Building an Organization to Accelerate Data Utilization

Hideki Ninomiya
Datack Inc

Profile

Career

- Tokyo University(Doctor of Medicine)
- Kansai Medical University
  - neurosurgeon
- Medley, Inc.
  - Medical Information
  - Telemedicine
- 3idea, Inc.
  - Data Science
- Datack, Inc.

Skill

- SQL, R, Python
- AWS
- epidemiology, data engineering, machine learning, natural language processing

Service

Real World Data Analysis

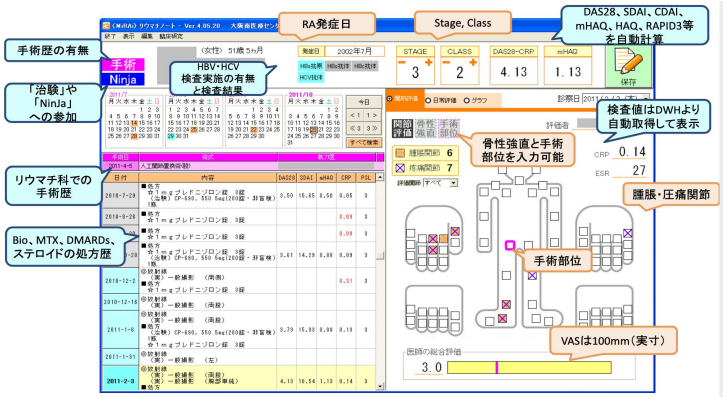
- Epidemiology, Data Engineering
- Claims data, Registry, EHR
- Database study

Real World Evidence

Building a clinical database in collaboration with hospitals

Goal: Application to Approval

- Area
  - Rheumatoid Arthritis



Today’s topic

- Why do we use RWD?
- Challenges in utilizing RWD
- Four Keys to Organization Building
- Practices

DATACK

Why do we use RWD?

Why do we use RWD?

DATACK

First of all, we need to clarify the purpose of RWD utilization.

Advanced real-world-evidence analytics can play an important role across the pharma value chain.

R&D

Identify unmet need:
- Inform research decisions

Innovate in trial design:
- Use synthetic control arms

Improve trial design:
- Define inclusion/exclusion criteria and end-points
- Optimize site selection
- Accelerate recruitment

Accelerate time to market

Refine formularies:
- Determine optimal dosing based on patient response

Monitor real-world outcomes:
- Quantify unmet need
- Understand safety and efficacy profiles

Market access

Improve evidence of economic value:
- Demonstrate economic value of treatment to payer
- Compare trial data with real-world evidence (RWE) to strengthen dossier
- Enable outcomes-based pricing

Improve formulary position:
- Achieve better patient access
- Show efficacy and safety through head-to-head in silico trials

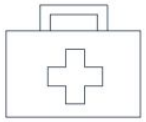
Achieve label expansion:
- Use RWE to eliminate need for new randomized clinical trial

Sales and marketing

Improve targeting of commercial activities:
- Target underdiagnosed patients
- Identify "super responders"
- Identify patients likely to switch or discontinue
- Inform design of patient services/solutions

Refine commercial strategy:
- Shape product positioning
- Understand healthcare-provider (HCP) decision making and impact on outcomes
- Sharpen understanding of influence networks

Build clinical-decision-support systems:
- Provide recommendations at point of care based on predictions of outcomes, risk, or disease progression



Medical

Improve pharmacovigilance:
- Monitor real-world usage for safety and adverse events
- Rapidly create granular view on benefits/risks

Strengthen evidence for differentiation:
- Analyze efficacy in under-studied populations
- Identify subpopulations for which effect outperforms trials

Improve effectiveness of medical affairs:
- Monitor unmet patient need at HCP level

Improve adherence:
- Support personal engagement to drive adherence and capture patient-reported outcomes with digital tools

How pharma companies are applying advanced analytics to real-world evidence generation | McKinsey

Why do we use RWD?

DATACK

green: start work → red: high impact, challenging

R&D

Identify unmet need:
- Inform research decisions

Innovate in trial design:
- Use synthetic control arms

Improve trial design:
- Define inclusion/exclusion criteria and end-points
- Optimize site selection
- Accelerate recruitment

Accelerate time to market

Refine formularies:
- Determine optimal dosing based on patient response

Monitor real-world outcomes:
- Quantify unmet need
- Understand safety and efficacy profiles

Market access

Improve evidence of economic value:
- Demonstrate economic value of treatment to payer
- Compare trial data with real-world evidence (RWE) to strengthen dossier
- Enable outcomes-based pricing

Improve formulary position:
- Achieve better patient access
- Show efficacy and safety through head-to-head in silico trials

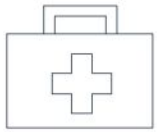
Achieve label expansion:
- Use RWE to eliminate need for new randomized clinical trial

Sales and marketing

Improve targeting of commercial activities:
- Target underdiagnosed patients
- Identify "super responders"
- Identify patients likely to switch or discontinue
- Inform design of patient services/solutions

Refine commercial strategy:
- Shape product positioning
- Understand healthcare-provider (HCP) decision making and impact on outcomes
- Sharpen understanding of influence networks

Build clinical-decision-support systems:
- Provide recommendations at point of care based on predictions of outcomes, risk, or disease progression



Medical

Improve pharmacovigilance:
- Monitor real-world usage for safety and adverse events
- Rapidly create granular view on benefits/risks

Strengthen evidence for differentiation:
- Analyze efficacy in under-studied populations
- Identify subpopulations for which effect outperforms trials

Improve effectiveness of medical affairs:
- Monitor unmet patient need at HCP level

Improve adherence:
- Support personal engagement to drive adherence and capture patient-reported outcomes with digital tools

How pharma companies are applying advanced analytics to real-world evidence generation | McKinsey

Challenges in utilizing RWD

Challenges in utilizing RWD

DATACK

1. Insufficient human resources and organization
    - Today's main topic

2. Quality and Quantity of RWD
    - Bias, Confounding
    - Lack of required clinical data
    - Not enough patients

3. Regulation
    - Approval
    - pharmacovigilance

Frequent bias, confounding

Category/Subcategory	Percentage (%)※2	論文数※1
Confounding	63.2	74
Confounding by indication	32.5	38
Time-dependent confounding	6.0	7
Unmeasured/residual confounding	28.2	33
Healthy user/adherer effect	5.1	6
Selection bias	47.0	55
Protopathic bias	3.4	4
Losses to follow-up (informative censoring)	2.6	3
Depletion of susceptibles (prevalent user bias)	10.3	12
Missing data	17.9	21
Measurement bias	46.2	54
Misclassification bias	43.6	51
Misclassification of exposure	23.9	28
Misclassification of outcome	28.2	33
Time-related bias	30.8	36
Immortal time bias	25.6	30
Immeasurable time bias	3.4	4
Time-window bias	2.6	3
Time-lag bias	0.9	1

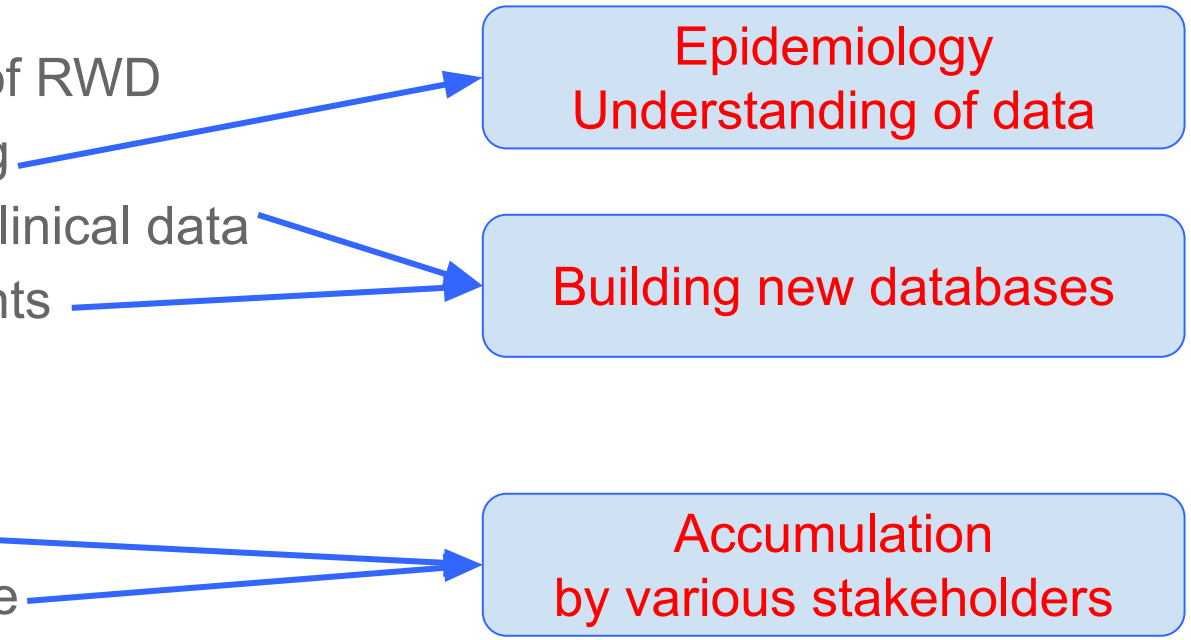
Prada-Ramallal, G., Takkouche, B. & Figueiras, A. Bias in pharmacoepidemiologic studies using secondary health care databases: a scoping review. BMC Med Res Methodol 19, 53 (2019).

Solutions to the Challenges

1. Insufficient human resources and organization
   - Today's main topic

2. Quality and Quantity of RWD
   - Bias, Confounding
   - Lack of required clinical data
   - Not enough patients

3. Regulation
   - Approval
   - pharmacovigilance



DATACK

Four Keys to Organization Building

Four Keys to Organization Building

1. A person who makes the strategy
2. Epidemiologists (RWD experts)
3. Each department, each person, gains experience
4. Create a system for using RWD

Required skill of Epidemiologists

DATACK

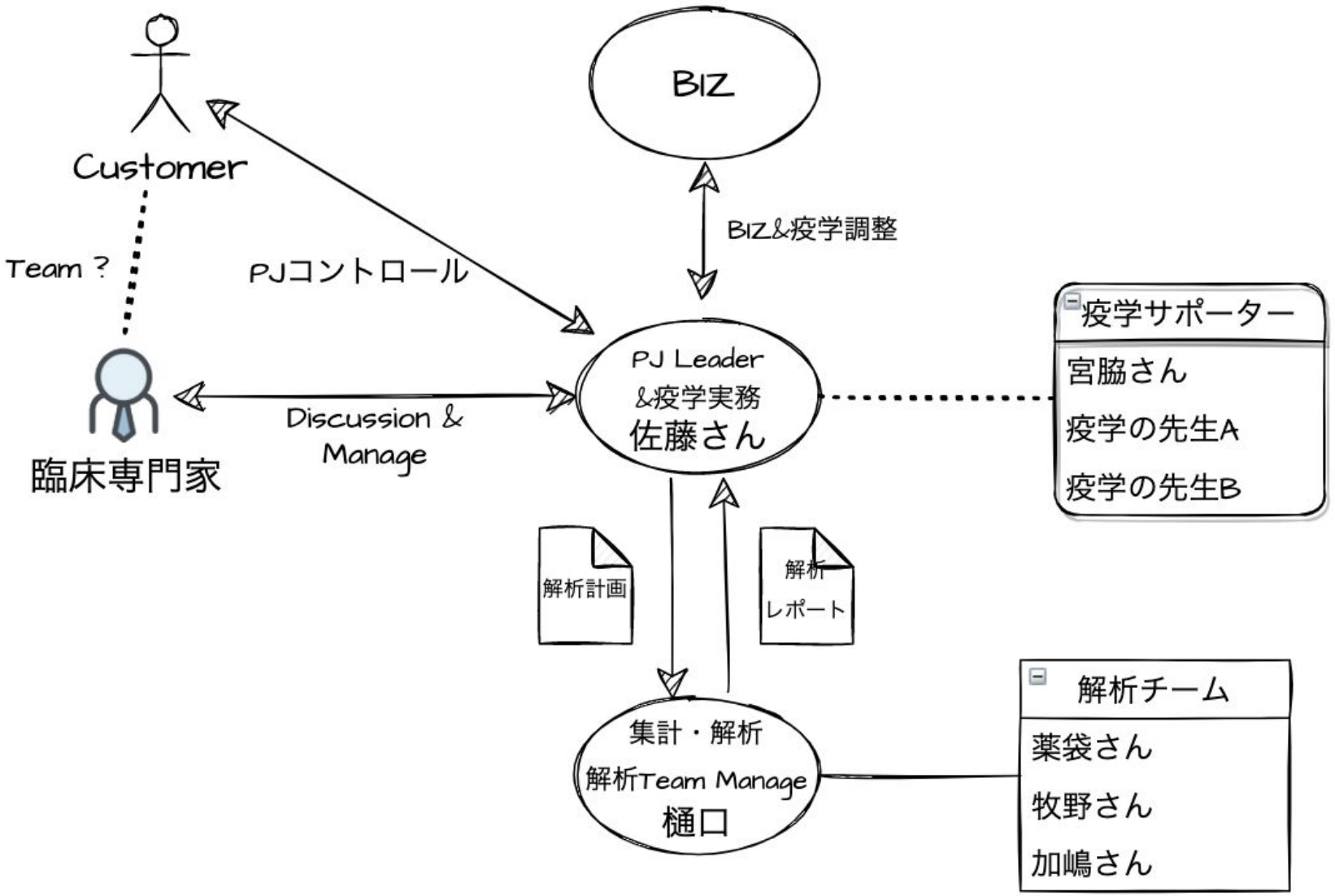
- Creating a research design
  - prospective research
  - database study
- Data handling
  - SQL, SAS, R…
  - Understanding of database and clinical practice
- Collaborating with clinical experts
  - Generating clinical questions
  - Making decisions
- Paper writing skills

DATACK

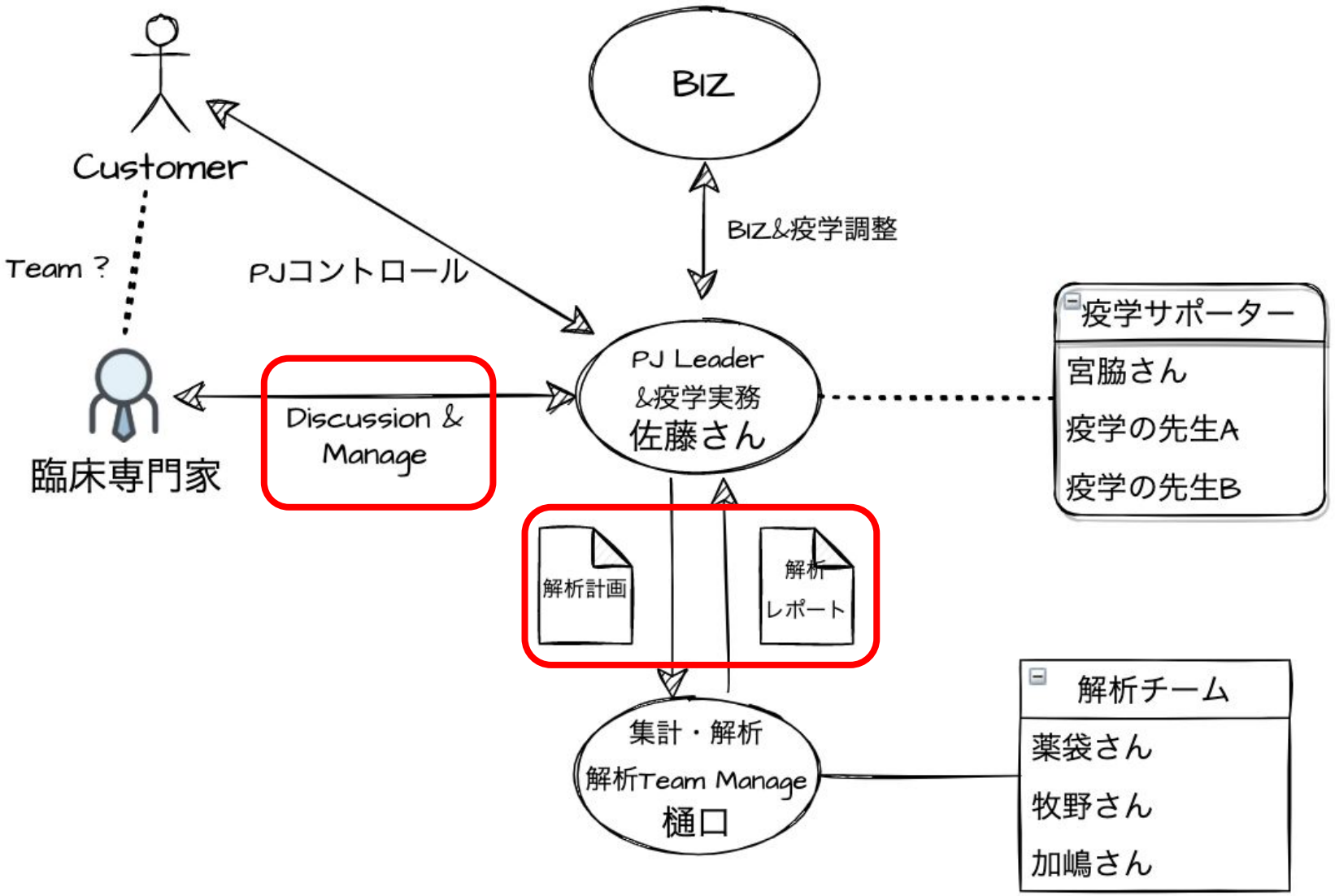
Practice

- Case: Datack
  - organization
  - workflow
- Extraction Request Form
- Code Set
- Analysis environment

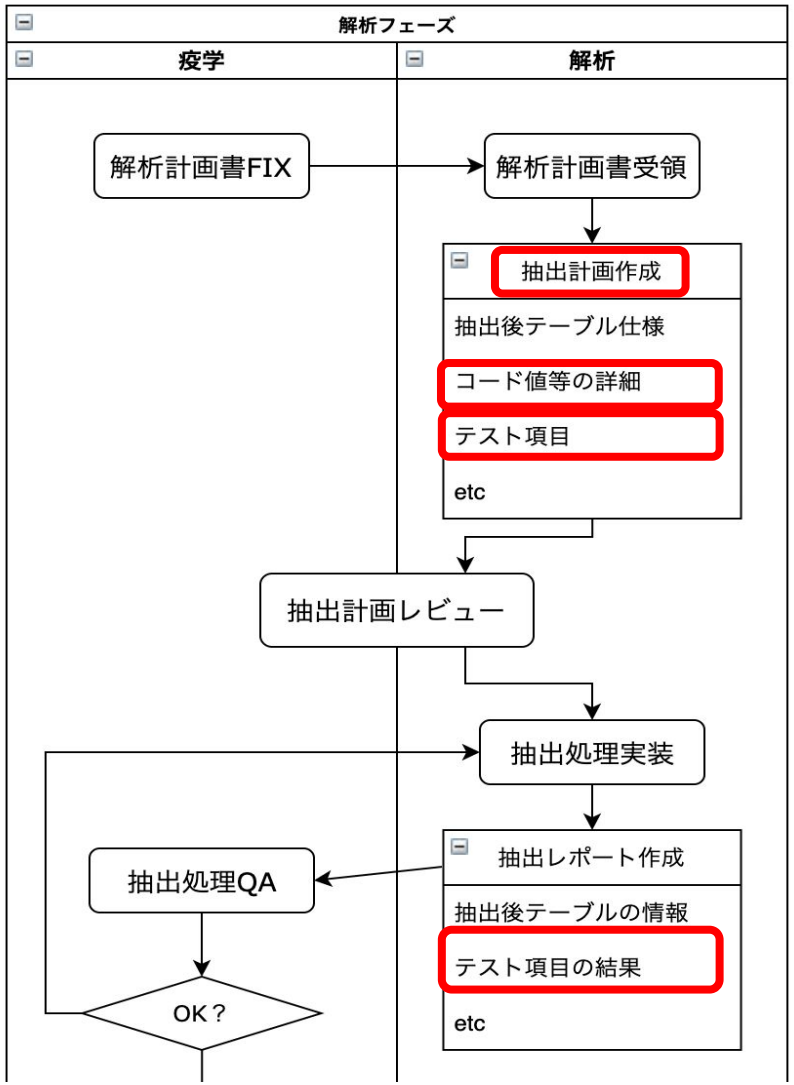
Datack Organization

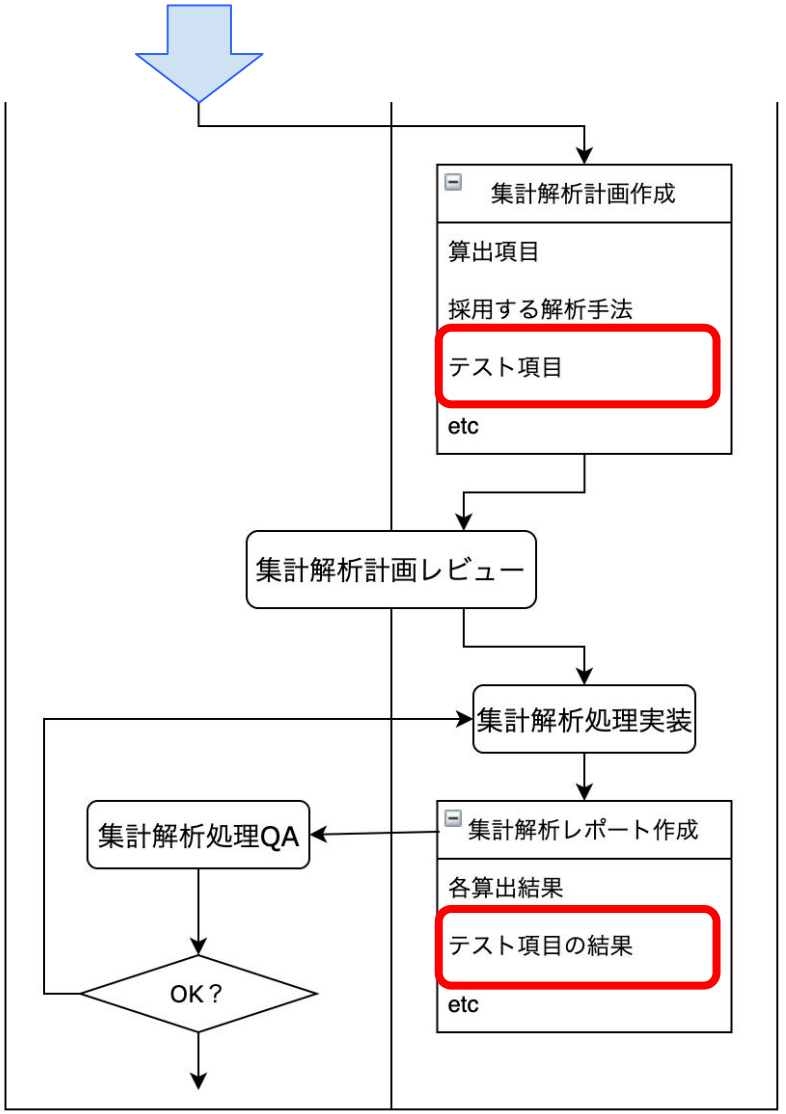


Datack Organization



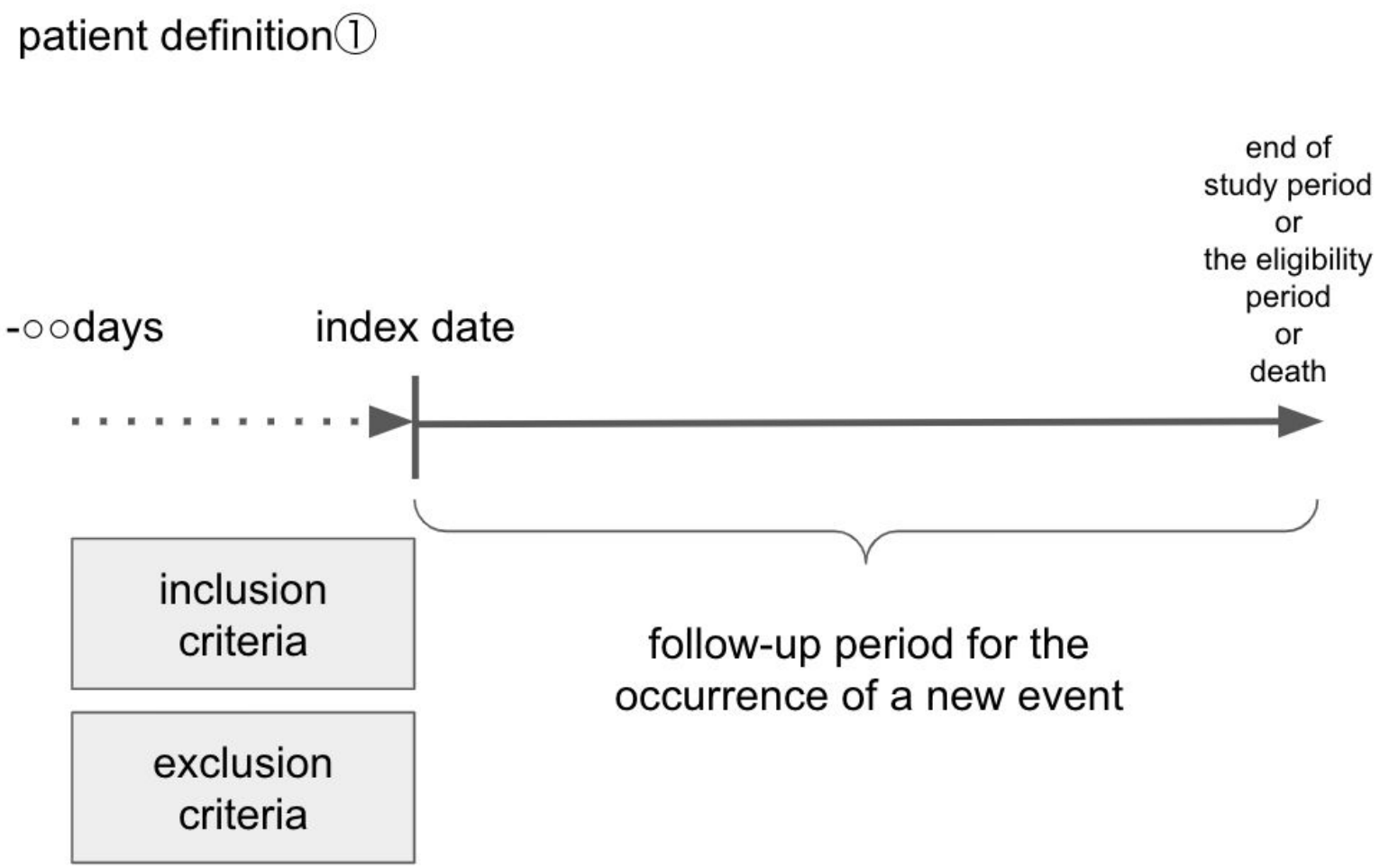
Workflow





右に続く

Extraction Request Form



Various Codes for Pharmaceuticals

DATACK

一般名	商品名	規格包装単位	薬価基準収載医薬品コード	個別医薬品コード（YJコード）	レセプト電算処理システム用コード	JANコード	調剤包装単位（GTIN-13）	販売包装単位（GTIN-14）	基準番号（HOT-13）	ATCコード
ロサルタンカリウム錠	ロサルタンカリウム錠２５ｍｇ「アメル」	２５ｍｇ／錠 ＰＴＰ １０錠×１０	2149039F1244	2149039F1244	622144801	4987058892039	04987058143971	14987058892036	1214482010101	C09CA01
		２５ｍｇ／錠 ＰＴＰ １４錠×１０				4987058892077	04987058143988	14987058892074	1214482010201	
		２５ｍｇ／錠 バラ ５００錠×１瓶				4987058892558	04987058214541	14987058892555	1214482010301	
ロサルタンカリウム２５ｍｇ錠	ロサルタンカリウム錠２５ｍｇ「ＮＰ」	２５ｍｇ／錠 ＰＴＰ １０錠×１０	2149039F1015	2149039F1201	622165301	4987190032300	04987190662507	14987190032307	1216530010101	
		２５ｍｇ／錠 ＰＴＰ １４錠×１０				4987190032324	04987190662521	14987190032321	1216530010201	
		２５ｍｇ／錠 バラ ５００錠×１瓶				4987190032348	04987190662538	14987190032345	1216530010301	
	ロサルタンカリウム錠２５ｍｇ「サンド」	２５ｍｇ／錠 ＰＴＰ １０錠×１０		2149039F1287	622177001	4987614408605	04987614408667	14987614408602	1217704010101	
		２５ｍｇ／錠 ＰＴＰ １４錠×５０				4987614408643	04987614408674	14987614408640	1217704010201	
		２５ｍｇ／錠 バラ ５００錠×１瓶				4987614408636	04987614408681	14987614408633	1217704010301	

いろいろな医薬品コード｜情報で医療をささえる　データインデックス株式会社

Code Set for Disease

DMARDs

Biological DMARDs (boDMARDS+bsDMARDS)

-TNF
インフリキシマブ (ATCコード：L04AB02)
エタネルセプト(L04AB01)
アダリムマブ（L04AB04）
ゴリムマブ（L04AB06）
セルトリズマブペゴル（L04AB05）
-IL6
トシリズマブ(L04AC07)
サリルマブ(L04AC14)
-T cell
アバタセプト（L04AA24）

Synthetic DMARDs

Targeted (tsDMARDS) =JAK阻害剤

トファシチニブ（L04AA29）
オルミエント（L04AA37）

Conventional (csDMARDS)

メトトレキサート（L04AX03）
レフルノミド（L04AA13）
ブシラミン（M01CC02）
ペニシラミン（M01CC01）
サラゾスルファピリジン（A07EC01）
ミゾリビン（ATCコードなし）
タクロリムス（L04AD02）
金チオリンゴ酸ナトリウム（M01CB01）
オーラノフィン（M01CB03）
メタルカプターゼ（M01CC01）
アクタリット（ATCコードなし）
イグラチモド（ATCコードなし）

Code Set for Disease, Procedure

＜Disease：ICD-10 code＞

- MI：I21
- Angina pectoris：I20
- Stroke：I60-I64
  - Ischemic stroke：I63-I64
- Atrial fibrillation：I48
- Heart failure：I50

＜CABG or PCI（K code）＞

- Ｋ５４６ 経皮的冠動脈形成術
- Ｋ５４７ 経皮的冠動脈粥腫切除術
- Ｋ５４８ 経皮的冠動脈形成術（特殊カテーテルによるもの）
- Ｋ５４９ 経皮的冠動脈ステント留置術
- Ｋ５５０冠動脈内血栓溶解療法
- Ｋ５５０－２ 経皮的冠動脈血栓吸引術
- Ｋ５５１ 冠動脈形成術（血栓内膜摘除）
- Ｋ５５２ 冠動脈、大動脈バイパス移植術
- Ｋ５５２－２ 冠動脈、大動脈バイパス移植術（人工心肺を使用しないもの）

Analysis environment requirements

DATACK

- Reproducibility of analysis
- Easy to review code
- Centralized data management and pre-processing
- Version control
- Storage, Memory, Process speed
- Security

Reproducibility & Replicability

	dataset	study design
Reproducibility	Same	Same
Replicability	Different	Same
	Same	Different
	Different	Different

Analysis Environment

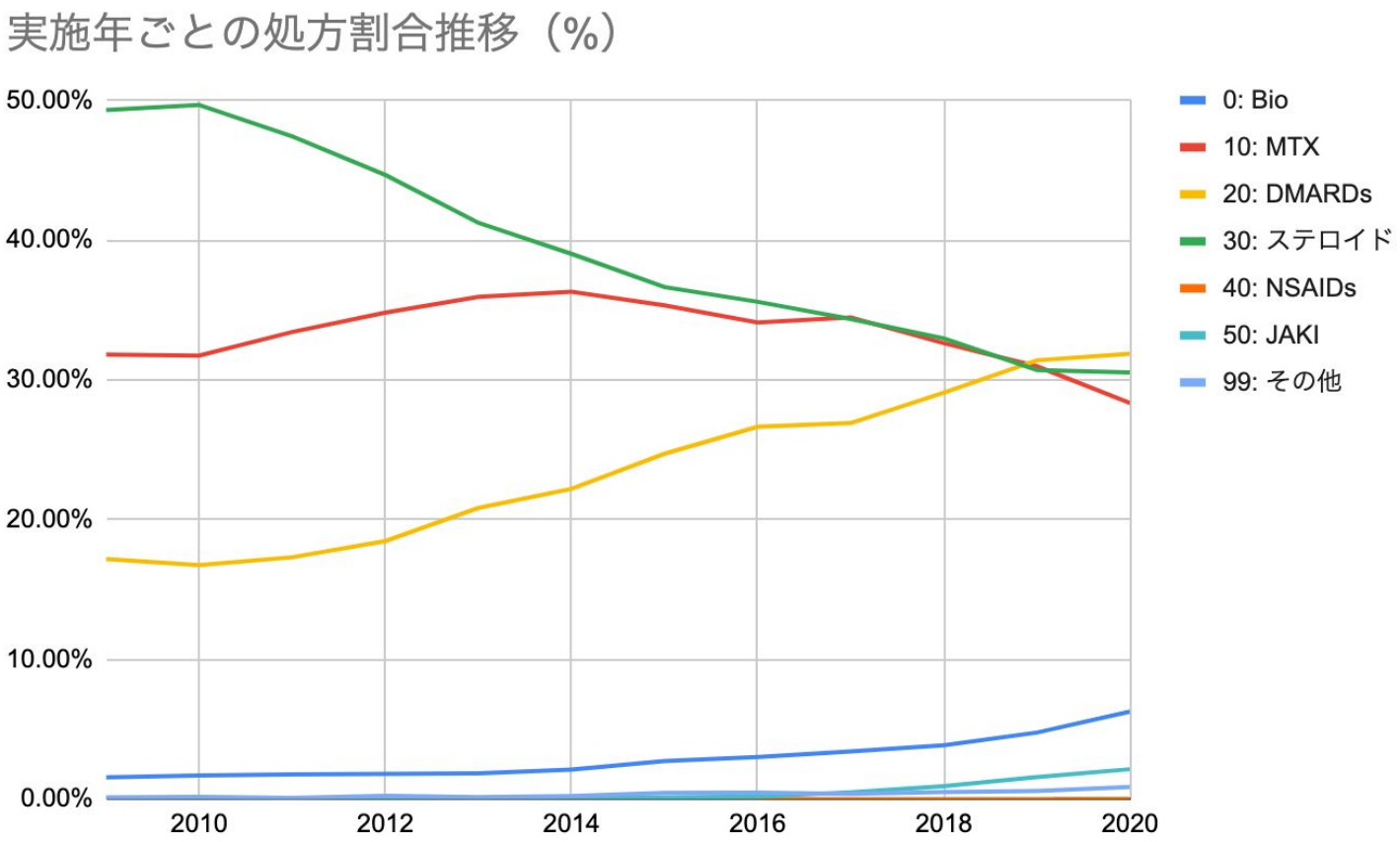
DATACK

data storage	 Amazon Simple Storage Service (Amazon S3)
database	 Amazon Athena  Amazon Redshift
analysis server	 Amazon Elastic Compute Cloud (Amazon EC2)
language	MySQL R Python
version control system	Bitbucket
execution environment	jupyter

Test of results

Based on clinicians and previous studies, it is important to determine indicators to check the validity of the tabulation and analysis.

- Percentage of patients using bDMARDs
  - Clinician's prediction: 20%



Easy to review

- Check the work of others
  - data
  - script
  - code
  - period

In [41]:
```
import yaml

import pandas as pd

from datack.path import user_home_path
from datack.sql import athena_quire_base


def load_athena_config():
    yaml_file = user_home_path("datack/config/athena_mirai.yml")
    with open(yaml_file) as f:
        config = yaml.load(f.read(), Loader=yaml.SafeLoader)["mirai"]
    return config


def athena_quire(sql_query, dtype=None):
    """Execute sql_query to select in Athena.
    Args:
        sql_query(str): SQL you want to execute.
    Return:
        pd.Dataframe: Result of sql_query
    Examples:
        >>> df = athena_quire('''SELECT * FROM sample_db LIMIT 1000
''')
    """
    config = load_athena_config()
    return athena_quire_base(sql_query=sql_query, config=config, dtype=dtype)
```

In [9]:
```
athena_quire('show tables')
```

```
EngineExecutionTime:      0.0 secondsDataScannedInBytes:      0.0 MB
EngineExecutionTime:    0.365 secondsDataScannedInBytes:      0.0 MB
```

Out[9]:

	m_drug
0	m_patient
1	t_disease
2	t_hisdrug
3	t_hisexamlab
4	t_hisradio
5	t_ope
6	t_ranote
7	t_rapanel

Is the data correct?

Easy to review

DATACK

Is there a disease other than RA?

```
In [12]: dis.disease.value_counts()
```

```
Out[12]: RA                    2327
         PMR                     37
         SLE                      9
         SjS                      8
         SSc                      5
         PM                       3
         MCTD                     3
         Behcet                   2
         DMy                      2
         IgG4RD                   1
         潰瘍性大腸炎関連関節炎          1
         Name: disease, dtype: int64
```

```
In [66]: # 2020年について、各患者で一番古い処方日を選ぶ
         first_odrymd_2020 = drug_2020.groupby('pid', as_index=False).odrymd.min()
         first_odrymd_2020.head(3)
```

Out[66]:

	pid	odrymd
0	341	2020-06-18
1	383	2020-01-22
2	634	2020-03-25

Is the period correct?

```
In [72]: # 同一患者で2020年で一番古い日の処方データに、複数種類の薬剤が含まれている場合の取り扱いについて
         # 00(Bio), 10(MTX), 20(DMARDs), 30(ステロイド), 40(NSAIDs), 99(その他)の順に優先する
         # JAKは分類していない。
         # 例えばBioとMTXが同時に処方されていた場合、その患者はBioとして集計する
         # 下記は分類ごとの処方数
         first_drug_2020.groupby('pid').bunrui.min().value_counts()
```

```
Out[72]: 10    1248
         20     535
         30     325
         0      314
         1       65
         99      29
         Name: bunrui, dtype: int64
```

Is the condition correct?

Today's summary

- We need to clarify the purpose of RWD utilization.

- Challenges in utilizing RWD & Solution
  - Insufficient human resources and organization
  - Quality and Quantity of RWD
  - Regulation

- Four Keys to Organization Building

- Practices
  - Case: Datack
    - organization
    - workflow
  - Extraction Request Form
  - Code Set
  - Analysis environment

My beliefs

DATACK

Data can lead society in the right direction.