と計算できるので、左度検定統計量は

$$X_1^2 = -2 \ln(\hat{\theta}_0) - (\hat{\theta}_0) = -2 \ln \frac{1}{\alpha} - \frac{1}{\alpha} = 2 \ln \frac{1}{\alpha} - \frac{1}{\alpha}$$

37

## 「バイアス」を式で「表現する」

- バイアスの影響を抑えるには...

$$\textcircled{1} \quad E(\bar{X}_A - \bar{X}_B) = (\tau_A - \tau_B) + \left\{ \frac{\sum_{j=1}^J (\alpha_j \sum_{i=1}^I n_{Aij})}{\sum_{j=1}^J \sum_{i=1}^I n_{Aij}} - \frac{\sum_{j=1}^J (\alpha_j \sum_{i=1}^I n_{Bij})}{\sum_{j=1}^J \sum_{i=1}^I n_{Bij}} \right\} + \left\{ \frac{\sum_{i=1}^I (\delta_i \sum_{j=1}^J n_{Aij})}{\sum_{j=1}^J \sum_{i=1}^I n_{Aij}} - \frac{\sum_{i=1}^I (\delta_i \sum_{j=1}^J n_{Bij})}{\sum_{j=1}^J \sum_{i=1}^I n_{Bij}} \right\}$$

$$\textcircled{2} \quad (\tau_T - \tau_A) + \lambda(1-\lambda) \left( \frac{1-s-t}{\lambda s + (1-\lambda)(1-t)} \right) (\mu_p - \mu_n) + (\tau_A - \epsilon) \left( \frac{(1-\lambda)(1-t)}{\lambda s + (1-\lambda)(1-t)} \right)$$

38

AMED班会議資料より 一部改変

## 用語説明 「レセプト」

- ある個人の診療毎の請求書。診療内容を把握できる（単位：枚）
- 診療1回毎、毎月が発生。異なる人物には異なるレセプトが発生すると同時に、同一人物の複数医療機関受診、複数月受診、入院と外来、保険者変更、・・・でも異なるレセプトが発生することに注意
- ある個人の診療を経時的に追跡するためには、複数枚のレセプトを名寄せする必要がある  
(1枚1枚のレセプトも複数の「レコード」に分割されて格納されている)

40

### Problem 1.4. (a)

Microsoft Excel を用いて平均値 10、標準偏差 5 の正規分布から無作為に 10 観を 10000 回抽出し 10 観の標本の平均値が 10 であるという確率の数値につき、スコア検定、Wald 検定（左度検定）を行った。同様、20 観を 10000 回抽出、100 観を 10000 回抽出、50 観を 10000 回抽出、100 観を 10000 回抽出してスコア検定、Wald 検定（左度検定）を行った。下の表は今回のシミュレーションの結果である。10000 回のうち帰無仮説が有意水準 0.05 で棄却された回数を示している。スコア検定では 5 つの観例数とも  $\alpha$  レベの確率が 0.05 を下回っており、 $n=20$  以上では 0.05 に近くになっている。一方で Wald 検定では、5 つの観例数とも  $\alpha$  レベの確率が 0.05 を上回っており、 $n=10, 30, 50$  では 0.05 より明らかに大きくなっている、 $n=100$  では 0.05 に近くになっている。

	n=10	n=20	n=30	n=50	n=100
スコア検定	426110000	465110000	478110000	472110000	473110000
Wald 検定	566110000	729110000	854110000	968110000	917110000
左度検定	713110000	611110000	566110000	507110000	465110000

Wald 検定統計量は、

$$X_0^2 = n(\hat{\theta}) \left( \frac{\partial \ln(\hat{\theta})}{\partial \theta} \right)^{-1} (\hat{\theta}) \left( \frac{\partial \ln(\hat{\theta})}{\partial \theta} \right)^{-1} n(\hat{\theta})$$

$$X_0^2 = \left( \frac{2n-2}{2n-1} \right) \left( \frac{2n-2}{2n-1} \right) \left( \frac{2n-2}{2n-1} \right) \left( \frac{2n-2}{2n-1} \right) \left( \frac{2n-2}{2n-1} \right) \left( \frac{2n-2}{2n-1} \right)$$

となるので、フィッシャー情報行列  $I_{\theta}(\theta, \theta)$  は、

$$I_{\theta}(\theta, \theta) = \begin{pmatrix} \frac{2n}{2n-1} & 0 \\ 0 & \frac{2n}{2n-1} \end{pmatrix}$$

39

## 用語説明 「名寄せ」

- 複数のファイルやレコードに分置された同一人物等のデータを、キー変数を用いて一つに紐付けすること（単に紐付けとも）
- 同一人の名寄せにおいては、  
「同じ人を同じ人と識別する」（＝同じ人物を見逃さない）  
「異なる人を異なる人と識別する」（＝異なる人を紐付けない）  
の2点を満たす必要がある。

## 「レセプトデータ」とは？

- 「レセプトデータ」の意味は個々人で違う!?  
レセプト情報・特定健診等情報データベース（NDB）  
NDBオープンデータ  
国保データベース（KDB）  
などなど
- 同じ「レセプトを用いた分析」でも内容は異なる

## 皆が「NDB」と言っている データには何種類がある

- ① 匿名化されていない日本全体のレセプトデータベース（**幻想**）  
厚生労働省が有する匿名化後のデータベース（**現実**）
- ② 匿名化後データベース(①-現実)から切り出され、多数のファイル（＝ミンチ化）の形で研究者へ渡されるデータ  
（当研究班が考察対象とした“**研究分析用の”NDB**）

### ミンチ肉はステーキ肉に戻せるか

データベースから匿名化して切り出した情報をもらい受けるため、再度データベースの構築が必要だが、再構築が難しい

- ・ 独特の欠陥を抱える2種類の個人IDと1種類のレセプト通番を手がかりに、一年分で 約15億レセプト、300億レコード、3TBのデータの名寄せを試みる

### 料理人が牛の解体まで行う必要がある

牛を丸ごと一頭分持ってこられて、「必要な部位を自分で切り出し、使える形に自分で成形してくれ」と言われているようなもの

- ・ データが巨大すぎるため、どこをどう切り出せば目的のデータが手に入るかの手法を示す必要

## 今後の学会の予定

- 公衆衛生学サマーセミナー ～現在の研究内容と今後の展望～（演題未定）  
2017年8月18日
- 国際疫学会 ～ Early surgical intervention and mortality in patients with hip fractures: analysis using Japanese nationwide administrative database ～  
2017年8月19日～22日
- 日本臨床疫学会 ～レセプト情報・特定健診等情報データベース（NDB）の活用  
：日本の糖尿病通院患者数～ 2017年9月30日10月1日
- 日本公衆衛生学会 ～レセプト情報・特定健診等情報データベース（NDB）の活用  
：糖尿病診療の全体像～ 2017年10月31日～11月2日
- 日本糖尿病学会近畿地方会 症例発表
- 日本糖尿病学会総会 日本の糖尿病診療の全体像