



関西広域 医療データ人材教育拠点形成事業  
Kansai Union / Kyoto University Education Program  
for Digital Health Innovation

令和4年12月7日

第13回匿名医療情報等の提供に関する専門委員会

参考資料 1

# 学生実習におけるサンプリングデータセット 利用について



京都大学  
KYOTO UNIVERSITY

# 申出の概要①

**【研究の名称】 医療データ利活用人材育成に資するデータ利用法の開発**

**【担当者】 京都大学医学部附属病院 医療情報企画部 教授 黒田 知宏**

## **【目的】**

医療データの効果的な利活用の推進と共に、次世代医療基盤や保健医療データプラットフォームに蓄積されるデータ分析を担う人材育成の必要性が高まっている。NDBデータについてもデータを正しく取り扱え、有効な分析ができる人材の育成が望まれている。今回の研究では、医療データ人材育成に資するデータ利用法の開発を行う

# 申出の概要②

## 【概要】

我々は文部科学省「医療データ人材育成拠点形成事業」において、医療データの利活用推進に資する人材育成プログラムを実施しており、本研究はその一環としてNDBデータ分析の初学者が分析手法を習得するために必要な教育プログラムの開発を行う。具体的にはサンプリングデータを用いて、NDBデータを分析する際のデータハンドリングの定型化について検討する。また、開発された手法について「医療データ人材育成拠点形成事業」に参加する大学院生を対象に有用性についての検証を行う。この検証にあたり、利用者申請をしていない学生が実習可能なデータセットが必要となるため、サンプリングデータセットを加工し、「レセプト情報・特定健診等情報の提供に関するガイドライン」の公表基準に準拠したデータセットを作成する。

## 【公表予定】

本研究は、研究拠点形成費等補助金（先進的医療イノベーション人材養成事業）における、「医療データ人材育成拠点形成事業」の一環として行われるものであり、成果の公表は当事業に関する報告書及び関連学会での発表・公表等の形式にて行う予定である。  
また、本研究で作成されたデータセットについては、学生実習において公開(利用)可能とする予定である。

## 概要

### 【背景・目的】

- 次世代医療基盤法・保険医療データプラットフォーム稼働等医療データに関する周辺環境の変化に対応可能な人材育成
- 医療データに触れられる環境での人材育成

### 【要件】

- 大学院修士課程又は博士課程の人材育成コースの設置
- 一定期間で知識と技術を習得できる研究コースの設置
- 2020年4月開講
- 複数の大学が連携して教育を実施
- 補助金終了後も発展的・継続的な活動を行う

### 【期間】

- 補助期間 3年 + 2年

# 関西広域 医療データ人材教育拠点形成 プログラム

Kansai Union / Kyoto University Education Program for Digital Health Innovation (KUPE-DHI)

京都大学 滋賀大学 京都府立医科大学 奈良県立医科大学 大阪大学 関西医科大学 近畿大学 和歌山県立医科大学 神戸大学 兵庫医科大学 鳥取大学、滋賀医科大学、兵庫県立大学



# 2021度の取り組み状況

---

## ①京都大学 以下のコースにおいて2021年度 受講生受け入れ

- ・医療データ取扱専門家育成コース
- ・社会変革型医療データサイエンティスト育成プログラム:DHIEP
- ・ビジネス特化型インテンシブコース : dot.b

## ②2021年度外部聴講生の受け入れ

## ③教育コンテンツの改良

- 医療情報を利用可能なセキュアな実習環境の構築
- NDBサンプリングデータセットの医療情報学実習での利用開始

## ④オンラインイベント開催

## ⑤教科書作成

## ⑥KUEP-DHI Instagram開始

# 医療データ取扱専門家育成コース概要

○京都大学医学研究科と情報学研究科に修士課程追加履修コースとして設置

○対象者

医学研究科(医科学専攻・人間健康科学系専攻・社会健康科学系専攻)および  
情報学研究科(各専攻)大学院生

○2020年4月開講

○コンテンツ

基礎科目:全体を理解するために必要な知識

・計算機科学概論・医学概論・統計学

必修科目:医療データサイエンスに必要な知識

・医療情報システム学・医療データ分析学・医療情報法制学・医療情報学実習

選択科目:医療データサイエンスに関わる関係科目

・関連研究科の関連講義を指定

※教育コンテンツ(テキスト・e-learning等)の作成

※実習(京大病院の電子カルテデータを用いた教育を実施予定)

※連携大学から受講可能とし、各大学のコース設計に活用いただけるようにする



# 受講状況(医療データ取扱専門家育成コース)

2021年度 受講生: 21名

(内訳)	情報学研究科修士学生	6名
	社会情報学専攻	4名
	先端数理科学専攻	2名
	医学研究科修士学生	12名
	社会健康医学系専攻	3名
	人間健康科学系専攻	9名
	医学研究科博士学生	3名
	医科学専攻	1名
	人間健康科学系専攻	1名
	社会健康医学系専攻	1名

# 修了要件

基礎科目	計算幾科学概論	情報科学基礎論
		情報分析・管理論
	医学概論	病態生理学
		医学基礎I
		疾病論
	統計学	臨床医学概論
		医療統計学
		保健医療統計学
		数理統計学 -データサイエンス 1- データ科学概観

情報系出身者の履修要件	医学系出身者の履修要件
情報系学部出身者は基本的には履修不要	要履修 2単位
要履修 2単位	医学系系学部出身者は基本的には履修不要
学部で統計科目履修済の場合は履修不要	

必修科目	医療情報システム学	医療情報学
	医療データ分析学	C20医療情報学
	医療情報法制学	医療情報法制学
	医療情報学実習	医療情報学実習

要履修 2単位	要履修 2単位

選択科目	基盤人材志向科目	情報システム設計論
		情報セキュリティ
	統制人材志向科目	基礎医療倫理学
		情報法制学 (開講未定)
	活用人材志向科目	ビッグデータ医科学特論
		人工知能特論
		データ科学展望I/II/III/IV 内二つ 複雑系科学 -データサイエンス 2-

要履修 2単位	要履修 2単位
---------	---------

・本コースは京都大学医学研究科と情報学研究科に追加履修コースとして設置する。

・当該研究科における既存の講義のうち、本コースの構成要素となる講義については、これらを必修単位の一部に数え、追加的に法制を含む核となる幾つかの講義と実習を履修することを、修了要件とする。

・本コースの修了には**16単位**の取得が必要である。(ただし、基礎科目の履修条件によって、実質的には10~14単位の取得が求められる。)

# 必修科目講義内容概要

## 医療情報学実習

第1回	システム設計・DB設計	医療データ（種別）， Excel => RDB, DB設計
第2回	システム設計・DB設計	SQL発展
第3回	システム設計・DB設計	R 発展 with SQL
第4回	システム設計・DB設計	Python with R & SQL
第5回	オープン DB を利用した医療統計データ処理	オープンデータ（Excel） => RDB, SQL整理, R分析, Python 表示
第6回	オープン DB を利用した医療統計データ処理	オープンデータ（Excel） => RDB, SQL整理, R分析, Python 表示
第7回	NDB データの取り扱い	NDB サンプルデータ加工済データ
第8回	NDB データの取り扱い	NDB サンプルデータ加工済データ
第9回	臨床研究データを利用した疫学分析	臨床試験データ
第10回	臨床研究データを利用した疫学分析	臨床試験データ
第11回	医療データと機械学習	
第12回	医療データと機械学習	
第13回	経営分析	
第14回	経営分析	

# 必修科目講義内容概要

## 医療情報法制学

第1回	医療情報に関わる倫理①	生命倫理
第2回	法学基礎	
第3回	現在の医療を取り巻く制度について	保険医療制度および公費医療について
第4回	世界の医療制度、医療政策について	
第5回	医療情報に関わる倫理②	情報倫理
第6回	情報法基礎	個人情報保護法・プライバシー保護・GDPR
第7回	情報法基礎	不正競争防止法（営業機密保護） 著作権法 サイバーセキュリティ基本法
第8回	医療情報の利用者が知っておくべき情報法制について	薬機法
第9回	医療情報の利用者が知っておくべき情報法制について	e文書法・3省3ガイドライン・プロバイダ責任制限法・電気通信事業法
第10回	医療情報の利用者が知っておくべき情報法制について	医事法（医師法・歯科医師法・医療法）
第11回	医療情報の利用者が知っておくべき情報法制について	次世代医療基盤法・HIPAA等
第12回	医療情報の利用者が知っておくべき情報法制について	倫理指針関連・各種ガイドライン等
第13回	どうすれば医療ビッグデータが利用できるのか	利用可能なデータおよび利用申請から利用開始まで
第14回	医療情報を取り巻く環境について	最近の動向について



# 今回の作成した実習用のデータセットについて

11

## NDBサンプリングデータセット

医科入院外：高額レセプト除去後、患者の1%を性別、年齢層別に抽出

医科入院：高額レセプト除去後、レセプト数の10%を抽出

DPC：レセプト数の10%抽出

調剤：高額レセプト除去後、レセプト数の1%抽出

※傷病名コード・医薬品コード、診療行為コード、診断群分類については出現回数の少ないものから総出現回数の0.1%に達するまで匿名化。ただし、「医科入院」「医科入院外」における診療行為コードについては、少ないものから順に0.01%に達するまで匿名化されている。



## 今回作成するデータセット

・傷病名コード・医薬品コード・診療行為コード・診断群分類について  
すでに出現頻度が低いものは匿名化されているが、さらに今回**サンプリングデータセット内での出現頻度が10回以下のものについてはレコード自体を削除**する。



## 提供を受けたデータセット

12

**データ提供対象**：2011年1月診療分～2015年10月診療分のサンプリングデータセット（医科入院外、医科入院、DPC、調剤）

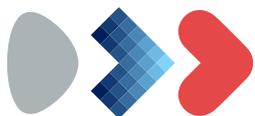
**医科入院**：IR（医療機関）・RE（レセプト共通）・HO(保険者)・**SY(傷病名)**・**SI(診療行為)**  
・**IY（医薬品）**・・・

**医科外来**：IR（医療機関）・RE（レセプト共通）・HO(保険者)・**SY(傷病名)**・**SI(診療行為)**  
・**IY（医薬品）**・・・

**DPC**：IR（医療機関）・RE（レセプト共通）・HO(保険者)・**BU(診断群分類)**・**SY(傷病名)**  
・KK（患者基礎）・**SI(診療行為)**・**IY（医薬品）**・・・

**調剤**：YK(薬局)・RE（レセプト共通）・HO(保険者)・SH(処方基本)・CZ(調剤情報)  
・**IY（医薬品）**・・・

**赤字**がデータ加工対象コードを有するレコード



# 加工方法①

診療行為コードとして10回以下の出現率のコードを認めた場合、該当のコードを有するIDのレコードは全レコードにおいて全削除とした。

## 例) 診療行為コードに10回以下の出現コードを認めた場合

通番1	通番2	レコード種別	診療識別	診療行為コード	数量	点数	回数	...
1	111122XXX	SI	12	160038010			144	1
12	353535XXX	SI	12	160203810			58	1
7	111111XXX	SI	14	190217970			230	2
4	222222XXX	SI	11	190217970			68	1
8	333333XXX	SI	11	140049910			18	
...	252525XXX	SI	13	150290210			280	
...	...	SI	12	140019710				
...	...	SI	12	...			65	
...	...	SI	12	...			12	
...	...	SI	25	...			33	
...	...	SI	14	...			21	
...	...	SI	12	...			46	
...	...	SI	11	...			75	
...	...	SI	11	...			88	
...	...	SI	12	...			12	

①診療行為コード190217970が出現率10回以下だった場合、診療行為コード190217970を有する通番2「111111XXX」と「222222XXX」のレコードをSIレコードから削除する。  
 (コードをnull化するだけの場合、点数等で同定されるリスクを除外するため、レコード毎の削除を行った。)

通番1	通番2	レコード種別	診療識別	診療行為コード	数量	点数	回数	...
1	111122XXX	SI	12	160038010			144	1
12	353535XXX	SI	12	160203810			58	1
8	333333XXX	SI	11	140049910			18	3
...	252525XXX	SI	13	150290210			280	1
...	...	SI	12	140019710			15	2
...	...	SI	12	...			65	1
...	...	SI	12	...			12	1
...	...	SI	25	...			33	1
...	...	SI	14	...			21	1
...	...	SI	12	...			46	2
...	...	SI	11	...			75	1
...	...	SI	11	...			88	1
...	...	SI	12	...			12	1



## 【第1回 2021年11月11日】受講生 19名（オンデマンド含む）

- 1) 糖尿病患者に関する集計
  - ①糖尿病の病名を持つIDを抽出
  - ②糖尿病の治療薬を処方されているIDを抽出

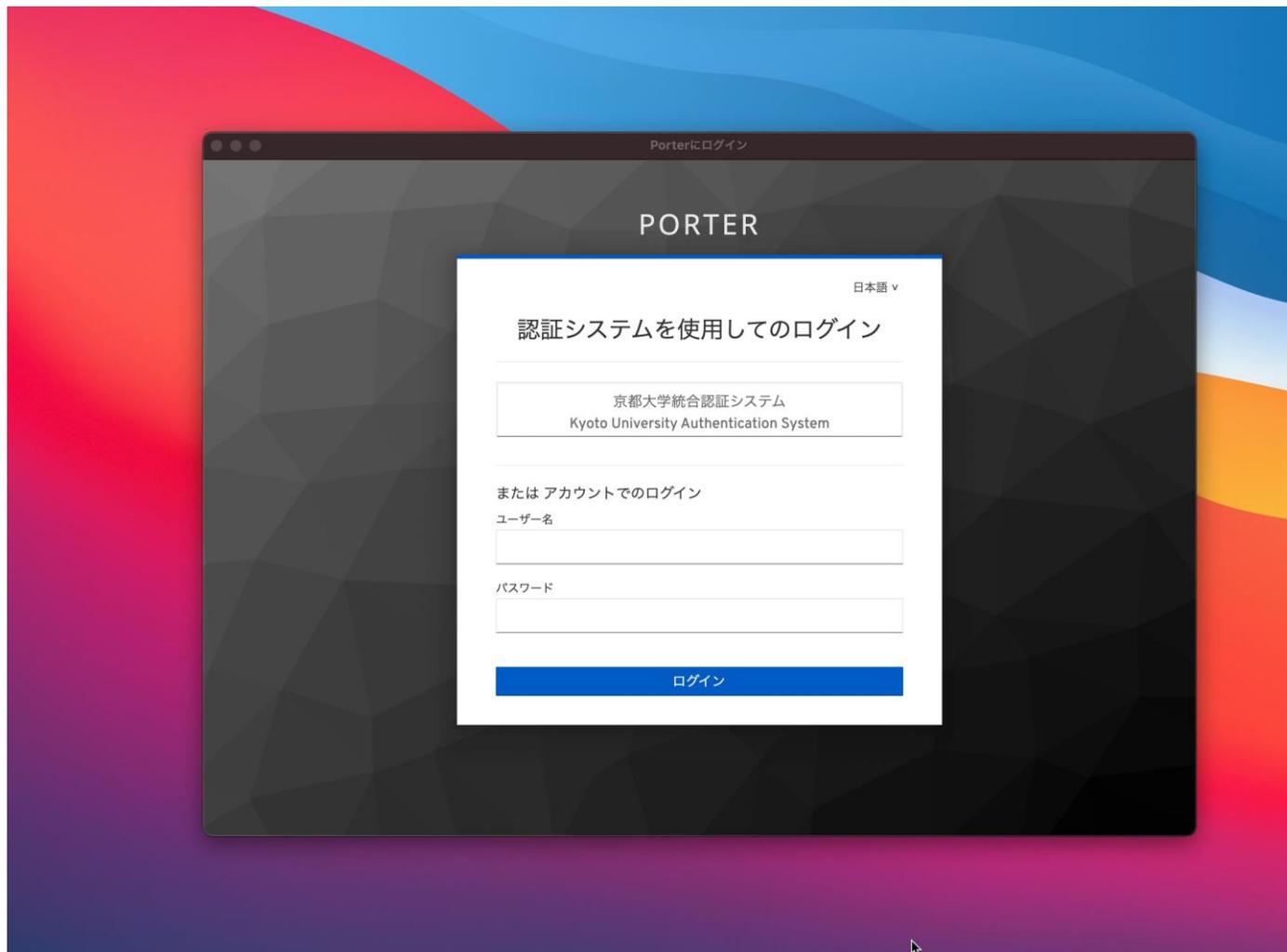
➤ ①かつ②のIDを集計
- 2) 1) と同様の集計を膝関節症と貼り薬で実施

## 【第2回 2021年11月25日】受講生 19名（オンデマンド含む）

- 3) サンプルングデータセットを用いた論文を題材とした実習  
<https://pubmed.ncbi.nlm.nih.gov/26990125/>



# 実習環境整備





## 今後の予定について

- KUEP-DHI事業は継続中であり、医療情報学実習は2022年度も開講予定（利用期間延長申請済）
- 今後も今回作成させて頂いた実習用データセットを大学院での実習および他の人材育成事業（社会人向け含む）でも利用継続させて頂きたい（加工済データセットの教育目的での制限なしでの利用は可能でしょうか？）