

# 医学概念・知識連結データベースの仕様と構築・管理方法

医学概念・知識連結データベース作業班

2024年10月7日

医学概念・知識連結データベースの仕様と構築・管理方法.....	1
概要.....	3
はじめに.....	3
仕様.....	3
2つのエントリ情報.....	3
属性情報(基本).....	4
属性情報(サブデータ固有).....	5
病名データ固有の属性情報.....	5
医薬品データ固有の属性情報.....	5
部位データ固有の属性情報.....	5
検査データ固有の属性情報.....	5
リンク情報.....	6
リンク先リソース候補.....	6
リンク先リソースのライセンス.....	7
信頼度フラグ.....	8
構築方法.....	8
初期構築.....	8
サイズ(用語数).....	9
未対応部分.....	9
データ管理の方針.....	9
項目推定.....	10
管理方法.....	11
メンテナンス手順.....	12
共通手順.....	12
信頼度フラグ.....	13
メンテナンスで利用するタグ.....	13
病名データ(DISEASE).....	13
実施内容.....	13
実施手順.....	13
基本情報.....	13
出現形.....	14
出現形_flag.....	14
出現形よみ.....	14
出現形よみ_flag.....	14
正規形.....	14
正規形_flag.....	15
属性情報.....	15
TYPE:否定.....	15
リンク情報.....	15
TO:ICD-10.....	15

TO:ICD-10_flag.....	15
TO:Lilak.....	15
TO:Lilak_flag.....	16
TO:日本医学会医学用語辞典.....	16
TO:日本医学会医学用語辞典_flag.....	16
医薬品データ (MEDICATION).....	16
実施内容.....	16
実施手順.....	16
部位データ (BODY).....	18
実施内容.....	18
実施手順.....	18
基本情報.....	18
出現形.....	18
出現形_flag.....	19
出現形よみ.....	19
出現形よみ_flag.....	19
正規形.....	19
正規形_ICD-11 (WIP).....	20
正規形_flag.....	20
TREE.....	21
TREE_flag.....	21
検査データ (TEST).....	21
実施内容.....	21
実施手順.....	21
基本情報.....	21
出現形.....	21
出現形_flag.....	21
正規形.....	21
正規形_flag.....	22
TREE.....	22
TREE_flag.....	22
おわりに.....	22
Appendix.....	22
リンク先リソースごとのライセンス情報.....	22
メンテナンス作業時のQA.....	24
病名データ.....	24
医薬品データ.....	26
検査データ.....	36

## 概要

SIP「統合型ヘルスケアシステムの構築」において、医療情報の統合・分析ために開発された医学概念・知識連結データベースの構築、管理方法について述べる。大規模かつ、変化し続ける複数のリソースと対応をとるために、様々な工夫がなされている。なお、本データは現在も構築中であり、本稿の内容も作業途中の内容を含むことに留意されたい。

## はじめに

SIP「統合型ヘルスケアシステムの構築」のミッションである「知識発見」と「医療提供」の循環を実現するためには、電子カルテ(EHR)や患者記録(PHR)など様々なテキストから医療知識を抽出し、統合する必要がある。そこで、我々は、既存の医療・医学関連の辞書やリソースをまとめた大規模な医学概念・知識連結データベースを構築・維持している。

すでに、海外には、SNOMED-CT, UMLSなどの巨大オントロジーがすでに存在している。しかし、維持管理に膨大なメンテナンスコストが必要とされている。そこで、我々は、近年生成AIなどの自然言語処理技術が急速に発展していることを踏まえ、大規模ではあるができるだけ、省コストで運用できるように設計を行った。

本データは、以下の4つのサブデータから構成される。

1. 病名データ: 症状や病名を収載。ICD-10や11, 標準病名に紐付け可能である。
2. 医薬品データ: 医薬品名を収載。一般名, ATCコード, KEGGコードに紐付け可能である。
3. 部位データ: 人体部分表現を収載。代表的な解剖学辞書の表現に紐付け可能である。
4. 検査データ: 検査表現を収載。代表的な臨床検査項目の分類に紐付け可能である。

本稿では、これら4つのデータの仕様と半自動化された構築方法および、人手によるメンテナンス方法を述べる。

## 仕様

### 2つのエントリ情報

本データは1つの出現形の表記(以降、この1つの表記のことをエントリと呼ぶ。)に対して、属性情報とリンク情報という2つの情報を付与している。それぞれの情報内では、各項目ごとに信頼度を付与している。

2つの情報は以下の通りである。

- 属性情報: 出現形に関する情報を付与する。例えば、よみがな、出現頻度などが属性情報にあたる。サブデータによっては追加で固有の属性情報を付与する。
- リンク情報: 出現形を別の用語体系に紐づけるためのリンク情報を付与する。例えば、本データでは、各出現形について、対応するICD-10のコードを記載する。この対応関係を、リンク情報と呼ぶ。リンク情報は、同概念であるかどうかを判定するために用いることを想定している。例えば、2つの出現形が同じICD-10へのリンク情報を持っていると、ICD-10の観点からは2つの出現形は同じ概念(同義)とみなせる、と考える。このようなリンク先を複数保持しており、様々な観点からの同義を扱えるのが、本データの特徴である。

## 属性情報(基本)

すべてのサブデータに共通で付与する属性情報は下表の通りである。サブデータによっては、次節で述べる固有の属性情報を加える。

	ラベル	概要	タイプ
基本情報	ID	出現形に対応する一意に割り当てられたID.	文字列
	出現形	医療テキストから抽出した、用語の出現形(ユニーク)	文字列
	出現形よみ	出現形のよみ	文字列
	正規形	<p>出現形を正規化した表記。本データは、既存リソースへのリンクが目的であり、正しい正規形や標準的な表現を提案するものではないことに注意されたい。あくまで、ビューアー上での表示や、概念の理解の補助を目的に正規形を設けている。</p> <p>サブデータごとに正規形のベースは異なる。それぞれの正規形のベースは以下の通りである。</p> <ul style="list-style-type: none"> <li>● 病名データ:標準病名マスター病名</li> <li>● 医薬品データ:一般名</li> <li>● 部位データ:ICD-11のエクステンションコードをベースとした表現</li> <li>● 検査データ:医師国家試験出題基準の「主な検査項目の表記」をベースにした表現</li> </ul> <p>出現形に対して複数の正規形が該当する場合、最大3つまでを半角セミコロン (;) で接続し列挙する。</p>	文字列
	TREE	最大4桁の記号による分類。ビューアー使用時のツリー構造の表示に用いる。	文字列
翻訳情報	出現形EN	出現形の英語訳。	文字列
	出現形ZH	出現形の中国語訳。	文字列
関係情報	REL:IS-A	上位下位関係。	文字列
	REL:PART-OF	部分全体関係。	文字列
	REL:CAUSAL	因果関係。	文字列
その他	用語難易度	用語の難易度	整数
	YEAR	エントリの収集年度	整数
	FREQ:GLOBAL	大規模医療テキストに基づく頻度	整数
	FREQ:LOCAL	診療記録に基づく頻度	整数
	SOURCE	エントリの抽出元の医療テキスト	文字列

## 属性情報(サブデータ固有)

サブデータごとに追加で付与する属性情報を示す。

### 病名データ固有の属性情報

名称	ラベル	概要	タイプ
症状フラグ	TYPE:症状	出現形が症状を表す表現かどうかのフラグ。 「寒気」などは1,「がん」などの症状でなく,病名は0.	ブーリアン
診断名フラグ	TYPE:診断名	出現形が病名(鑑別されるような診断名)かどうかのフラグ。「高血圧症」などの場合は,症状フラグ,診断名フラグともに1となる	ブーリアン
否定フラグ	TYPE:否定	出現形がある病名を否定する表現の場合,1とする.例: 非糖尿病.	ブーリアン

### 医薬品データ固有の属性情報

名称	ラベル	概要	タイプ
メーカー・製造会社名	TYPE:メーカー名	医薬品の製造メーカー名.	文字列

### 部位データ固有の属性情報

部位データには固有の属性情報は付与しない。

### 検査データ固有の属性情報

名称	ラベル	概要	タイプ
材料	TYPE:材料	検査材料の名称.臨床検査マスターの材料名称から引用.	文字列
測定法	TYPE:測定法	検査の測定法.臨床検査マスターの測定法名称から引用.	文字列
単位	TYPE:単位	検査結果値の単位.臨床検査マスターの参考単位から引用.	文字列
	TYPE:物質検査／臨床全般検査		
	TYPE:診療科		
	TYPE:施設		
	TYPE:基準値		
	TYPE:臨床判断値		
基準値	TYPE:基準値	標準的な値の範囲.	文字列

## リンク情報

### リンク先リソース候補

リンク情報の紐づけ先候補として、下表記載のリソースを選定している。リソース名のラベルは、TO:ICD-10のように、TOから開始される名称になっている。「病」「薬」「検」「部」列はそれぞれ病名データ、医薬品データ、検査データ、部位データの略称である。チェックマークはその行のリソースと紐付けられることを表す。1つのエントリが紐づけ先リソースの複数の項目に対応する場合、最大3つまでを半角セミコロン (;) で接続し列挙した。

リソース名	概要	リンク	病	薬	検	部
TO:日本医学会医学用語辞典	日本医学会が出版している日英辞書。	見出し語	✓	✓	✓	✓
各種厚労省標準マスター関連用語集	MEDISが公開しているマスター。		✓		✓	
TO:CLIDAS	循環器に関する臨床情報データベース。		✓			
TO:J-CaseMap	自治医科大・東大が開発している辞書。		✓			
TO:Lilak	医学知識連結DB。東京大学が開発したオントロジー。	疾病名または部位コード	✓			✓
TO:薬効分類	総務省が定めた日本標準商品分類番号にしたがって分類された、医薬品の効能に関する分類。	分類番号		✓		
TO:JAPICID	医薬商品に関するコードへの紐付け。			✓		
TO:KEGGエントリ	京都大学が管理している医薬品DBへのID。	エントリID		✓		
TO:PubChem	アメリカNCBIが管理する化学物質データベースのID。	SID		✓		
TO:ATC	WHOによる薬品の分類コード。	ATCコード		✓		
TO:JLAC10	日本臨床検査医学会が制定した臨床検査項目の分類コード(第10回改訂版)				✓	
TO:JLAC11	日本臨床検査医学会が制定した臨床検査項目の分類(第11回改訂版)					
TO:ICD-10	WHOが制定した国際的な病名分類(第10回改訂版)。	ICDコード	✓			

TO:ICD-11	WHOが制定した国際的な病名分類第11回改訂版).	ICDコード	✓			✓
TO:HPO	Human Phenotype Ontology (HPO) 研究用のフェノタイプ用語集.	ID	✓			
TO:SNOME D-CT-IPS	SNOMED-CTの縮小版. SNOMED-CT/International Patient Summary Terminology	ID	✓			
TO:MedDRA/J	医薬品規制調和国際会議(ICH)が制定した医薬用語集の日本語版.	見出し語	✓			
TO:UMLS	アメリカNLMが公表している, 医療概念のメタソーラス. CUIを収載.	ID	✓	✓	✓	✓
TO:一般名処方マスタ	厚生労働省が策定している医薬品の一般名処方の標準的な記載.	一般名コード		✓		

## リンク先リソースのライセンス

リンク先のリソースによって、ライセンスの取り扱いが異なる。本データでは、最小セットとしてライセンスを持たない利用者でも利用可能なリソースへのリンクを収録する。これらのリソースへのリンク情報は、どの利用者でも利用できる。加えて、個別リソースのライセンスを持つ利用者には、そのリソースへのリンク情報を加えたデータを提供する。

最小セットとして収録するリソースは以下の通りである。

- 最小セットとして収録するリソース
  - TO:SNOMED-CT-IPS
  - TO:ICD-10
  - TO:HPO
  - TO:薬効分類
  - TO:一般名処方マスター
  - TO:JLAC10コード

下記のリソースは、利用者がライセンスを保有している場合、リンク情報を付け加えて提供する。

- ライセンスが必要なリソース
  - TO:Lilak
  - TO:J-CaseMap
  - TO:UMLS
  - TO:ICD-11
  - TO:MedDRA/J
  - TO:JAPICID
  - TO:検査分類1(検査マスター)
  - TO:検査分類2(検査マスター)
  - TO:KEGGエントリ

下記のリソースのリンク情報は、原則公開しない。

- 原則非公開のリソース
  - TO:日本医学会医学用語辞典

## 信頼度フラグ

属性情報とリンク情報には信頼度フラグが付与されている。信頼度は、人手によるチェックがなされた度合いにより、以下の五段階に分類される。開発・メンテナンスを通じて、信頼度DをAにあげていくことが求められる。

信頼度は、属性情報とリンク情報の各項目ごとに設定する。つまり、読みがなは信頼度Dであるが、ICD-10コードは信頼度がSといった情報を保持できる。

信頼度	概要
S	信頼度最大。既存のリソースと照合済み。
A	信頼度高。人間3名以上がチェック済。
B	信頼度中。人間2名がチェック済。
C	信頼度低。人間1名がチェック済。
D	信頼度低。自動推定された結果。

## 構築方法

### 初期構築

以下の2つの方法でエントリを構築する。

対象となる医療テキストに言語処理基盤を適用し、用語を抽出する。この際、頻度を集計し、頻度1は除く。医療テキストとしては、以下を用いた。

1. J-Stage(日本語症例報告)
2. 万病辞書構築時の用語頻度

1は、J-stageに公開されている日本語論文から症例報告に該当するテキストを抽出・収集したデータである。

同時に、他のリソースからエントリのマージも行なった。著作権などの問題で、これらが使えなくなった場合を想定し、由来を保持している。このため、使えないリソース部分を削除したリリースが可能である。

### サイズ(用語数)

現在の各データのサイズは以下の通りである。

サブデータ	エントリ数
病名データ	534778
医薬品データ	113256

部位データ	40133
検査データ	45624

## 未対応部分

現状未対応な部分は下表の通りである。

カテゴリ	名称
関係情報	REL:IS-A
	REL:PART-OF
リンク情報	TO: J-CaseMap
	TO:AIH

## データ管理の方針

膨大なエントリーのため、全てのデータを高い品質に保つのは高コストである。

そこで、頻度の高い部分の品質は高く、それ以外は頻度に応じて、品質改善の優先度を下げる。このため、全体のデータを高頻度ブロック、中頻度ブロック、低頻度ブロックの3つのブロックに分ける。それぞれのブロックの意義は次のとおりである。

- 高頻度ブロック: 2000語。人手によるチェックを行う。品質のチェック、方法の検討に用いる。
- 中頻度ブロック: 高頻度ブロックを除く上位20000語。人手によるチェックを行う。
- 低頻度ブロック: それ以外。人手によるチェックを行う。

このように3つのブロックにわけた後、以下の手順をふむ。

### ・ステップ0: 高頻度ラベル作成

高頻度ブロックのラベルを人手で作成する。

### ・ステップ1: 中頻度にラベル伝播

高頻度ブロックを教師データとし、中頻度ブロックを自動推定。人手でチェック。

### ・ステップ2: 低頻度ラベル伝播

高頻度ブロックと中頻度ブロックを教師データとし、低頻度ブロックを自動推定

### ・ステップ3: ランダムマスキング

頻度に応じて、1%をサンプリングし、エントリをマスクする。残り99%で1%を推定し、現状のエントリと異なったものが推定された場合は、チェックする。以降、ステップ三を繰り返す。

この手順イメージを以下に示す(図1)

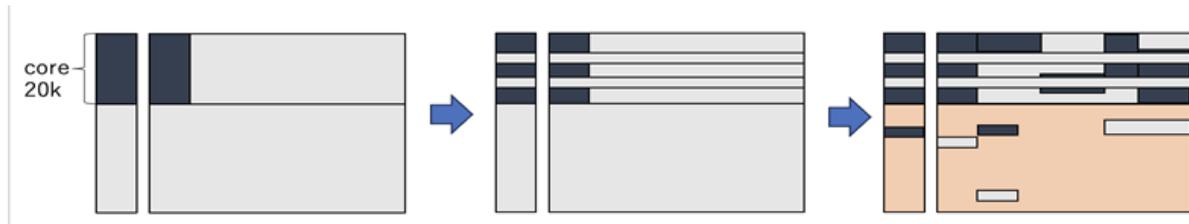


図1: ラベル伝播のイメージ

## 項目推定

自動推定の方法には大きく分けて次の3つの方法がある。

- 教師あり手法
  - 値が埋まっており、かつ、信頼度フラグの高い内容を教師ラベルとして、信頼度フラグの低い項目のラベルを推定する。推定に当たっては、生成モデルである日本語T5を用いた。
- 教師なし手法
  - 既存のオントロジーやターミノロジーと文字列が類似しているものを探し、それを項目とする。例えば、日本医学会医学用語辞典は、出現形と類似した日本医学会医学用語辞典のエントリを探し、その項目をそのまま用いた。
- 既存ツールの適用
  - 既存の紐づけツールが存在するリソースについては、ツールを利用して紐づけを実施した。

これら3つの手法を項目により使い分けて、項目の推定を行った。例えば、よみがなは、教師あり手法を用いて、学習モデルが読み仮名を生成する。よみがなに教師なし手法を使えないのは、「糖尿病2型」のよみ「とうによびょうにがた」があったとして近い出現形の文字列「糖尿病1型」のよみに対して、「とうによびょうにがた」を使えないことから明らかである。

一方、分類やクラス分けといった項目は教師なし手法がうまく機能する。例えば、HPOや日本医学会医学用語辞典は、すでに対象となる用語集があるため、ここから最も近いものを探すという教師なし手法を用いた。

最後に、すでに既存のツールが存在し、それを用いることができる場合がある。例えば、UMLSへの紐付けはQuickUMLSなど、ID(UMLSではCUIと呼ばれる)マッピングツールが存在し、これをそのまま用いた。

## 管理方法

管理は次の2パターンを想定している。

- メンテナンス: データ構築またはメンテナンス作業。業務としてデータのクオリティ向上に勤める専門家がもっぱら従事する。

- コメント: データチェック作業. 不特定多数のユーザが不定期にチェックし, データのクオリティについてコメントを行うことを想定している.

それぞれの役割に応じてアプリケーションを用意している. メンテナンスには「knitboat」という図2に示すウェブアプリを用いて作業を行う.

図2: 概念DBマネージャの使用イメージ. 本事業においては各種ユーザがデータを開覧・編集する. データには折に触れて新規に用語が追加されるが, 用語の詳細情報には間違いが含まれるため, それを人手で修正する. ユーザはデータの追加や進捗管理を実施する「管理者」と用語詳細を編集する「作業者」の2種類とする. 作業者は用語の詳細を編集する「作業」と, 作業済みの用語を確認する「確認」が可能で, 管理者が作業者ごとに「作業」担当の用語と「確認」担当の用語を任意に割り当てられる.

コメントは, より軽微なスマートフォンでも動作するビューアー(図3)を用いた作業を想定している.



図3:コメントのためのビューアー。iPhoneやAndroid上で動作し、用語の検索およびコメントができる。不正なエントリの情報のピックアップによる品質改善に用いる。

## メンテナンス手順

本節では、人手によるメンテナンス作業の手順について述べる。本データにおけるメンテナンス作業とは、自動推定を利用して初期構築されたエントリに対して、人手でメンテナンスを行う作業のことを意味する。

まず、各サブデータ間で共通のメンテナンス手順について説明する。次に、各サブデータごとの詳細なメンテナンス手順を述べる。

### 共通手順

本節では、各サブデータ間で共通のメンテナンス手順を述べる。

## 信頼度フラグ

人手によるメンテナンスを実施した項目は、信頼度を更新する。メンテナンス作業実施後に、作業実施前に入力されていた信頼度を一段階上げた値を入力する。また、作業実施前に入力されていた信頼度が「D」以下の場合、作業実施後は「C」を入力する。例えば、作業実施前に信頼度が「D」または空であった場合、作業実施後は「C」とする。ただし、人手によるメンテナンスによって付与できる信頼度は「A」が最高となるため、作業実施前に信頼度が「A」であった場合、作業実施後も「A」のままとする。

## メンテナンスで利用するタグ

メンテナンス作業では次のタグを利用する。

- 出現形\_flag:[TYP]
- 正規形:-1
- 正規形:[ERR]

各タグの詳細な運用はサブデータごとに異なるが、ここでは全般的な運用ルールを述べる。出現形\_flag:[TYP]は、タイポすなわち入力時のミスと判定できる出現形に付与する。正規形:-1は、出現形が、エントリとして妥当ではあるが適切な正規形が見当たらない場合に付与する。正規形:[ERR]は、出現形がエントリ構築時の抽出エラーであることが明らかで、エントリとして妥当ではないと判断した場合に付与する。

## 病名データ (DISEASE)

### 実施内容

自動推定によって構築されたデータに対して、メンテナンス対象の各列を人手でメンテナンスする。メンテナンス対象は、各属性情報のうち信頼度フラグが「S」以外のものとする。

- 基本情報
  - 出現形
  - 出現形\_flag
  - 出現形よみ
  - 出現形よみ\_flag
  - 正規形
  - 正規形\_flag
- 属性情報
  - TYPE:否定
- リンク情報
  - TO:ICD-10
  - TO:ICD-10\_flag
  - TO:Lilak
  - TO:Lilak\_flag
  - TO:日本医学会医学用語辞典

### 実施手順

#### 基本情報

基本情報のメンテナンス手順について、各ラベルごとに述べる。

#### 出現形

出現形は、メンテナンス時には変更しない。医療テキストから抽出されたものをそのまま利用する。

## 出現形\_flag

出現形の表記を確認し、入力時の誤字・脱字・変換ミスと思われる表記であれば、出現形\_flagの値を [TYP] とする。これらに該当しない場合、空欄のままとする。例えば、以下の例では正しい表記は「亢進」であるが、出現形では「高進」となっているため、出現形\_flagを [TYP] とする。

例:

出現形:「甲状腺機能高進」  
出現形\_flag:[TYP]

## 出現形よみ

自動推定によって得られた出現形のよみがなが正しいかを確認し、誤りがあれば修正する。また、かっこ書きで英語略称などが付されている場合、かっこのよみがなは付与しない。例えば以下の例では、出現形において、「慢性腎臓病」の後に略称が書かれているため、出現形よみはかっこを除いた「まんせいじんぞうびょうしーけーでいー」とする。

例:

出現形:「慢性腎臓病(CKD)」  
出現形よみ:「まんせいじんぞうびょうしーけーでいー」

## 出現形よみ\_flag

[共通手順](#)に従い信頼度を更新する。

## 正規形

出現形に対する正規形を入力する。病名データの正規形は、標準病名マスターに収録されている病名を参考に行っている。正規形は以下の基準で入力する。

1. 標準病名マスターを検索して、出現形に該当する標準病名を入力する。
2. 標準病名マスターを検索しても出現形に該当する標準病名が見つからない、もしくは1つの標準病名に絞り込めない場合、「-1」を入力する。
3. 出現形が抽出エラーとみなせる場合、[ERR] を入力する。

それぞれの基準について詳しく述べる。

まず1について。出現形に最も近いと思われる標準病名を[標準病名マスター病名検索](#)で検索し、該当する標準病名があれば正規形として入力する。出現形が略語で表記されている場合、インターネット上で検索し、その略語が表す病名を特定する。その際、略語が表す病名が1つに絞り込める場合、それをもとに標準病名マスター病名検索で標準病名を検索する。略語が表す病名が1つに絞り込めない場合、正規形に「-1」を入力する。

次に2について。標準病名マスター病名検索で検索しても出現形に該当する標準病名が特定できない場合、正規形に「-1」を入力する。標準病名が特定できないケースとしては、以下のような事例がある。

例:

- 出現形が表す症状が複数想定される:「皮膚症状」「臓器障害」など
- 出現形から病名が特定できない:「疾患」「慢性疾患」「後遺症」など
- 出現形から症状がある部位が特定できない:「逆流」「潰瘍」「穿孔」「腫大」「結石」など
- 出現形が病名でない語:所見、変化を示す用語(上昇、減少など)、検査所見(～像、～陰影など)

次に3について。エン트리構築時に出現形として抽出された語が、抽出範囲の誤りによって正しく語を抽出できていない場合や、病名・症状以外を表す語である場合、正規形に [ERR] を入力する。

例:

- 医療と無関係とみなせる語:「A」「故障」「災害」「透」「声」など

- 途中で語が途切れているもの:「悪性腫」など
- 病名と治療名を組み合わせた語:「骨折治療」「腎性貧血治療」など
- 照応表現:「標題癌」など(症例報告などで、報告の標題に挙げた病名を参照する表現。「標題癌」の場合、報告の標題に具体的な癌の名称が書かれており、文書中でそれを参照するときに「標題癌」と表記することがある)

## 正規形\_flag

[共通手順](#)に従い信頼度を更新する。

## 属性情報

TYPE:否定

出現形が否定表現を含むかどうかをブーリアンで入力する。否定表現である場合、「1」否定表現でない場合「0」とする。否定の接頭辞「非」が付くことにより後続の語が否定されている出現形を否定表現とみなした。

例:

出現形:「非糖尿病」「非癌」「非肝硬変」

## リンク情報

TO:ICD-10

ICDコードを入力する。自動推定によってすでにICDコードが入力されている場合、人手で正規形を利用して標準病名マスターを検索し、入力済のICDコードが正しいかを判定する。入力済のICDコードが誤っていた場合、正しいICDコードに修正する。自動推定によってICDコードが入力されていない場合、人手で正規形を利用して標準病名マスターを検索し、該当するICDコードを特定し、入力する。

TO:ICD-10\_flag

[共通手順](#)に従い信頼度を更新する。

TO:Lilak

自動推定されたLilakとの紐づけ結果を確認・修正する。メンテナンス作業実施前にTO:Lilak\_flagの値が「D」であり、かつ正規形が [ERR] 以外の行を対象とする。TO:Lilak列には、自動推定によって出現形と最も類似すると判定されたラベルが入力されている。作業者は、出現形からラベルの妥当性を判断し、自動推定によって入力されたラベルが誤っていれば修正する。複数のラベルが該当する場合、最大3つまで半角セミコロン (;) でラベルを接続し列挙した。

例

出現形:「腎機能の悪化」

TO:Lilak:「S0055574:腎機能異常;S0004256:腎機能障害」

TO:Lilak\_flag

[共通手順](#)に従い信頼度を更新する。ただし、作業者が判断に迷った場合、一時的にflagを「C?」として別の作業者と協議ののち確定させた。

TO:日本医学会医学用語辞典

自動推定された日本医学会医学用語辞典との紐づけ結果を確認・修正する。メンテナンス作業実施前にTO:日本医学会医学用語辞典\_flagの値が「S」以外であり、かつ正規形が [ERR] 以外の行を対象とする。TO:日本医学会医学用語辞典列には、自動推定によって出現形と最も類似すると判定されたラベルが入力されている。

作業者は、出現形からラベルの妥当性を判断し、自動推定によって入力されたラベルが誤っていれば修正する。複数のラベルが該当する場合、最大3つまで半角セミコロン (;) でラベルを接続し列挙した。

TO:日本医学会医学用語辞典\_flag

[共通手順](#)に従い信頼度を更新する。ただし、作業者が判断に迷った場合、一時的にflagを「C?」として別の作業者と協議のち確定させた。

## 医薬品データ (MEDICATION)

### 実施内容

出現形の正規形、分類の付与を行う。正規形は、原則として医薬品の一般名を用いる。(例外ルールは下記実施手順を参照)

分類は、医療用医薬品の薬効分類(<https://www.kegg.jp/brite/jp08301>)に沿って、出現形の分類を行う。

### 実施手順

1. 出現形を確認し、この出現形が医薬品に関連する出現形かどうか判定する。出現形が医薬品に基づかないものであれば、正規形は-1とする(コードについては正規形が-1のときに入力するコードを参照)。抽出ミスであれば、正規形に[ERR]を入力する。出現形にタイポがあったら、出現形\_flagの部分に[TYP]といれるようにする。
  - a. 抽出ミスと思われるものは[ERR]。(例:「アスピリン」が「アスピリ」とかになっている場合など)
  - b. その略形が実際に使われるのであれば、[ERR]にはしない。ただ、酸化マグネシウムが、「カマ」や「カマグ」になっている場合などは-1にしない
  - c. 「と」、「、」で複数の医薬品が繋がれている場合には基本的には抽出ミスとして「ERR」となるが、例外として複数の医薬品の組み合わせがなんらかのレジメンを意味する場合には「-1\_レジメン」とする。

#### 正規形が-1のときに入力するコード

医薬品でないもの:  
(-1\_非医薬品)アレルギーなどの原因となりそうなものも含む  
素材:天然ゴム, 合成樹脂, リネン  
衣類:手袋, マットレス  
動物:ミニチュア・ダックスフンド  
飲食物  
農薬系  
元素:水素, 一酸化炭素, ヒ素  
装具系(-1\_装具):マウスピース, 補助人工心臓, 下肢装具  
治療(-1\_治療):アロマセラピー, レーザー, 鍼灸, イオン, PET検査, など  
レジメン(-1\_レジメン):レジメンの場合には-1\_治療ではなく-1\_レジメンとする

2. 正規形には基本的には一般名を入力するが、1つに決まらないケースや一般名を入れられないケースがある。

#### 一つに決まらないケース

同格: 正規形が複数あるが、同じものを言っているもの(アスピリン→アスピリン;アスパロン;アセチルサリチル酸)  
配合剤: 正規形が複数あり、出現形がいくつかの成分を含んでいるもの(配合剤など, アドエア→サルメテロールキシナホ酸塩;フルチカゾンプロピオン酸エステル)  
曖昧: 正規形が複数あるが、出現形が曖昧であることによるもの(オザグレル→オザグレル塩酸塩水和物; オザグレルナトリウム)

#### 一般名を入れられないケース

KEGG DGROUPを利用(例, 出現形が降圧薬であれば, 血圧降下薬)

そのルールに関して以下に示す。

## A. 複数ある場合の原則ルール

候補が3個以下:  
 同格:一つを選択...ルール:一般名として適切であるものを選ぶ  
 配合剤:+で繋ぐ  
 曖昧な場合:;で繋ぐ

候補が4個以上:  
 同格:一つを選択  
 配合剤(combination drugs):[COM]  
 曖昧な(ambiguous)場合:[AMB]  
 \*曖昧とは正規化先が複数あってまとめられない場合

### \*同格で一つに選択する際の留意点

この際、それまでに選んでいた代表的な正規形がある場合にはそれを選ぶようにする。(例、出現形:アスピリン、一般名:アスピリンとすでに選んでいた場合、出現形:低用量アスピリン、一般名:アスピリン、とする。ちなみに、一般名にはアスピリン、アスパロン、アセチルサリチル酸がある、<https://www.kegg.jp/entry/D00109+ja>)

## B. 正規形を一般名に紐付けできない場合の原則ルール

(一般名に紐付けできるもの以外の)

正規形についてはkegg dgroupの名称を使う。

当てはまるものがなかった場合、keggの薬効分類番号に用いている名称を使う。そこにも当てはまるものがない場合、添付文書上の薬効分類名を用いる。

\*生薬の一般名にはカタカナ、漢方の正規形には、漢字を用いる。

### 3. 入力フラグについて、

- a. C(2名チェック)だった正規形を直した場合には、C!(2名で意見の相違)にする。Cがあっていた場合には、CからB(2名チェック)にする。
- b. D(1名チェック)だった正規形を直した場合には、C(2名チェック)にする。

### 4. 出現形が医薬品に基づいていれば(正規形が-1で始まるものでないならば)、この出現形に対応する適切な薬効分類を<https://www.kegg.jp/brite/jp08301>のツリーから探す。この薬効分類は次の例に沿って適切な大きさのものを選択する。以下に選択の原則を示す。

- a. 1つの薬効分類で表せるものを選択する。
- b. 分類がいくつかの分類にあてはまる場合は以下のルールに則る。
  - i. 分類がいくつかの分類に渡ってしまう場合で、3つまでならセミコロンで繋ぐ。
  - ii. 4つ以上になったら、上位概念にする。(上位概念は上位2桁目まで、6.1\_抗生物質などの概念が入力可能な最上位概念)
  - iii. 内服と外用が別れてしまう場合などは、内服優先でTREEを選択する。(薬効分類にはどちらもセミコロンで区切って入力する。)
  - iv. 上位概念があまりにもばらけてしまったら、-1(しかし、できるだけ-1は付与しない)。一階層目の分類ではカバーできない場合など、ツリーの中に適切なものが見つからない場合には「9\_その他」とする。
- c. 眼科用医薬品や歯科用医薬品などの分類は、出現形からそれが読み取れる場合以外にはTREEには利用しない。(薬効分類には、内服、外用、眼科、の順で最大3つまで入力する。この順は原則として、全身→局所とする。外用の中でも軟膏類や吸入があるが、この場合は軟膏を優先する。)

以下、例:

出現形	分類	理由
ステロイド	2.4.5_副腎ホルモン剤	「2.4.5_副腎ホルモン剤」にはエピネフリンなどのステロイド以外も含まれてしまうが、一つ下の分類(例えば、2452 コルチゾン系製剤、2454 フッ素付加副腎皮質ホルモン製剤、2456 プレドニゾン系製剤などもステロイドとみなせる)では細かすぎる。
造影剤	7.2_診断用薬(体外診断用医薬品を除く)	「721 X線造影剤」に近いが、網羅しているわけではない。(例えば、mri用造影剤のガドレル酸メグルミンは7290である。)そのため、一つ上の階層「7.2_診断用薬(体外診断用医薬品を除く)」とする。
抗精神病薬	1.1.7_精神神経用剤	「44_中枢神経系用薬」には全身麻酔剤、解熱鎮痛消炎剤などが含まれてしまうが、その下の階層だと狭すぎるため、精神神経用剤という項目がより合致すると考えられるので変更

免疫抑制剤	9_その他	「3999 他に分類されないその他の代謝性医薬品」にできるかもしれないが、一部の抗がん剤など免疫抑制剤にくらわられる医薬品があるため。
鎮痛薬	1.1.4_解熱鎮痛消炎剤	外用薬の「264 鎮痛、鎮痒、収斂、消炎剤」や歯科用の「273 歯科用鎮痛鎮静剤」は含めず、より一般的な「1.1.4 解熱鎮痛消炎剤」を設定する。

#### 参考資料

KEGG DRUG Database

[https://www.kegg.jp/kegg/drug/drug\\_ja.html](https://www.kegg.jp/kegg/drug/drug_ja.html)

承認されたバイオ医薬品（一般名など）

[https://www.nihs.go.jp/dbcb/approved\\_biologicals.html](https://www.nihs.go.jp/dbcb/approved_biologicals.html)

日本薬局方名称データベース

<https://jpdb.nihs.go.jp/jp/>

## 部位データ (BODY)

### 実施内容

自動推定によって構築されたデータに対して、メンテナンス対象の各列を人手でメンテナンスする。メンテナンス対象は、各属性情報のうち信頼度フラグが「S」以外のものとする。

- 基本情報
  - 出現形
  - 出現形\_flag
  - 出現形よみ
  - 出現形よみ\_flag
  - 索引構造
  - 正規形
  - 正規形\_flag

### 実施手順

#### 基本情報

#### 出現形

[病名データ](#)と同様

#### 出現形\_flag

[共通手順](#)に従い信頼度を更新する。ただし、出現形内でひらがなが使用されている場合、以下の基準に応じて出現形\_flagを [TYP] とする。

- 以下に該当する場合、[TYP] とはしない
  - 当てはまる漢字がなく、ひらがなとするしかない場合
  - ひらがなで利用されることが一般的な場合
  - 例
    - 「うつ病」
    - 「骨粗しょう症」
- 以下に該当する場合、[TYP] とする
  - 意図的にひらがなを利用しているとみなせない場合
  - 例
    - 「こう内炎」

## 出現形よみ

[出現形よみ](#)と同様

## 出現形よみ\_flag

[共通手順](#)に従い信頼度を更新する。

## 正規形

出現形に対する正規形を入力する。部位データでは、以下の基準で正規形を決定する。

1. SOURCE列に「BODY」を含む場合、出現形をそのまま正規形とする
2. 出現形の信頼度が「S」以外の場合、以下の手順で正規形を決定する
  - a. 基準1で正規形として採用された正規形のリストに該当する語があれば、それを正規形とする
  - b. 基準2aで該当する語が無い場合、以下いずれかとする
    - i. 新たに正規形を作成する
    - ii. -1 とする
    - iii. [ERR] とする

それぞれの基準について詳しく述べる。

まず1について。SOURCE列に「BODY」を含む行は、すでに既存のリソースと照合済みの他のリソースのエントリをマージして得られたエントリである。したがって、これらのエントリの出現形は正規形と一致しているため、信頼度は「S」である。

次に2について。基準1で得られた正規形をまとめると、部位を表す語の正規形のリストが得られる。出現形の信頼度が「S」ではない場合、基準1で得た正規形のリストから該当する語を探し、正規形とする。基準1で得た正規形のリストに該当する語がない場合、新たに正規形を作成するか、-1を付与するか、[ERR]を付与する。どの対処が適しているかは、事例ごとに判断する。以下にメンテナンス作業時に議論した事例を挙げる。

- 正規形を作成
  - 「～組織」:「組織」より前の語から正規形を作成する
    - 例:「胎盤組織」の場合、「胎盤」を正規形とする
  - 正規形内の数字は漢数字を用いる
    - 例:出現形「第1腰椎」の場合、正規形「第一腰椎」
- [ERR]
  - 出現形が表す部位が、2つ以上の部位に当てはまる
  - 「～細胞」:細胞レベル以下は部位に該当しない
    - 例:「星状細胞」
  - 「～の」:「の」より前の語は部位とみなさない
    - 例:「血管の」
  - 文章の見出しと思われる表現
    - 例:「28関節」
  - 明らかに部位ではない表現
    - 例:「2階」
  - 成分名
    - 例:「タンパク質」
- -1
  - 全身に分布しているもの
    - 例:「皮膚」「血管」など
  - 「正中部」「血流」「全層」「洞内」「軟部」
  - 「～周囲」
    - 例:「肛門周囲」「ストーマ周囲」
  - 肝臓か肺のいずれか判別できないが部位とみなせる表現
    - 例:「S2」

- 病変の部位
  - 例:「結腸癒痕」「癌病変部」
- 病名になりうる表現
  - 例:「血栓」
- 人工物
  - 例:「人工弁」「義歯」「インプラント」「人工肛門」
- 正確な部位は特定できないが、何らかの部位の省略とみなせる表現
  - 例:「左上6番」「5番」
- 体液
  - 例:「尿」「粘液」
- 医療器具の名称
  - 例:「カテーテル」
- 医療技術の名称
  - 例:「結紮」

また、作業者が出現形に対する正規形の判断に迷った場合、一時的に正規形flagを「C?」として別の作業者と協議のちに確定させた。

### 正規形\_ICD-11 (WIP)

部位データの正規形をICD-11のエクステンションコードをベースにしたものに変更した。エクステンションコードとは、ICD-11から新たに追加された章であり、重症度や時間関係などを付加的に付与することができる。エクステンションコードのうち、「XA」ではじまる解剖学的詳細 (Specific Anatomic Detail) と呼ばれるカテゴリは人体の部位を表すため、部位データの正規形として採用した。

エクステンションコードは内部に階層構造をもつ。すべての階層を正規形として採用すると正規形の粒度として細かすぎるため、エクステンションコードを持つ要素のうち、最上位の階層とその1つ下の階層にある要素だけを正規形に採用した。

また、ICD-11 はICD-10のような日本語訳がまだ存在しない。そのため、日本語の正規形として、英語を機械翻訳により日本語に翻訳し、明らかな翻訳の誤りを修正した語彙を利用している。

### 正規形\_flag

[共通手順](#)に従い信頼度を更新する。

### TREE

索引構造を入力する。部位データの索引構造は、『ネッター解剖学アトラス第7版』による階層構造に依拠する。正規形をもとに、最も適した階層を付与する。正規形が「-1」の場合、出現形をもとに階層を付与する。出現形から該当する階層が不明な場合、「8\_その他」とする。正規形が [ERR] の場合、階層は付与しない。

### TREE\_flag

[共通手順](#)に従い信頼度を更新する。

## 検査データ (TEST)

### 実施内容

自動推定によって構築されたデータに対して、メンテナンス対象の各列を人手でメンテナンスする。

- 基本情報
  - 出現形
  - 出現形\_flag
  - 出現形よみ
  - 出現形よみ\_flag

- 正規形
- 正規形\_flag
- TREE
- TREE\_flag

## 実施手順

### 基本情報

### 出現形

[病名データ](#)と同様

### 出現形\_flag

[共通手順](#)に従い信頼度を更新する。ただし、検査データにおいては大文字・小文字の誤りも [TYP] とした。

- 大文字・小文字を誤っているため [TYP] とした出現形の例
  - 「Spo2」(正しくは「SpO2」)
  - 「Ph」(正しくは「pH」)

### 正規形

出現形を参照し、最も適切と思われる正規形を付与する。まず、検査項目の正規形の候補となる表現をリスト化し正規形参照資料とした。検査データにおいては、厚生労働省による[医師国家試験出題基準](#)の「主な検査項目の表記」に挙げられている項目を正規形とした。次に、正規形参照資料を元に、以下の基準でデータのメンテナンス作業を実施した。

正規形の判定は、以下の基準で行う。

1. 正規形参照資料に該当する語があれば、その語を正規形とする。
2. 1に該当する語が無い場合、以下いずれかの対処をする。
  - a. 正規形として新たに項目を立てる。[臨床検査項目基本コードセット](#)に含まれる検査項目を参考に、出現形が一般的な検査および検査項目とみなせる場合、正規形とする。
    - i. 出現形が「検体＋検査項目」で構成される語の場合、頻出する語は正規形とする。例えば「動脈血pH」はそのまま正規形とする。
    - ii. 出現形が「ST」「QT」など検査特有の表現は検査項目とみなし正規形「心電図検査」とする。
  - b. -1を付与する。
    - i. 1つの出現形に対して、4つ以上の正規形が対応すると判断できる場合、-1を付与する。
    - ii. 出現形が検査結果を含む表現と判断できる場合、-1を付与する。例えば「APTT延長」はAPTTが延長しているという検査結果を表す。
    - iii. 出現形が検体の種類を表す表現の場合、-1を付与する。例えば「血液」、「喀痰」、「羊水」などは検体の種類である。
  - c. [ERR] を付与する。
    - i. 出現形が複数の語の場合(例:「血圧と心拍数」)
    - ii. 抽出範囲を誤っている場合(例:「リンパ球の」)
    - iii. 検査結果値や陽性／陰性を含む場合(例:「総ビリルビン2. 2」「インフルエンザ抗原検査陽性」)

### 正規形\_flag

[共通手順](#)に従い信頼度を更新する。ただし、検査データについては作業実施前に信頼度が「S」の場合であってもメンテナンス対象とした。作業実施前に信頼度が「S」であった場合、作業実施後の信頼度は「C」とした。また、作業者が判断に迷った場合、一時的にflagを「C?」として別の作業者と協議ののち確定させた。

## TREE

索引構造を入力する。検査データの索引構造は、「検査分類1」「検査分類2」による階層構造に依拠する。正規形をもとに、最も適した階層を付与する。正規形が「-1」の場合、出現形をもとに階層を付与する。正規形が[ERR]の場合、階層は付与しない。出現形が検査名に該当する表現ではあるが、適切な階層が無い場合、「10\_不明」とする。

## TREE\_flag

[共通手順](#)に従い信頼度を更新する。ただし、作業者が判断に迷った場合、一時的にflagを「C?」として別の作業者と協議ののち確定させた。

## おわりに

本稿では、SIP「統合型ヘルスケアシステムの構築」のために開発された医学概念・知識連結データベースの構築、管理方法について述べた。大規模かつ、変化し続ける複数のリソースと対応をとるために、言語処理技術を用いた半自動化、および、品質維持のための工夫についても紹介を行った。

## Appendix

### リンク先リソースごとのライセンス情報

[リンク先リソースのライセンス](#)で述べた通り、本データ内で紐付けているリンク先のリソースは、それぞれライセンスが異なる。リソースによって、会員登録が必要なもの、有償での購入が必要なもの、二次利用の制限があるものなどが存在する。下表に2024/06時点での各リソースのライセンス情報へのリンクを掲載する。本データのリンク情報を利用した二次利用の際には、各リソースのライセンスを確認のこと。

	参考URL	入会の必要有無	金額	最小セットへの収録
TO:Lilak	<a href="https://www.str-v.com/en/ontology.html">https://www.str-v.com/en/ontology.html</a>	無し	販売元に確認	
TO:J-CaseMap	<a href="https://www.naika.or.jp/j-casemap/">https://www.naika.or.jp/j-casemap/</a>	有り	9000円	
TO:日本医学会医学用語辞典	<a href="https://jams.med.or.jp/dic/mdic.html">https://jams.med.or.jp/dic/mdic.html</a>	有り	0円	
TO:UMLS	<a href="https://www.nlm.nih.gov/research/umls/index.html">https://www.nlm.nih.gov/research/umls/index.html</a>	有り	0円	
TO:SNOMED-CT-IPS	<a href="https://www.snomed.org/international-patient-summary-terminology">https://www.snomed.org/international-patient-summary-terminology</a>	有り	0円	✓
TO:ICD-10	<a href="https://www.who.int/standards/classifications/classification-of-disease">https://www.who.int/standards/classifications/classification-of-disease</a>	無し	0円	✓

	<a href="#">s</a>			
TO:ICD-11	<a href="https://icd.who.int/en">https://icd.who.int/en</a>	無し	0円	
TO:HPO	<a href="https://hpo.jax.org/">https://hpo.jax.org/</a>	無し	0円	✓
TO:MedDRA/J	<a href="https://www.jmo.mri.jp/">https://www.jmo.mri.jp/</a>	有り	会員区分ごとに異なる	
TO:薬効分類	<a href="https://www.soumu.go.jp/main_content/000294493.pdf">https://www.soumu.go.jp/main_content/000294493.pdf</a>	無し	0円	✓
TO:JAPICID	<a href="https://www.japic.or.jp/service/iyaku/index.html">https://www.japic.or.jp/service/iyaku/index.html</a>	販売代行業者を通じて購入 <a href="https://www.data-index.co.jp/">https://www.data-index.co.jp/</a>	販売代行業者に確認 <a href="https://www.data-index.co.jp/">https://www.data-index.co.jp/</a>	
TO:KEGGエントリー	<a href="https://www.kegg.jp/kegg/kegg_ja.html">https://www.kegg.jp/kegg/kegg_ja.html</a>	無し	0円	✓
TO:PubChem	<a href="https://pubchem.ncbi.nlm.nih.gov/">https://pubchem.ncbi.nlm.nih.gov/</a>	無し	0円	
TO:ATCコード	<a href="https://atcddd.fhi.no/atc_ddd_index/">https://atcddd.fhi.no/atc_ddd_index/</a>	無し	0円	
TO:一般名処方マスター	<a href="https://www.mhlw.go.jp/seisakunitsuite/bunya/kenkou_iryou/iryohoken/shohosen_240401.html">https://www.mhlw.go.jp/seisakunitsuite/bunya/kenkou_iryou/iryohoken/shohosen_240401.html</a>	無し	0円	✓
TO:検査分類1	<a href="https://www2.medis.or.jp/master/ke/nsa/index.html">https://www2.medis.or.jp/master/ke/nsa/index.html</a>	無し	0円	
TO:検査分類2	<a href="https://www2.medis.or.jp/master/ke/nsa/index.html">https://www2.medis.or.jp/master/ke/nsa/index.html</a>	無し	0円	
TO:JLAC10コード	<a href="https://www.jslm.org/committees/code/index.html">https://www.jslm.org/committees/code/index.html</a>	無し	0円	✓

## メンテナンス作業時のQA

人手によるメンテナンス作業時に作業員から生じた疑問点と、その解決法について掲載する。

## 病名データ

No	質問	回答
1	TREEの分類、どれを参考にしたらよいか	標準臨床検査マスター(ただし、内視鏡検査、画像検査などは含まれません)
2	出現形がBMI→参考資料にない→正規形はBMI; 体格指数; ボディマス指数と入力でよいか	正規形にない場合どうするか→正規形は適宜参考資料に追加可(一般的な検査であれば問題なし) BMI、体格指数、ボディマス指数 セミコロン(;)を使用するときは、すでに入力されているものと同様に半角入力をお願いします(正規形の英数字等は原則"全角"でそろえてください)。
3	正規形Sで出現形HbA1cであるが、グリコヘモグロビン分画;グリコヘモグロビンA1c(NGSP)正規形参考資料には:HbA1cがあり。CRPも同様。	HbA1c CPR
4	出現形が肝機能→TREEはメインのAST・ALT・γGTPの項目でいいか	複数のTREEの入力が可能か→TREEは1つしか表示できないため、作業者の判断で1つだけ選択(より自然に使う方を選択、深く考えすぎる必要ありませんので、同等であれば直感で選択していただいてOK)
5	27腎機能→TREEは複数該当? 生化と一般検査/尿一般検査	同上
6	出現形CD→Cdはカドミニウムの略であるが、大文字小文字なのでERRと考えていいのか	夫文字のCDはクロストリジウム・ディフィシル(現在はクロストリジオイデス・ディフィシル)の略ではないでしょうか? 複数該当する場合は1でもよい気がします。 ERRにしましょう
7	CD抗原やCDトキシンとなればクロストリジウム・ディフィシルとも考えたんですが、CDだと菌名になってしまうので、-1でなくERRかと思い、検査項目ではまるものと考えてCdであればカドミニウムで検査項目に当てはまるかと思いました。菌名やウイルスの表現の場合、ERRでなく-1としていいのか。	出現形:CD→ERR 菌名・ウイルス名の扱い→ERR
8	病名や症状はERRでいいか(NIDDM・骨折・痛み・精神症状など)	ERR
9	スコア(SOFA、DAS、JOASコアなど)指標は、[ERR]でいいか、-1で残すのか。	-1で残してください
10	HbA1など、途中で切れているものは[ERR]でいいか。	抽出範囲のERR
11	~のなど、部位や病名の時と同様[ERR]でいいか。	「~の」という場合は、ERR



No	質問	回答
1	ステロイドに関しては、仕様書では、2.4.5_副腎ホルモン剤をふるようになっていましたが、ID1, 13は-1となっていました。先日お伺いした際にも-1のまま話がすすんでいましたが、ID1, 13も仕様書通りに変更してよいでしょうか。	正規形が-1の場合でも、できるだけTREE に入力する
2	ID167 の糖尿病治療薬は 3.9.6_糖尿病用剤をふっても問題ないのかとも思いましたが、やはりここは-1が最適なのでしょうか。	正規形が-1の場合でも、できるだけTREE に入力する
3	タクロリムスなど、内服、外用とどちらも存在する場合、内服の方の薬効分類を優先してTREEに入力してよいでしょうか？-1にするべきでしょうか？また、その際、薬効分類の欄に内服、外用ともに記載することでよいでしょうか？	<p>・案1: 内服と外用では内服優先、外用は基本使わない(当初1つにしたいと思っていたため) しかし、むしろタクロリムスなど、外用でよく使われる薬も、さらに、複数の付与がOKという語があったため、</p> <p>・案2: 薬効分類に3つまでなら入れられるので、内服、外用どちらもいれる→TREEは-1に-</p> <p>案2: 基本的に内服、外用を優先し、歯科、眼科を例外的な扱いにするのはいかがでしょうか？(歯科、眼科は局所だが、外用、内服ともに全身に使うため)</p> <p>TREEの原則として、どれか一個しかふることができない。 TREEはできるだけ-1にしたくないので、薬効分類のうちどれか代表的なものをふりたい。 例: 2144のACE阻害薬であれば、これがメインと考えられるので、2144 →タクロリムスの場合、TREEがわれてしまう。これは、内服優先で、TREEをふる。(-1にはしない)</p>
4	出現形で、NSAIDsのsが抜けているような場合も[ERR]をふってよいでしょうか？ ID96	<p>複数形か単数形か、として考えられると思うので、単数形としてみてもいいかと思います。</p> <p>→[ERR]ではない。</p> <p>原則:[ERR]ですが、抽出ミスのみで、人間の間違いは[ERR]ではない</p> <p>例: NSAIDDSはNSAIDsとみなしてよい</p>
5	ID 111 ベクロニウム臭化物 は現在製造販売されていないようです。その場合は正規形等への記入は-1でしょうか。	<p>辞書の使用例として、過去の文書の解析があるため、現在製造されていなくとも辞書には含めたいと思っています。</p> <p>→現在製造されてなくとも-1にはしない。</p>
6	TREEには内服優先で1つのみ入力して、(外用、眼科用)はTO:分類の欄に内服分類とともに入力してよいか？	TREE⇒内服優先で1つのみ TO:分類⇒内服、外用、眼科用記入可
7	PET、CT等の検査名の場合は、-1_治療 を入力してよいか	検査名も-1_治療 で入力する
8	内服の薬効分類を上位にもっていった2桁にしても2つ以上になり、さらに外用、眼科用がある場合は、内服の数を削ってでも外用、眼科を入れてもよいでしょうか？	薬効分類: 内服、外用、眼科用の薬効分類が存在する場合は、それぞれ1つ以上入力する。
9	ID13 ステロイド剤 で 外用が 264 軟膏クリーム類 と2290 吸入があり優先順位はどうすべきか	外用の中では、優先順位 軟膏類→吸入(全身⇒局所の順にする)
10	出現形が一般名の場合は、正規形は単剤を入力で大丈夫か？配合剤は、分類のところに入れれば入れる方向でよいか？	出現形が成分名の場合は、正規形は単剤で入力。配合剤は分類に入れれば入れる

11	候補となる成分が二つ存在する場合正規形に(;)でつなげて記入してもよいのか	候補となる成分が2つ以上ある場合には、3つまでなら;で繋いでいれる。
12	D440 ベタメタゾン の場合、ベタメタゾンを販売名(総称名)をとらえれば、一般名はベタメタゾンのみ。ベタメタゾンを一般名の一部ととらえるとベタメタゾンプロピオン酸エステル等ほかの一般名も候補となる。販売名をとらえるか、一般名をとらえて正規形(-1)とするべきかどうか。	ベタメタゾンを一般名の一部とすると、候補があまりにも多くなってしまうため、販売名とする(インスリン、とかも同じルールに従えると思います。インスリン→インスリングルルギンなどにすると、候補が大量になってしまう。)
13	正規形は、情報処理の段階でのみ使われるデータなのでしょうか。それとも、辞書を使用した時に人の目に入る、表示されるデータになりますか？	例えば、シンプルに利用するケースとしては、"タキソール注射液30mg", "タキソール注射液100mg", "パクリタキセル注30mg/5mL「NK」"のようなテキスト表現(表現形とか出現形といいます)があったら、"パクリタキセル", "パクリタキセル", "パクリタキセル"のように表現形のゆらぎを正規形(または、標準形ともいいます)丸め込む作業に使われます。プログラムに辞書を読み込ませて、入力された表現のうち、辞書中の最も近い表現形を探索して、標準形を出力させます。そのため、出力結果として、正規形が出力されるため、処理結果として人が見るものになり得ります。
14	~正規形について薬、~剤、~製剤等の入力については、何かに記載されているものを使用する等の原則はなしで、出現形に添って適宜決定してよいのでしょうか？	1、日本薬効分類名、2、keggクラス分類名の優先順位
15	例えば、出現形が TNF $\alpha$ の場合にそのままであればサイトカインということで正規形は -1_治療 になるのではと思いますが、できるだけ医薬品に寄せるように考えて、TNF阻害薬 にするほうがよいのでしょうか。	「TNF $\alpha$ 」だと意味が変わってしまうため、-1_治療のままとする
16	Ca → カルシウム剤(薬効分類より) K → カリウム補給剤(keg分類名より) Mg → マグネシウム製剤(keg分類名より) とするべきか？ 元素として(-1_非医薬品)とするべきか。	ノイズになってしまう可能性があるため、元素として扱う
17	正規形の入力規則について	正規形: 一般名が一つに特定できる場合、一般名を正規形に入力 一つに特定できない場合、KEGG DGROUPを用いる。 ・原則括弧記載はなし。 ・1つの薬品について、一般名が複数存在する場合はこれまで通り3つまでなら(;)でつなげて入力。・
18	薬効分類の入力規則について。数を絞るのか、-1とするのか。また、その方法について。	薬効分類:数を絞らずにすべて入力する。これまで通り、2桁までの上位にまとめるのも可。

19	販売中止となった薬品について。現時点での薬効分類表には記載がないため販売時点での薬効分類番号を検索することになるが、一般名、TREE、薬効分類すべて入力が必要か	正規形、TREE、薬効分類、全て入力する
20	漢方薬の正規形の入力規則について。エキス、丸製剤、漢字、カタカナ表記ものの優先順位	漢字のみで入力
21	出現形(鉄、銅、亜鉛等)は、正規形(鉄剤、銅補給剤、亜鉛補給剤、)として問題ないか。	問題なし
22	●経口糖尿病薬→糖尿病薬or(-1)、経口ステロイド、吸入ステロイド→副腎皮質ステロイドor(-1)経口、吸入、外用等つく場合はgroup名としては存在しないため(-1)としてよいか	経口糖尿病薬→糖尿病薬 経口ステロイド、吸入ステロイド→副腎皮質ステロイド とする
23	正規形の入力規則について	(一般名に紐付けできるもの以外の)正規形については原則として kegg dgroupの名称を使うが、うまく当てはまるものがなかった場合、keggの薬効分類番号に用いている名称を使う。 そこにもううまく当てはまるものがない場合、添付文書上の薬効分類名を用いるものとする。
24	一般名が4以上ある場合の優先順位は入力者が決めてよいか。	入力者が決めてよい。判断に迷う場合は相談する。
25	正規形の分類名について、 三環系抗うつ薬:keggグループ分類にあり 抗うつ薬 :kegg分グループ分類になし の場合、“抗うつ薬“という名が使われているということが確認できれば使用してもよいか	「抗うつ薬」という分類名がない場合、分類名に余分なものがついてしまっている場合で、修飾語を取ればそれにマッチするような名称になるのであれば、それを使用する
26	正規形フラグが A の場合に修正した際の 入力方法	A → A! とする
27	出現形にtypoがあった場合の、フラグへの入力は [TYP] でよいか	出現形にtypoあり → 出現形flag:[TYP]
28	出現形で ~眼軟膏、~吸入 等用途が限定されている場合、薬効分類には該当する用途のもののみ入力していたのですが、あっていますでしょうか？ 出現形:アシクロビル眼軟膏 正規形:アシクロビル TREE:1.3.1.9_その他の眼科用剤 薬効分類 1319(これに内服注射軟膏等の6250を加える必要があるか)	薬効分類には使用用途に該当する分類のみでよい

29	<p>低分子ヘパリン の正規形について。</p> <ul style="list-style-type: none"> <li>・”低分子ヘパリン”のグループ名、薬効分類名が存在しない。</li> <li>・グループとしてはヘパリンに属する</li> </ul> <p>この場合は、該当分類なしで正規形(-1)とするか。ヘパリンとすべきか。</p>	<p>低分子ヘパリン、未分画ヘパリンも、正規形は”ヘパリン” でまとめる</p>
30	<p>出現形:元素 の場合の正規形入力規則</p>	<p>基本は:-1_非医薬品</p> <p>カルシウム、カリウム、マグネシウム、ヨウ素、鉄 セレン、リン、亜鉛、銅:補給目的とする</p> <p>例外として バリウム:硫酸バリウム, プラチナ:白金化合物</p>
31	<p>出現形:ガドリニウム(元素) の場合の正規形入力規則</p>	<p>例外として ガドリニウム:造影剤 として扱う</p>
32	<p>正規形 一般名:アファチニブマレイン酸塩 現在の入力:アファチニブ flag A のように~マレイン酸、~臭化物等が抜けている場合は flagA でも修正していいか</p>	<p>修正してよい</p>
33	<p>出現形:リチウム(元素)の例外追加の可否</p>	<p>例外に追加 出現形:リチウム → 正規形:炭酸リチウム</p>
34	<p>出現形:ナトリウム(元素)の例外追加の可否</p>	<p>例外に追加 出現形:ナトリウム → 塩化ナトリウム製剤</p> <p>(塩化ナトリウム10%注、塩化ナトリウム補正用、塩化ナトリウム等の意味)薬効分類としては、生理食塩水も同じ番号となる</p>
35	<p>微量元素製剤(高カロリー輸液投与時)の扱いについて</p> <p>微量元素製剤については 鉄、亜鉛、銅、マンガン、ヨウ素、それぞれの補給としては考えない、でよいか。 (亜鉛の場合、ヒスチジン亜鉛と微量元素製剤が結果として同じ薬効分類に入ってしまうが)</p>	<p>微量元素製剤に含まれる 鉄、亜鉛、銅、マンガン、ヨウ素、は、それぞれの補給としては考えない</p> <p>例 出現形:鉄→ 薬効分類:3222(鉄化合物)3229(微量元素製剤)</p>

36	正規形が複数となる場合にどうするか？	<p>3個までなら、          同格：一つを選択・・・ルール：一般名として適切であるもの(例、アスピリンには、アスピリン、アスパロン、アセチルサリチル酸があるが、成分名として一般的なアセチルサリチル酸を選択する)から頻度の多いもの、          配合剤：+で繋ぐ。          曖昧な場合：;で繋ぐ。</p> <p>4つ以上：          同格：一つを選択          配合剤(combination drugs):[COM],          曖昧な(ambiguous)場合：正規化先が複数あって曖昧な場合(まとめられない場合(あまりない?))は[AMB]</p> <p>同格で一つに選択する際の留意点          この際、それまでに選んでいた代表的な正規形がある場合にはそれを選ぶようにする。(例、出現形：アスピリン、一般名：アセチルサリチル酸とすでに選んでいた場合、出現形：低用量アスピリン、一般名：アセチルサリチル酸、とする</p>
37	<p>生薬の正規形についてですが、これまでは漢字で入力していました。          局法名がカタカナ表記になっているため、カタカナに統一するべきか</p>	生薬の正規形：カタカナ表記
38	正規形：キノロン、ニューキノロン、フルオロキノロンの使い分けについて	ニューキノロン系は(フルオロキノロンではなく)ニューキノロン系抗生物質で、それにあてはまらないキノロン系はキノロン系
39	<p>出現形：吸入ステロイド 正規形：副腎皮質ステロイド と同じく          出現形：新規経口抗凝固薬 正規形：抗凝固薬としてよいか</p>	出現形：新規経口抗凝固薬 正規形：抗凝固薬としてよい
40	<p>出現形：～製剤、～抗体、～薬 など商品名、成分名ではない場合          正規形：～製剤、～抗体、～薬として出現形を優先するか、もし薬剤が1～3種類しか存在しない場合はできるだけ薬剤名を優先したほうがよいか</p>	出現形が、商品名、成分名でない場合には、出現形を優先する
41	<p>出現形：KM の場合、正規形：カナマイシン または 正規形：KM(胃腸薬)が正規形として候補だが、この場合(;)でつなげてよいのか。          略称ではない“KM”を優先すべきか</p>	略称ではないKMを優先
42	出現形：プレドニゾロン(P)の場合、正規形をERR、-1とするべきか	<p>抽出ミスとは考えにくい          出現形：プレドニゾロン(P) 正規形：プレドニゾロンとする</p>

43	<p>ヨウ素の扱いについてですが、 ID 2695 無機ヨード:無機ヨードすべて ID 1895 ヨウ素:補給としてのヨウ素剤 ID 718 ヨード:補給としてのヨウ素剤 と、現在しています。ヨードカプセル(放射性医薬品 甲状腺疾患診断薬)もあるのですが、ヨードは、元素としてのヨウ素としての扱いでよいか</p>	<p>ヨウ素に関しては、商品名として、ヨウ素~~~, 一般名として、「ヨウ素」単体があるということで、消毒、として扱う。 出現形 ヨウ素 :消毒として 出現形 ヨード :ヨードカプセル 出現形 無機ヨード: 無機ヨードすべてとする</p>
44	<p>ID2984 クロラムピシル 等動物用医薬品も、通常通り入力してよいか</p>	<p>動物用医薬品も通常通り入力</p>
45	<p>一般用医薬品 (例 ID 3114 ニコチンガム) の場合は (正規形:-1_治療 or -1_非医薬品)とするか、正規形:成分名 としてTREE、薬効分類:-1とするか</p>	<p>一般用医薬品についても、成分名が入るものについては、正規形にいれる</p>
46	<p>ホルモンについて 出現形:黄体ホルモン 正規形:プロゲステロン(生体内で産生される天然のホルモン) 出現形:黄体ホルモン製剤 正規形:黄体ホルモン製剤(添付文書の分類名より) と考えてよいか。 keggでは <a href="https://www.kegg.jp/medicus-bin/similar_product?kegg_drug=DG01986">https://www.kegg.jp/medicus-bin/similar_product?kegg_drug=DG01986</a> グループ名 卵胞ホルモン となっており、今日の治療薬でも 卵胞ホルモンという分類名?の中に各種卵胞ホルモン製剤が書かれている。 黄体ホルモン を分類名としてとらえて 出現形:黄体ホルモン 正規形:黄体ホルモン製剤 または 黄体ホルモン としたほうがよいか</p>	<p>卵胞ホルモン、黄体ホルモン、男性ホルモン等は分類名として扱う 例 出現形:卵胞ホルモン 正規形:卵胞ホルモン製剤 (卵胞ホルモンを分類名として扱う。正規形に 結合型エストロゲンとは書かない)  出現形:エストロゲン 正規形:結合型エストロゲン (卵胞ホルモン(天然型)=エストロゲンだが、エストロゲンの場合は一般名として扱う)  出現形:性腺刺激ホルモン 正規形:性腺刺激ホルモン製剤  出現形:ヒト絨毛性性腺刺激ホルモン 正規形:ヒト絨毛性性腺刺激ホルモン (出現形自体が一般名)</p>
47	<p>ID3619 出現形:メイロン(商品名) 正規形:炭酸水素ナトリウム のように、出現形が商品名の場合のTREE、薬効分類の入力方法</p>	<p>TREE, 薬効分類は出現形をもとにして、商品名を含む分類のみを入力する</p>
48	<p>出現形が "がん化学療法レジメン" の内容の場合の入力規則</p>	<p>・レジメンは正規形:-1_レジメン, と入れる ・「と」、「,」で複数の医薬品が繋がれている場合には基本的には抽出ミスとして正規形は「ERR」となるが、複数の医薬品の組み合わせがなんらかのレジメンを意味する場合には「-1_レジメン」とする。</p>

49	3862 出現形:カルジオリピン は 正規形:-1 flag A となっていました。 出現形が疾患、検査、等に関係がありそうな生体成分の場合は、正規形:-1_治療、としていたのですが、この理解で正しかったでしょうか？	出現形が疾患、検査、等に関係がありそうな生体成分の場合は、正規形:-1_治療、で正しい
50	タリウム、テクネチウム等の放射性医薬品として使われる元素については、今後は確認せず元素の例外として扱ってよろしいでしょうか。	確認なしで、例外としてよい
51	出現形:丸山ワクチン(日本医科大学から、有償治験として希望する患者さんに提供される。免疫療法)の場合、 丸山ワクチン自体は、医薬品として承認されておらず、その成分 結核菌熱水抽出物 を使って“アンサー-20”という医薬品が販売されています。 出現形:丸山ワクチン 正規形:結核菌熱水抽出物 TREE & 薬効分類:アンサー-20のものとしていますが、問題ありそうでしょうか。	問題なし
52	出現形:アミノレバンEN 正規形:アミノレバン(総称) と現在なっていますが、この場合(総称)がわかったほうがよければ 括弧 を残そうと思えますがいかがでしょうか。	( )はできるだけなくす方針なので、取ってしまってもよい
53	出現形:骨粗しょう症治療、出現形:漢方治療等、出現形が、~治療、となっている場合 正規形:-1_治療 正規形:骨粗しょう症治療薬、漢方薬 どちらがよいか。	出現形:~治療 の場合は 正規形:-1_治療 とする
54	4684 出現形:セロトニン作動薬 正規形: 5-HT1A-受容体作動薬 flagA となっていました。薬効分類に5-HT3、5HT4、も含めたほうがよいと考え、正規形:—1 に変更。 また、 125 出現形:カルシウム拮抗薬 は、以前はカルシウム拮抗薬の大半を占める“ジヒドロピリジン系カルシウム拮抗薬”を正規形としていたが、薬効分類にベンゾジアゼピン系カルシウム拮抗薬(ジルチアゼムのみ)も含めて“カルシウム拮抗薬”としている。 1つの系統だけに限定しない方向でよいか。	問題なし
55	アミノレバンの(総称名)の入力について確認しましたが、正規形に総称名(商品名)を使用すること自体は大丈夫なのでしょうか。 kegg記載 アミノレバンEN配合散 一般名:該当するものなし 総称名:アミノレバン 成分が多いため、一般名なしとなっているのであれば [COM] とするべきか	出現形:アミノレバン 正規形:[COM] とする

56	放射性の有無で、安定ヨウ素、放射性ヨウ素、と分けることが多いため、“ヨード剤”の場合は放射性ヨウ素を含める方向でよいでしょうか。 出現形:ヨード剤 正規形:—1 薬効分類:3221(ヨウ化カリウム、ヨウ素レシチン);4300(ヨウ化ナトリウム(放射性医薬品))	問題なし ヨウ素についてこれまでの出現形正規形 出現形 ヨウ素 :消毒として 出現形 ヨード :ヨードカプセル 出現形 無機ヨード:無機ヨードすべて 出現形 ヨード剤 :-1 薬効分類:3221(ヨウ化カリウム、ヨウ素レシチン);4300(ヨウ化ナトリウム(放射性医薬品))
57	250 出現形:エリスロマイシン 正規化先:(エリスロマイシン ; エリスロマイシンラクトビオン酸塩 ; エリスロマイシンエチルコハク酸エステル ; エリスロマイシステアリン酸塩) 4つ以上の正規化先があり規則では 正規形:[AMB] となる。 正規形:エリスロマイシン にしたくなりますが、先発の“エリスロシン”が成分名“エリスロマイシンエチルコハク酸エステル”でもあるので やはり正規形:[AMB]とすべきかなとは思いますが。規則通りでよろしいでしょうか。	エリスロマイシンの構造を保っている、エリスロマイシン、エリスロマイシステアリン酸のみを正規形とする。エステルなど共有結合により構造変化しているものについては、別とみなし正規形には入れない
58	976 出現形:サリチル酸 サリチル酸は～サリチル酸等も含めると正規形先が多数になるため正規形:サリチル酸 として扱ってもよいでしょうか 1564 出現形:フェノール 2882 出現形:クエン酸 についても同様に	問題なし
59	正規形が、前駆体のものや様々な修飾をされているものも含まれる。 例 出現形:セチリジン 正規形:セチリジン塩酸塩;レボセチリジン塩酸塩 例 1525 出現形:ジクロフェナク 正規形:ジクロフェナクナトリウム;ジクロフェナクエタルヒアルロン酸ナトリウム 例 2962 出現形:モルヒネ 正規形:モルヒネ塩酸塩水和物;アポモルヒネ塩酸塩水和物 (アポモルヒネ塩酸塩水和物はモルヒネから合成されるが、モルヒネ骨格もない。)	出現形:セチリジン 正規形:セチリジン塩酸塩 出現形:ジクロフェナク 正規形:ジクロフェナクナトリウム 出現形:モルヒネ 正規形:モルヒネ塩酸塩水和物 それぞれ、これらの出現形からレボセチリジン塩酸塩、ジクロフェナクエタルヒアルロン酸ナトリウム、アポモルヒネ塩酸塩水和物を結びつけなくても良いかと思いました。どうでしょうか？ (これらの出現形の各々から、それぞれの正規形を呼ばないのではないか、と思ったため。 例えば、レボセチリジン塩酸塩のことをセチリジンとは言わないのではないか?)
60	5256 出現形:非ステロイド性抗炎症薬(NSAIDs) の場合、とじ括弧がないため 正規形は [ERR]となっています。 [ERR]のままでもよいでしょうか。	問題なし
61	5309 出現形:CF	正規形優先順位は -1_レジメン, -1_治療の 順

	<p>大腸内視鏡検査(ColonoFiberscopy:CF) または、CF(シスプラチン+5-FU)療法 の選択肢があります。</p> <p>略語はかなりの数があり、今後もこのような例がありそうなので、-1_治療、1_レジメンの優先順位は</p>	
--	--	--

出現形よみのメンテナンス作業時に発生したQA.

番号	質問	回答
1	Aはえーではなくえいでしょうか？	Aはえーではなくえいの方がよさそうです。
2	例えば抗CD20モノクローナル抗体の数字はにゼロでしょうか？にじゅうでしょうか？英語読みのつーゼロでしょうか？	
3	同じ単語が繰り返し出てくる時があるのですがスキップしても良いのでしょうか？	基本的にはスキップせずをお願いします。例があれば教えてください。
4	非ステロイド抗炎症薬(NSAIDs)やインターフェロン(IFN)- $\alpha$ 、アンジオテンシン受容体拮抗薬(ARB)が繰り返し出てきたように思います。スキップせずにやっています。	同じ単語が繰り返しの場合もありますが、微妙に表現が異なっているので(括弧の部分がなかったり、括弧内の表現が違ったりなど...)、スキップせずにとすることで継続をお願いします。
5	2151行はA列が酸素-亜酸化窒素-セボフルランなのですが元素と麻酔薬なので非医薬品にするべきか-1にするべきかわからず	2151行は[ERR]のままでOKです。複数のものがカンマやハイフンや・などで繋がっている場合は[ERR]としていただければと思います。
6	G列がERRとなっていなくても連日、他薬などの単語の場合もERRで良いでしょうか？ また、4236行目の鍼灸や4246行目のリネンはERRになっていないのですがG列がプレドニゾロンやザナミビル水和物なっています。 こういった物は、-1_非医薬品に変更して良いのでしょうか？	変更をお願いします
7	4536行目ペグインターフェロン+リバビリンのように+は「と」と呼ぶのでしょうか？	このようなケースはERRとして読みは無視でオッケーです。
8	E列がEのものでG列の出現形が明らかに合ってなさそうなものがありますが、そのまま大丈夫でしょうか？ 例えば手袋やタオルの出現形が医薬品の名前やタルクになっているのですが、-1_非医薬品に変えても良いのでしょうか？	変更をお願いします

## 検査データ

検査データのメンテナンス作業時に発生したQA.

No	質問	回答
1	栄養関係の用語を残した方がよいですか？ 例：タンパク摂取量、水分摂取量、食事摂取量など	念のため-1で残しておいていただけますか？
2	TREEのラベル(1~9)に該当しない場合の対応 現時点では、多少項目に合わないものも1~9を選択することもあるようです。基本的に1~9に該当しなければ「10_不明」でよいでしょうか？多くは「不明」ではないため、「その他」を選択できればよいのですが。	
3	出現形：喀痰、髄液、羊水、血液など(検体の種類のみ)はERRとしてよいか	検体の種類の場合は、残してもらえますか？
4	【出現形に検体が明記されている事例】※今回判断に迷っている事例 出現形：動脈血pH、尿pH、胃内pH、髄液pHなど 「検体名+pH」 正規形は、いずれも「pH」とするか、「動脈血pH」「尿pH」のように主要なものは分けてもよいものでしょうか？	主要なものは分けてください