

MTA2023 (富良野)

# データベース活用のための人材育成 技術者育成の視点から

鈴木 英夫  
一般社団法人SDMコンソーシアム  
株式会社MoDeL  
千葉大学

*SDM*  
Consortium

©登録商標第6025526号

## 診療データの記録と検索の違い

### 診療データの記録

- 患者単位で記録・検索
- コードで記録・名称で表示
- 発生時刻で記録・時系列に表示
- 集計情報は不要
- 過去からの経過が重要
- 対象トランザクション数<10000
- 対象テーブル数>1000
- DB機能：挿入、更新、削除
- 検索機能：定型

### 診療データの検索

- 項目単位で検索
- 名称で検索
- 時間を条件として検索
- 集計情報は必要
- 過去からの変化が重要
- 対象トランザクション数>100000
- 対象テーブル数<100
- DB機能：検索
- 検索機能：非定型

## データベースのジレンマ

データベースは大量のデータを記録し、高速に検索できるしくみである

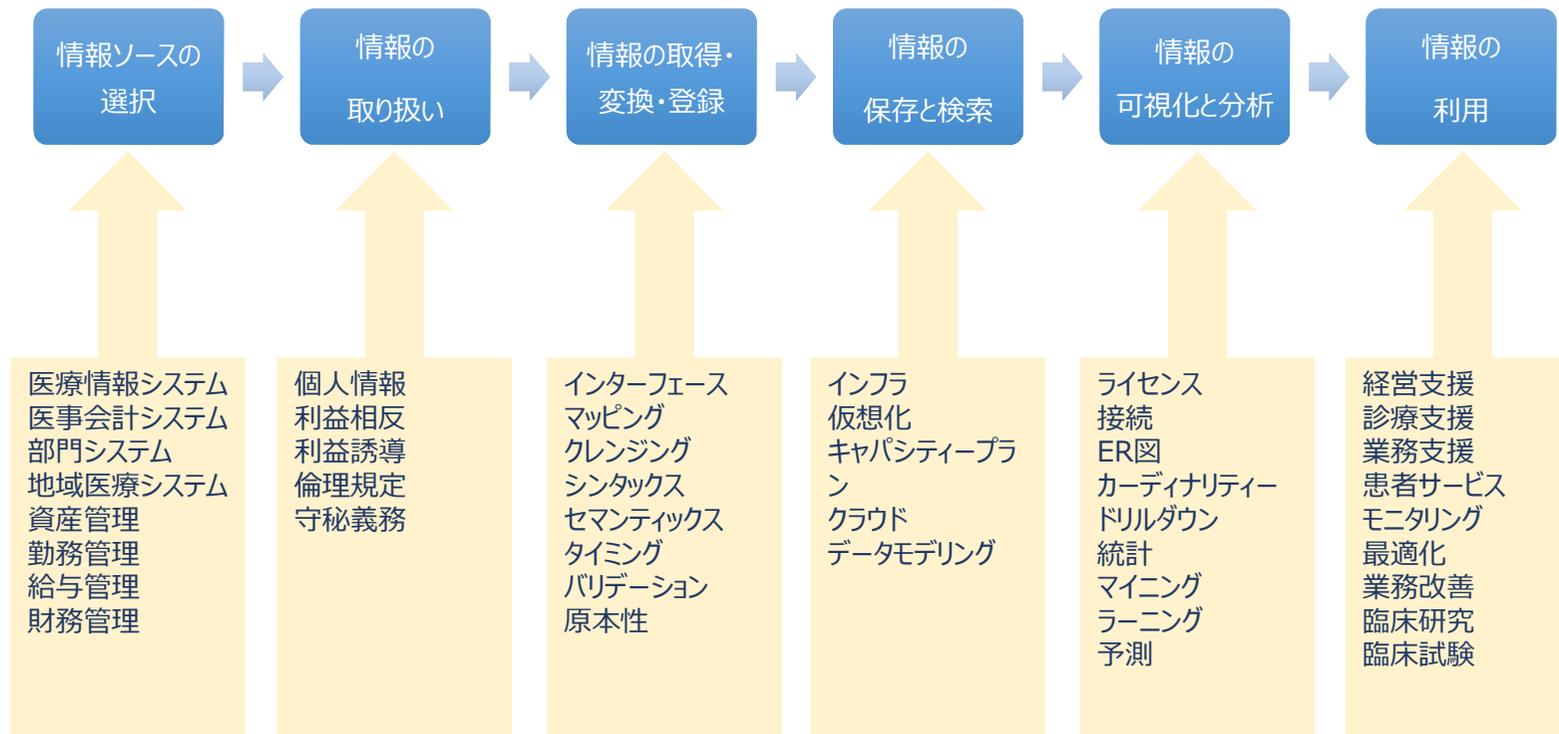
- データを記録する
  - 発生源のすべてのデータを特定の定義のもとで記録する
  - データベースを設計するのはアプリケーションを提供するベンダーである
  - アプリケーションは特定の目的のために開発され、ベンダーにより提供される
  - 特定のアプリケーションにより記録されたデータは同じアプリケーションだけが抽出可能である
- 記録したデータを検索する
  - 記録されたデータを検索し抽出する必要があるのはユーザーである
  - 特定のアプリケーションで記録されたデータの構造はユーザーには開示されない
  - ユーザーはベンダーが提供する特定のアプリケーションを通して、データの検索と抽出および可視化が可能となる
- データの管理義務の責任があるユーザーがデータを検索・抽出することができない

## どうすればユーザーがデータを管理できるか

### 汎用的なデータウェアハウスを構築し、すべてのデータを抽出管理する

- データの棚卸
  - それぞれのアプリケーションが有するデータの目録を作成する（種別、期間、容量など）
  - 各アプリケーションのデータベース定義書を取得して、版管理を行う
  - データ出力機能があるアプリケーションの場合は出力仕様（定義書）を取得する
- データ・サイエンティスト、データ・エンジニアの育成
  - 定期的にデータ目録の更新を行う管理者のスキル育成
  - データベースに関する基礎知識の習得
  - 各アプリケーションからの出力可能なデータの保存と管理、抽出
  - 構築したデータ・ウェアハウスを管理し、BIツールなどで可視化するスキル育成

## データ抽出から利用までのステップ



## データ・サイエンティスト、データ・エンジニアに必要な基本知識とスキル

- **医療の質に関する知識**
  - ストラクチャー・プロセス・アウトカム
- **医療に関する知識**
  - 診療録、診療報酬、プロブレムなど
- **業務フローと用語に関する知識**
  - 診察、オーダー、検査、処方など
- **医療情報システムに関する知識**
  - 電子カルテ、オーダー、看護、部門、医事など
- **データベースに関する知識**
  - RDB、SQL、ODBCなど
- **データ抽出、可視化ツールに関する操作スキル**
  - DWH、BIツールなど
- **分析に関する知識**
  - QI、PDCA、統計分析、マイニング、バランススコアカード
- **関連法に関する知識**
  - 個人情報保護法、医療情報の安全管理に関するガイドライン、病院機能評価

診療情報管理士

医療情報技師

## データサイエンティスト・エンジニアにおける重要なリテラシー

シンタックス = **文法**  
セマンティックス = **意味**

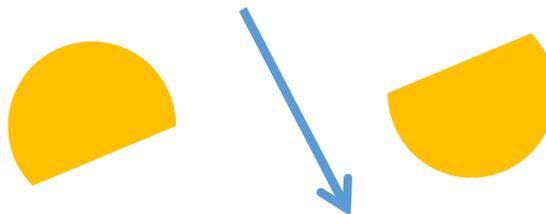
校正 = **文法**のエラーを直すこと  
校閲 = **意味**のエラーを直すこと

医療情報技師

診療情報管理士

検体検査と病理検査は同列？

上弦の月はどっち？

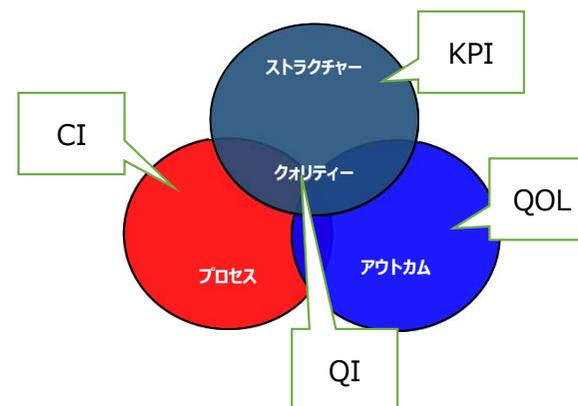


## 医療の質に関する知識：医療の質の評価方法

- 医療の質は、ストラクチャー、プロセス、アウトカムのトータルで測ることができる
  - ストラクチャーを測るにはKPI (Key Performance Indicator) を用いる
  - プロセスを測るにはCI (Clinical Indicator) を用いる
- アウトカムを測るには
  - 費用効果分析
    - 臨床におけるアウトカム（転帰、生存率など）とその費用の分析
  - 費用最小化分析
    - プロセスが違ってもアウトカムが同じ場合、その費用の比較分析
  - 費用便益分析
    - アウトカムに対する支払意思額の分析（他業界との比較が可能）
  - **費用効用分析**
    - **QOL（主観的価値観）とその費用の分析**
- QOLとは、フィジカル、メンタル、ソーシャルの主観的評価のトータルな評価

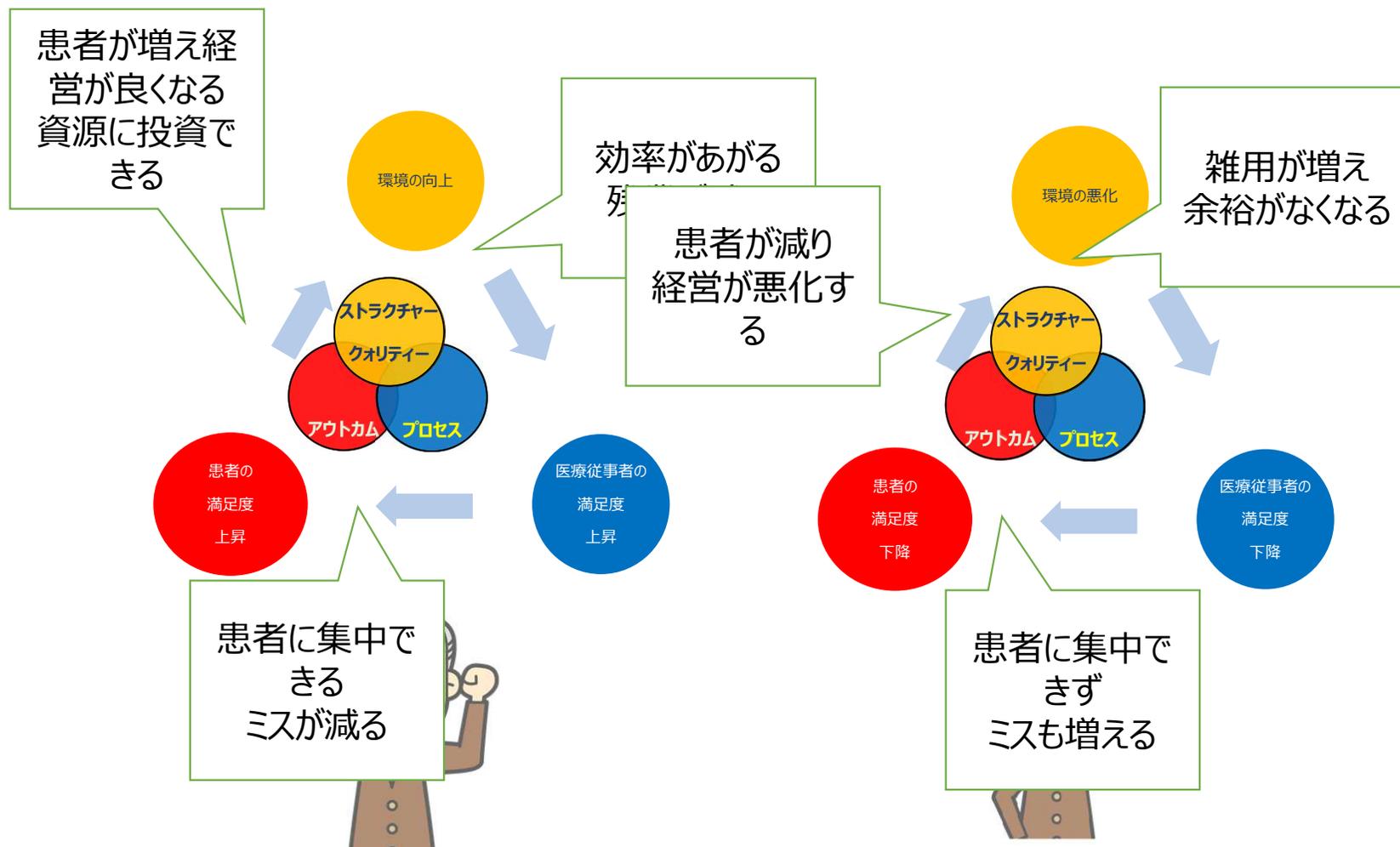
ADL(Activities of Daily Living)  
日常生活動作の客観的評価

QOL(Quality of Life)  
全人的な価値観の主観的評価

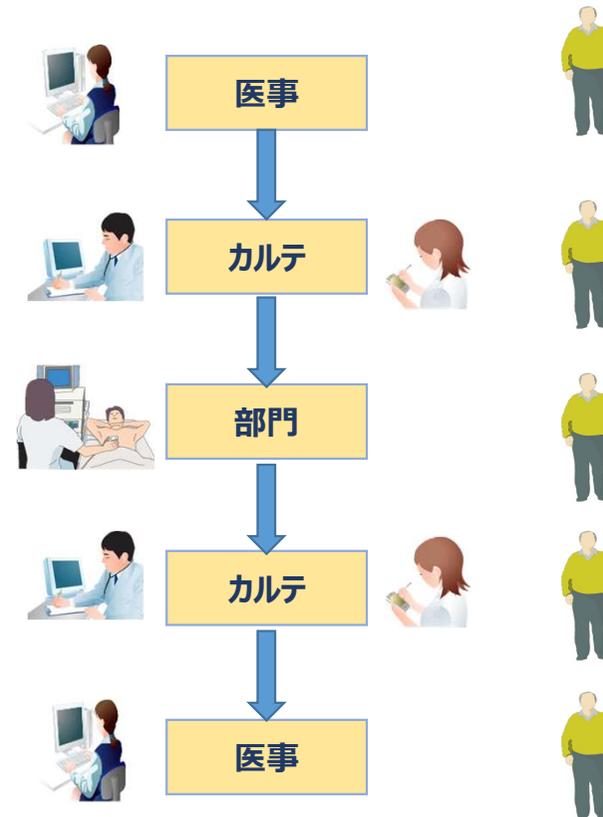
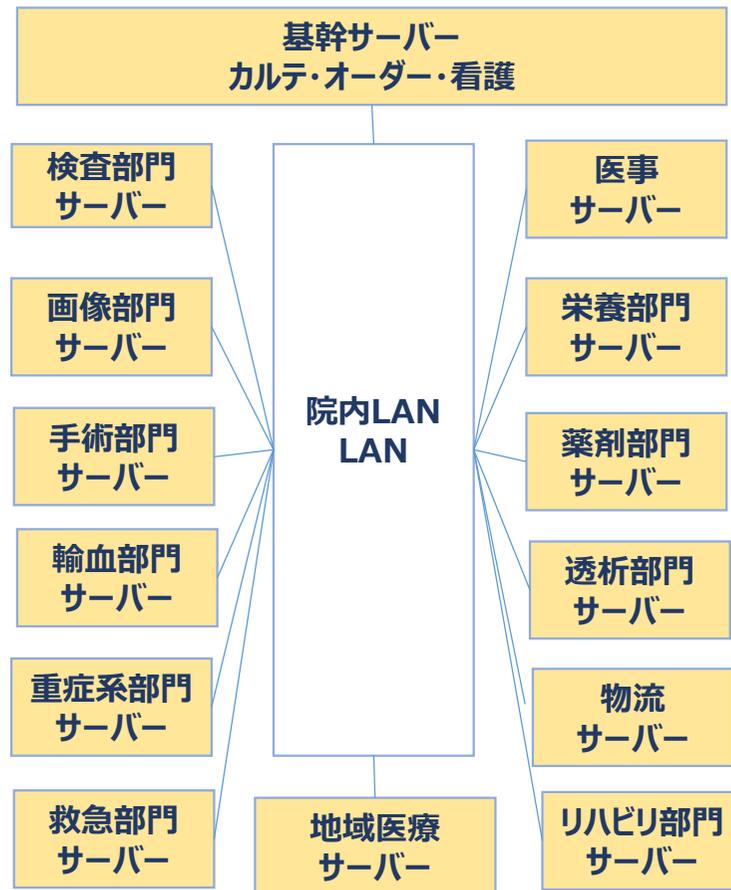


QOLとは、全人的な観点からの尺度である  
私は、寝たきりでも良いので長く生きたい  
私は、病気と治療と痛みに対する恐怖がないようにしたい  
私は、多少寿命が縮んでも、普通の生活を維持したい

# 医療の質に関する知識：好循環と悪循環

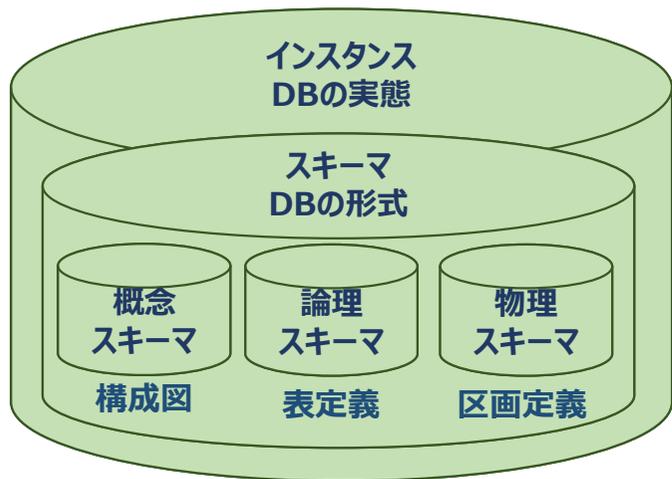


## 医療情報システムに関する知識：情報の流れ（外来）



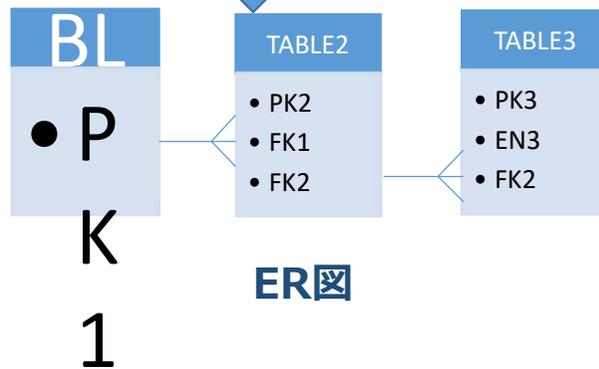
# データベースに関する知識：関係データベース（RDB）の構造

リレーショナルデータベース（RDB）：テーブルとテーブル間の関係を有したデータベースモデル



**論理スキーマ**  
・テーブル(表)  
・エンティティ (項目)  
・リレーション (関係)

項目名 (別名)  
型  
プライマリキー  
NULL+DEFAULT  
インデックス  
外部キー



## 用語

SQL：RDBにおけるデータの操作や検索を行うデータベース言語  
ER図：RDBにおいてテーブル間の関連を表すモデル図

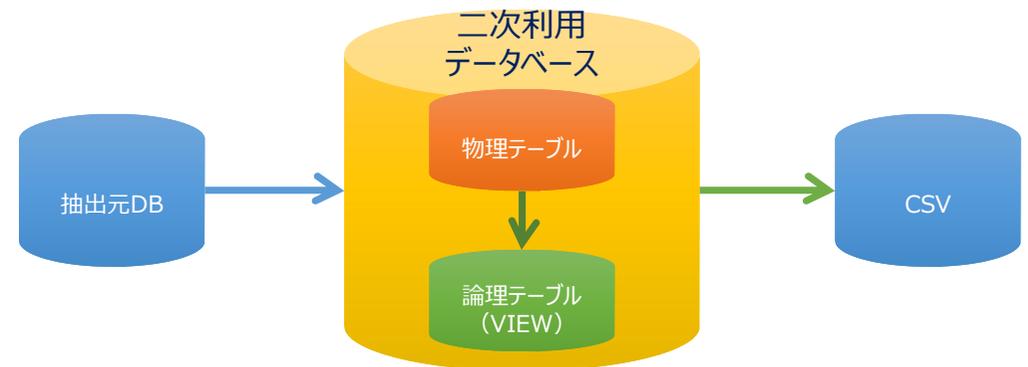
## データベースに関する知識：二次利用データベース設計手順

### • データベース設計手順



### データ抽出の手順

- ① データベースマネジメントシステム(DBMS)の選択
- ② データベース環境構築
- ③ 関係DBモデル設計
- ④ 物理DBテーブル設計
- ⑤ 論理DBテーブル(VIEW)設計
- ⑥ テーブル生成
- ⑦ テーブル更新
- ⑧ バージョン管理



## データベースに関する知識：設計における注意点

### データの重複が起きない設計であるか

RDBにおいては、プライマリキーにより、重複を判断しているため、プライマリキーが一意とならなければならない

- 例 1 プライマリキーに更新日時を設定する→ETLをやり直した場合に、同じデータを違うデータとして登録  
臨床研究におけるダブルカウントとなる
- 例 2 プライマリキーに患者IDを設定する→1患者が複数IDを持つこともあるので、別症例として登録  
同一症例を別症例として登録となる

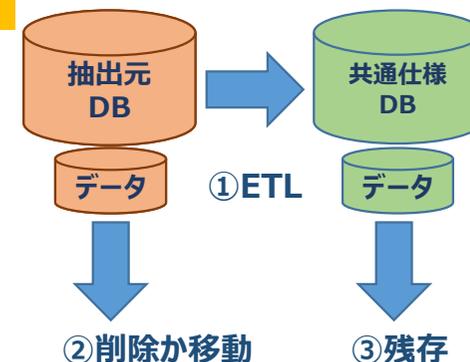
検体検査	検体検査
更新日時②	更新日時①
患者ID①	患者ID②
結果①	結果①

すべて同一患者・同一データ

### 削除データが存在しない設計であるか

ETL実行後、抽出元のデータが物理的に削除（物理削除）された場合に、削除されなければならない

- 例 救急患者が後から当院登録患者であることが判明し、移行して削除した→  
新規の患者IDで登録されたデータが残るため、同一症例を別症例として登録

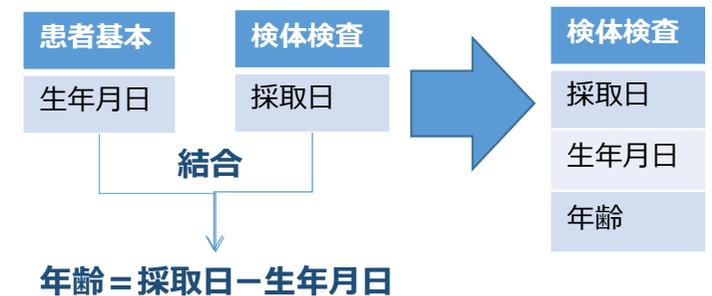


## データベースに関する知識：設計における注意点

### 検索の効率を考えた設計であるか

RDBにおいては、複数のテーブルを結合する負荷が大きいので、可能な限り非正規化を行う（階層を減らす）

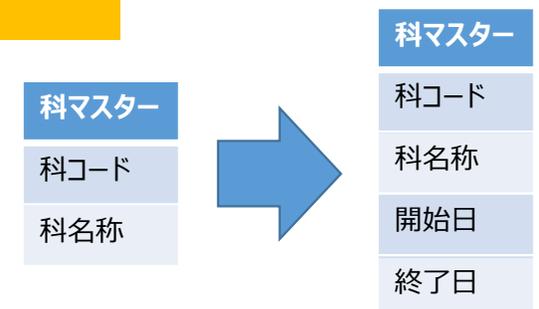
例 生年月日、性別を別テーブルで管理→検査時年齢や性別は、臨床研究において必須  
→年齢計算を行うために、必ずテーブル結合を行う必要がある



### マスターの世代管理がされているか

過去データのコードを新しいマスターで変換すると不整合が起きるのでマスターは世代管理が必要

例 診療科名が呼吸器科から呼吸器センターへ変更されたため、マスターは同一コードで名称を変更した  
→過去のデータを新しいマスターで変換したら、過去に存在しない呼吸器センターに変換された



## データベースに関する知識：設計における注意点

### データの欠損かどうかの確認

選択な項目にデータが存在するかどうかを確認する

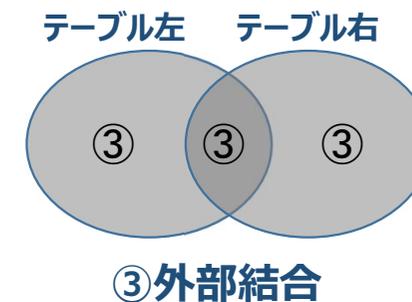
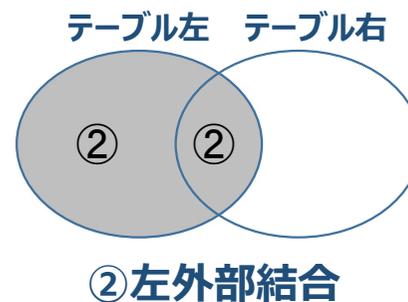
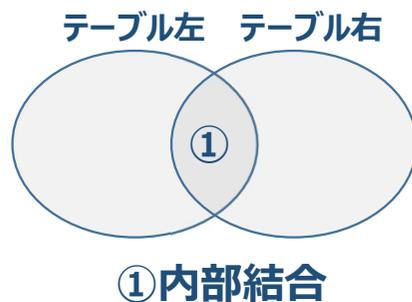
例 条件を指定した結果、一つの項目がすべて空白であったが、使用されていないデータなのかどうかの判断が  
つかなかった→項目は定義されているが、ETLの対象に入っていない場合もある

例の対処方法

設計の際にルールを決めてあるかを調べる。もしルールが無ければ、SQLにより全体を調べる必要がある

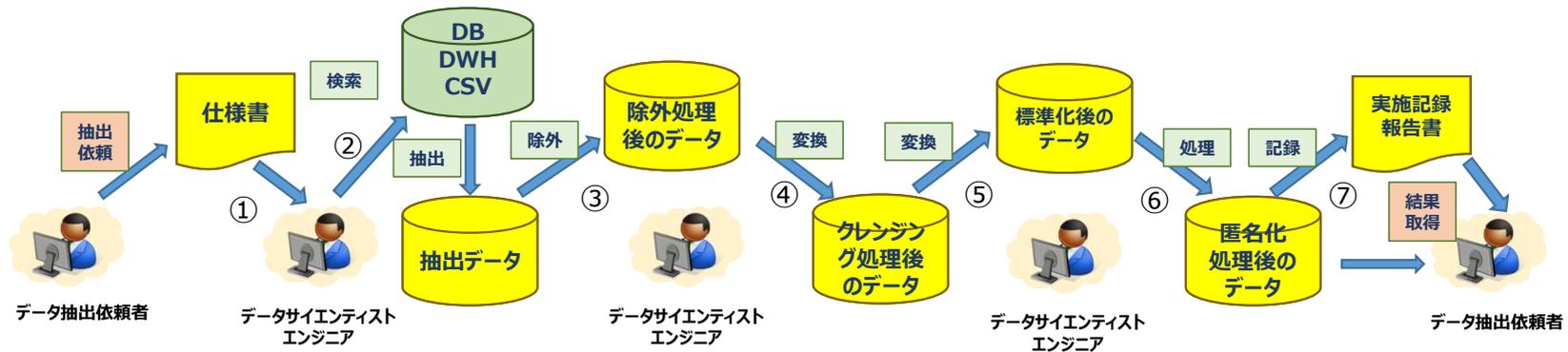
### 条件検索の順番を変えた場合に結果や検索速度が変わる

表結合の場合や、項目の選択順により、結果や検索速度が変わる



## データ抽出に関する知識：データ抽出の詳細手順

### 二次利用におけるデータ抽出から出力までの流れ

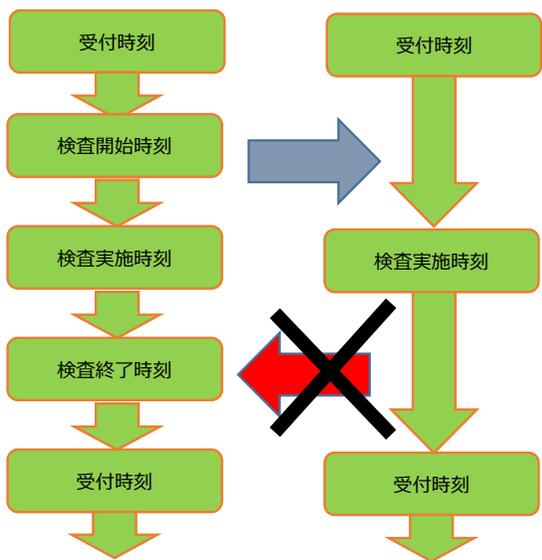


#### データ抽出の手順

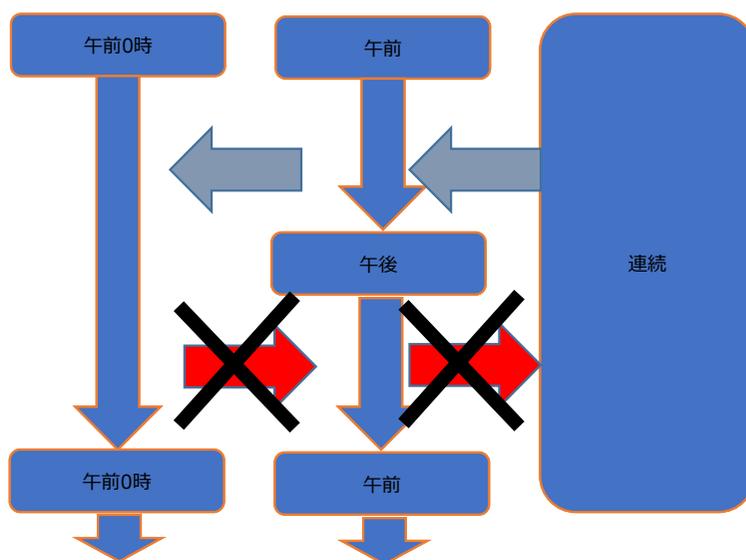
- ① 抽出の依頼から、対象データの抽出条件を抽出
- ② 対象データの存在する各社DBあるいはDWHなどから抽出条件により検索して保存する
- ③ 抽出したデータを確認し欠損値、異常値のデータを除外、出力形式へ変換して保存する
- ④ データのクレンジング処理（表記、単位などの統一など）
- ⑤ データの標準化処理（名称、コードなどの標準化）
- ⑥ 個人情報を匿名化するなど、個人情報除去処理して保存する
- ⑦ 各処理の手順、処理方法、各処理前後のデータ件数、匿名化処理方法などを記録する

# データ抽出に関する知識：粒度、頻度、精度

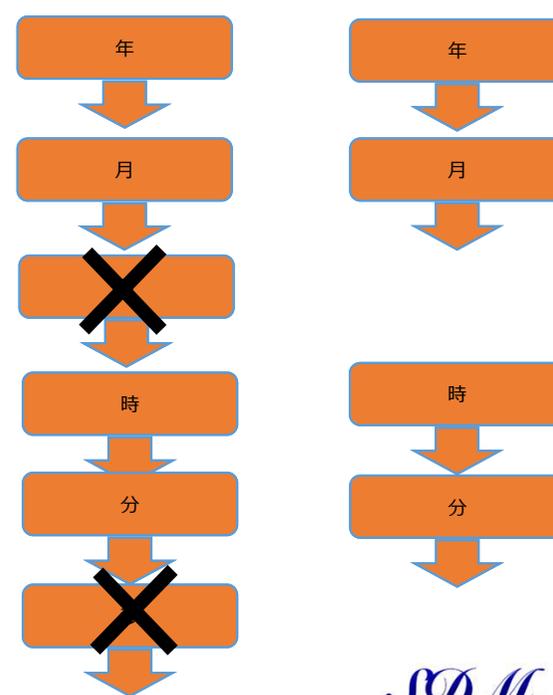
粒度



頻度

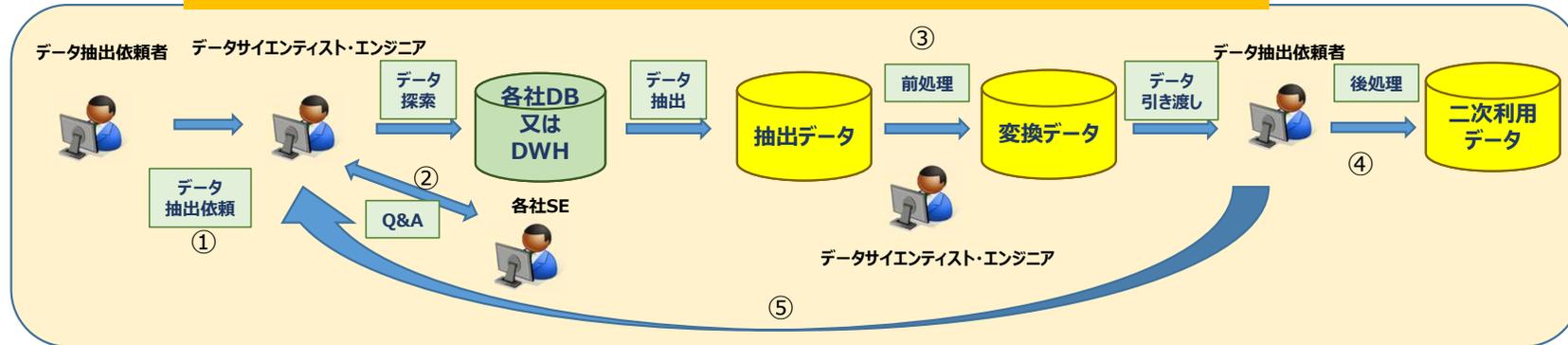


精度



## データ抽出に関する知識：データ抽出の実際（共通仕様DWHの有無）

共通化されていない診療データを直接対象とした場合、二次利用までに多くのプロセスが発生する



### 現状の問題点

- ① データ抽出依頼者からの依頼を理解し、要求データを特定するのが難しい
- ② 各社DBあるいはDWHは、各社独自方式であるため、データ抽出には、各社の支援が必要
- ③ 抽出したデータの形式から、依頼された形式への変換、および個人情報除去などの前処理が必要（前処理）
- ④ 引き渡されたデータを臨床研究に利用するためにデータ変換、並べ替え、結合などが必要（後処理）
- ⑤ 依頼したデータが不十分だった場合、再度依頼から行う必要がある

データ抽出から二次利用までのプロセスを削減可能



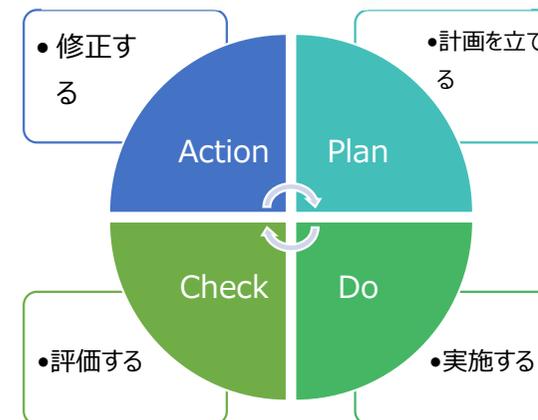
## 分析に関する知識：指標、可視化、分析は、何のために必要なのか

課題を明確にし：可視化

原因を見つけ出し：分析

解決案を考え：指標

改善する：PDCA



## データサイエンティスト・エンジニアに期待されること

### • 依頼内容の理解、課題の抽出

- 目的、データ抽出範囲、データ加工、出力範囲の決定
- 課題の抽出、分析方法、出力方法の決定

### • データ抽出・変換

- 医療情報システム、経営管理システムなどからデータを抽出し、変換して分析・可視化ツールに読み込ませる

### • 可視化・分析

- ツールを利用して、集計、表、グラフなどを作成する
- ツールを利用して、指標、統計、などから分析結果を得る

### • 出力、報告

- 可視化、分析の結果を出力する
- 報告書、プレゼン資料を作成する

## 病院の組織と職種（現状）とデータサイエンティスト・エンジニアの立ち位置

データサイエンティスト・エンジニアの組織、資格はない

スキルを持てば兼務が可能

需要 > 供給

### • 病院の組織

- 管理部門
- 診療部
- 看護部
- 薬剤科
- 検査科
- 放射線科
- リハビリテーション科
- 栄養科
- 事務部門
- 医事科
- **情報管理部門**
  - 情報システム管理（医療情報システムの運営・管理）
  - 診療情報の管理（診療録の点検、病名管理）
- 医療相談室
- **経営企画室**（集計、分析、戦略立案）
- 委員会

データサイエンティスト

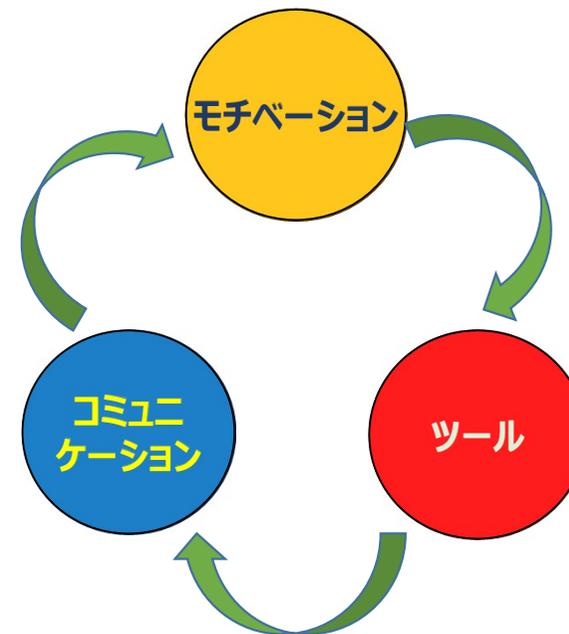
### • 医療関連の資格

- 国家資格
  - 医師、歯科医師
  - 保健師、助産師、看護師
  - 診療放射線技師、臨床検査技師
  - 理学療法士、作業療法士
  - 視能訓練士、臨床工学技士、義肢装具士、歯科技工士
  - 歯科衛生士、救急救命士、薬剤師
  - 言語聴覚士、管理栄養士
  - 社会福祉士、介護福祉士、精神保健福祉士、柔道整復師、あん摩マッサージ指圧師、鍼灸師
- 都道府県認定資格
  - 准看護師、ホームヘルパー、介護支援専門員
- 医療関係団体資格
  - 全日本病院きゅかい認定総合医（全日本病院協会）
  - 病院管理士（全日本病院協会）
  - **診療情報管理士**（日本病院会）
  - 日本糖尿病療養指導士（日本糖尿病療養指導士認定機構）
  - 消化器内視鏡技師（日本消化器内視鏡学会）
  - 臨床心理士（日本臨床心理士資格認定協会）
  - **医療情報技師**（日本医療情報学会）
  - 専門医（医学会認定）

出典：全日本病院協会HP

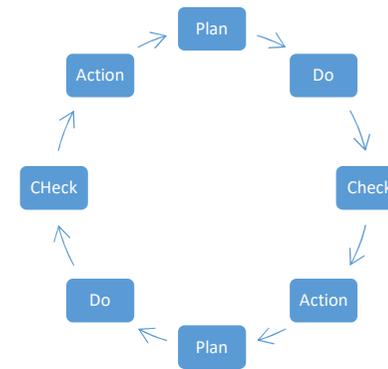
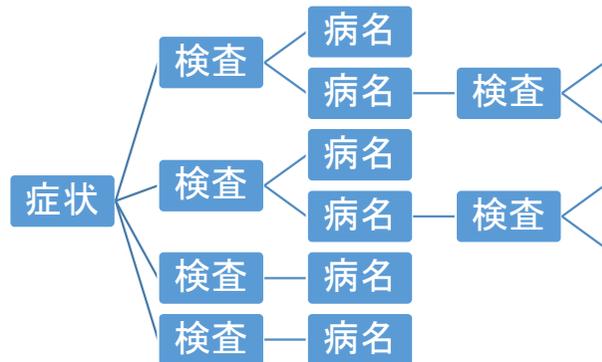
## データサイエンティスト・エンジニアを育てるための3つの要素

- モチベーションを与える
  - 依頼する・必要とする
  - 感謝する・ほめる
  - インセンティブを与える
- BIツールの導入
  - 依頼結果の可視化
  - 期待を超える成果
  - 新しい知見
- コミュニケーションの機会を増やす
  - ディスカッションし課題を解決する
  - 重要度、優先順位を話し合う
  - 対策を一緒に考える

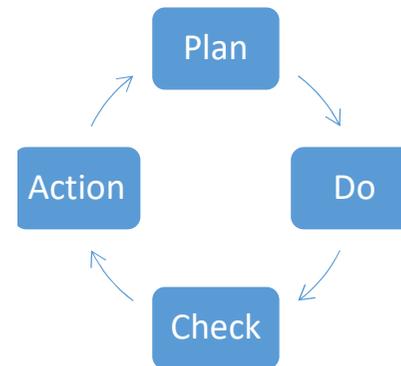
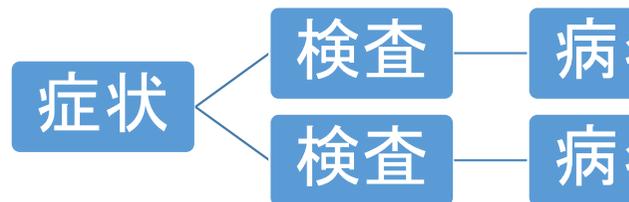


# 医療の質に関する知識：臨床経験の差

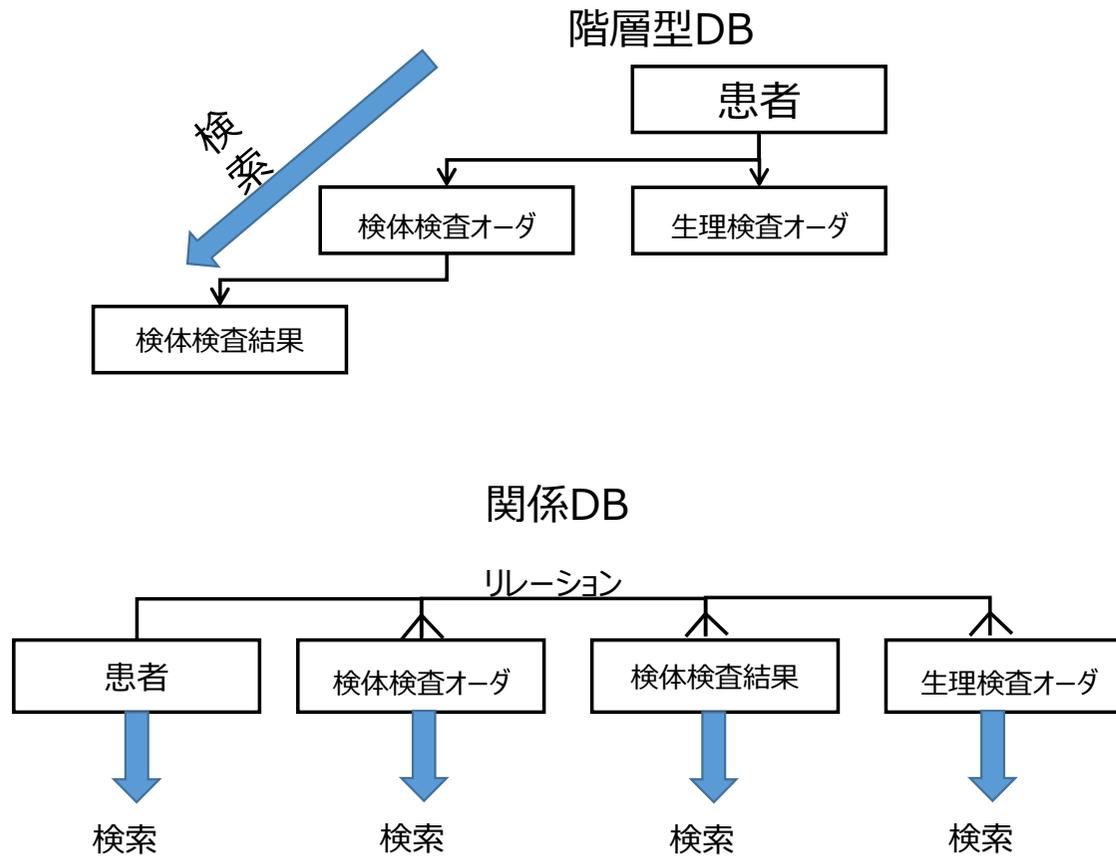
臨床経験が少ない場合



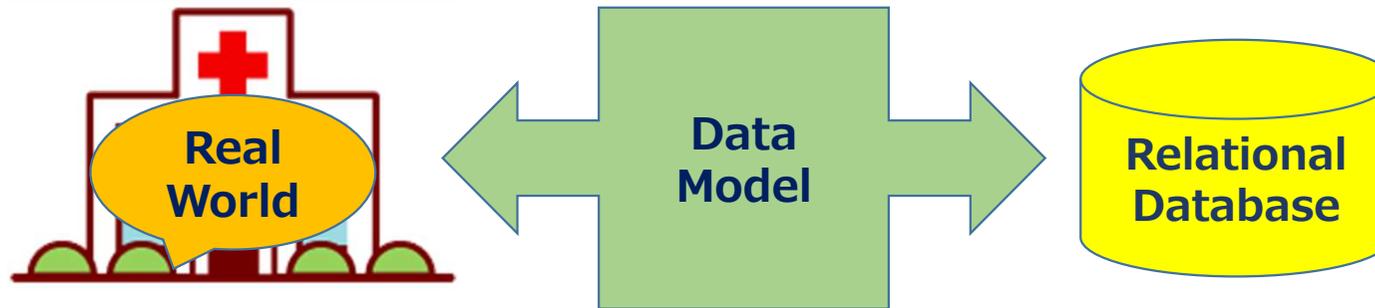
臨床経験が豊富な場合



## データベースに関する知識：階層型DBと関係型DB



## データベースに関する知識：データモデル

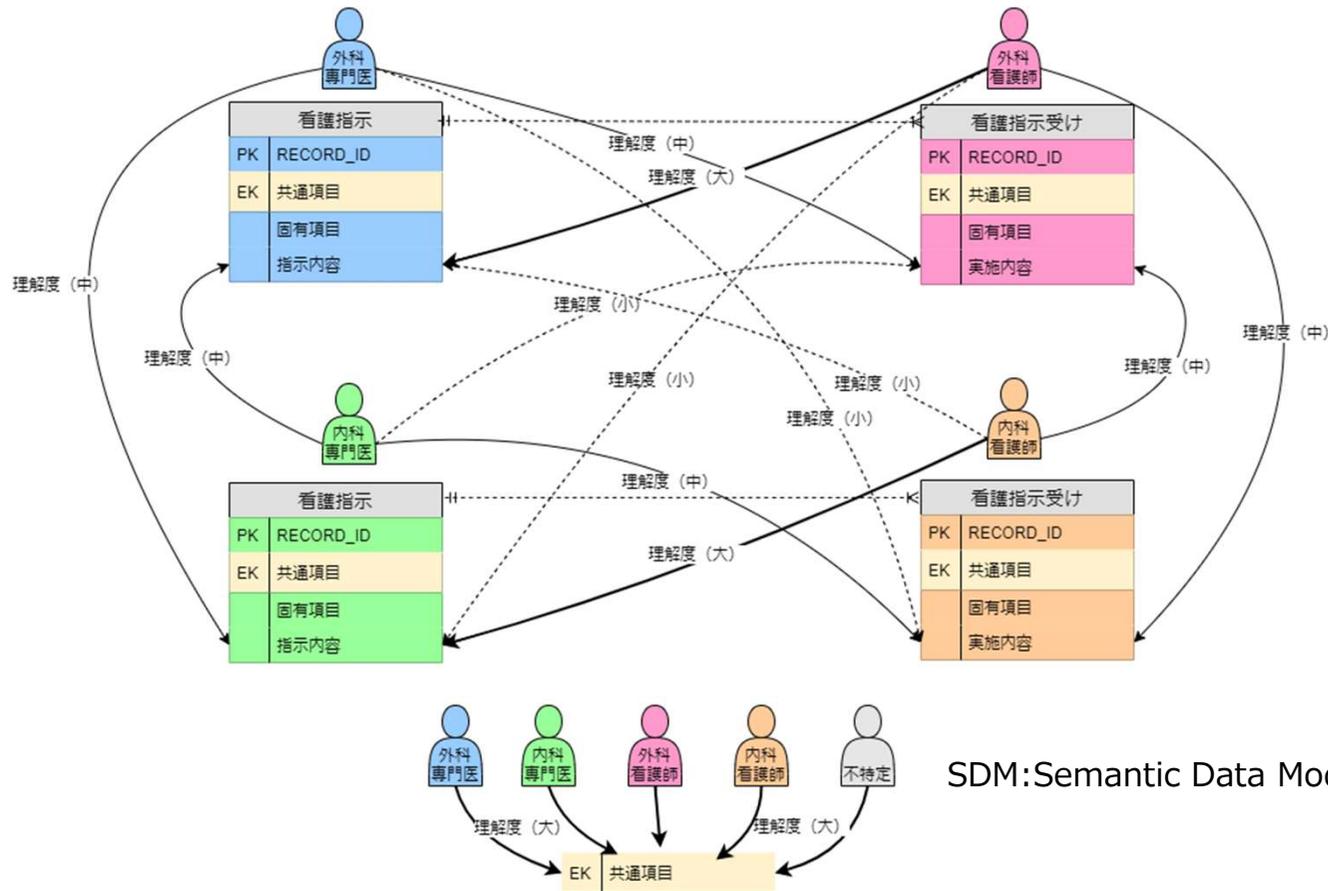


- データモデルとは、実世界で発生するデータ（リアルワールドデータ）を物理ストレージ上に保存するための手法のこと
  - 階層型データモデル、ネットワーク型データモデル、関係データモデル、オブジェクトデータモデル
- セマンティックデータモデル
  - RDBにおいて、リアルワールドデータとデータ間の関係性、およびデータの意味を概念的に一致させるモデル

## データベースに関する知識：SDMのデータモデル

### SDMで設計したデータモデルは、セマンティックデータモデリングを参考にして開発したヘルスケア共通データモデル（CDM）

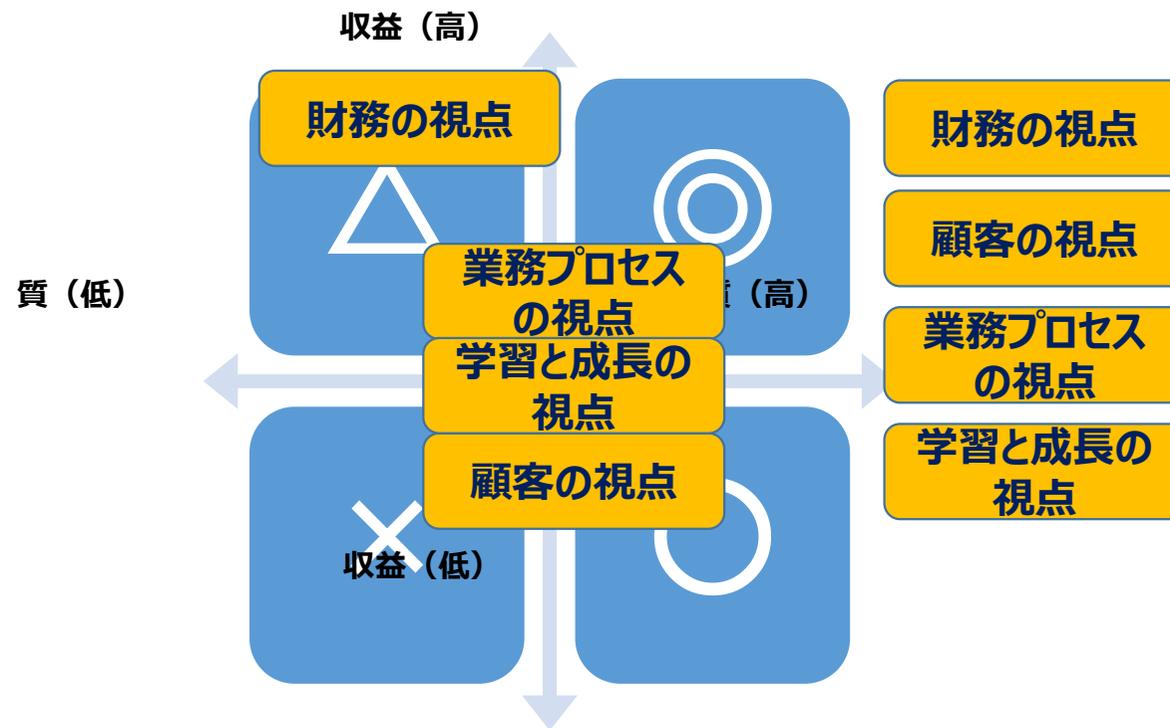
SDMにおける共通項目（一般用語）と固有項目（専門用語）の理解度



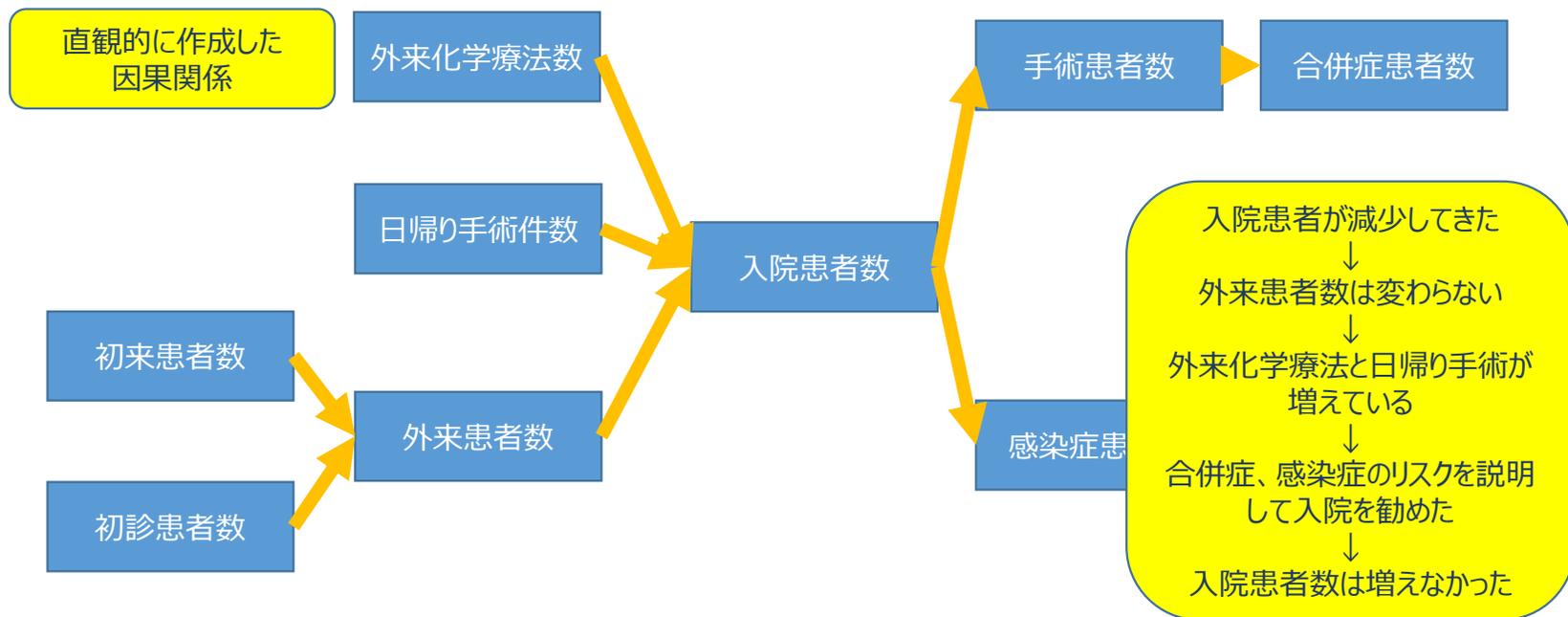
SDM: Semantic Data Model (一般社団法人SDMコンソーシアム)

## 分析に関する知識：医療におけるBSCのための課題選択

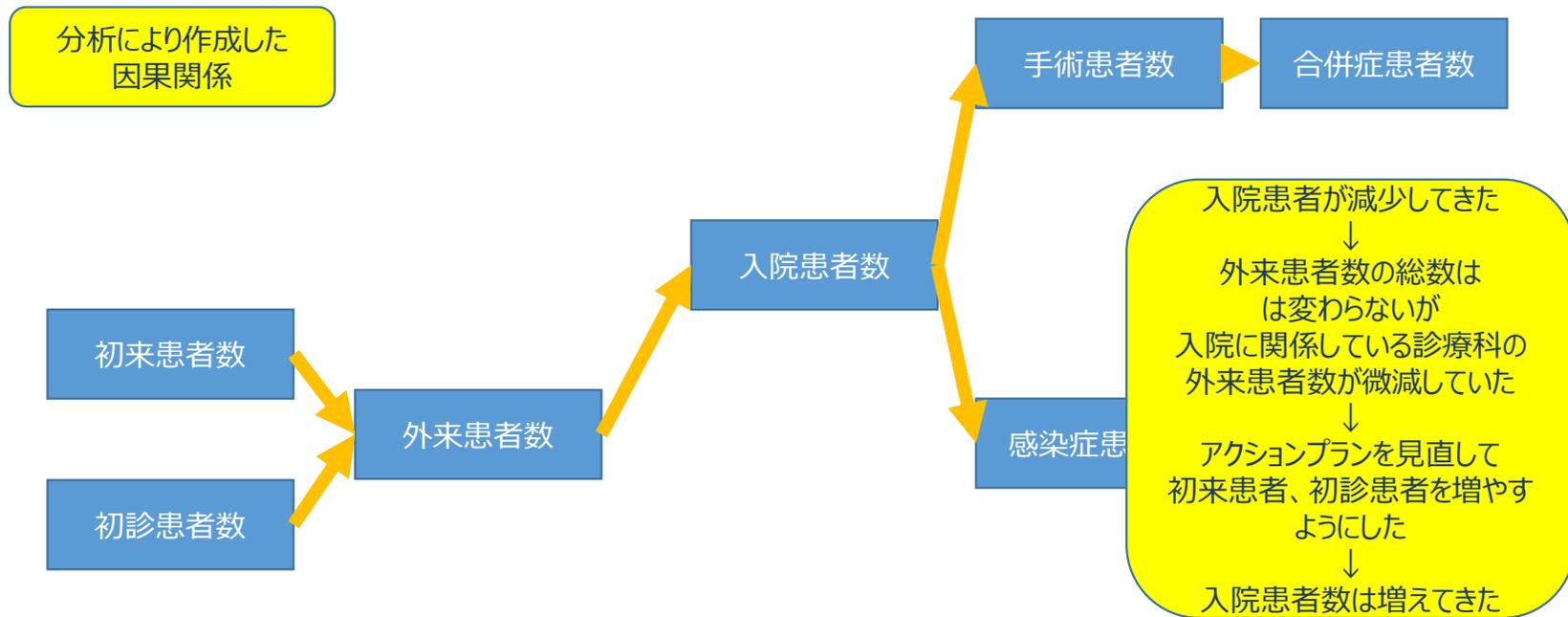
量的な基準だけではなく、質的な基準も加えて、課題の優先順位を決める



## 分析に関する知識：BSCにおける指標の因果関係の間違った定義



## 分析に関する知識：BSCにおける指標の因果関係の正しい定義



## 個人情報保護法（基本法） 次世代医療基盤法（特別法） 改正個人情報保護法

改正個人情報保護法（基本法）が施行され、病歴などの要配慮個人情報に関してオプトアウト（第三者提供）には、完全匿名化が必須となりました

6か月以下の懲役ま  
たは  
30万円以下の罰金



創薬のための治験や医療機器開発のための臨床試験には、病歴（要配慮個人情報）を含む診療録を必要とするので、完全匿名化が義務付けられることとなります

完全匿名化は技術的に不可能であることが有識者会議により証明されました

要配慮個人情報の第三者提供を可能とするためには「丁寧なオプトアウト」を必要とするという「医療分野の研究開発に資するための匿名加工医療情報に関する法律」（次世代医療基盤法：ビッグデータ法）が新たに特別法として施行されました

改正個人情報保護法において、仮名化が加わりました