

厚生労働科学研究費補助金（政策科学総合研究事業）
分担研究報告書

小児 AI 開発に向けた小児版医薬品辞書の整備に関する研究

研究分担者 荒牧英治

奈良先端科学技術大学院大学 研究推進機構情報科学研究科 特任認定教授

研究協力者 若宮翔子（同研究科 特任助教）

研究要旨

小児医療人工知能（小児 AI）開発のためには、小児医薬品の適正使用および安全対策推進のための情報を、添付文書や医療文書などのリソースから幅広く取得し、データベースの整備を行うことが重要である。添付文書や医療文書からの情報抽出は、自然言語処理技術を適用することにより実現可能であるが、そのためには、同一の医薬品に対する表現のバリエーションを正規化する必要がある。本研究では、医薬品の一般名と臨床現場などで用いられる表現を対応付けた医薬品辞書「百薬辞書」から、小児領域で使用実態のある医薬品に関するデータを抽出して「百薬辞書：小児版」とした。さらに、「百薬辞書：小児版」において、各医薬品の一般名に対応する表現のバリエーションを調査した結果、対応する表現数が 100 を超えるものもあることを示した。「百薬辞書：小児版」は、今後の小児 AI を支える自然言語技術開発のための重要なリソースになると考えられる。

A．研究目的

昨今、様々な医療情報や医薬品情報が電子化されつつある一方で、医薬品の適正使用に欠かせない薬剤添付文書（以下、添付文書）については整備が遅れている。特に、小児領域で汎用されている医療用医薬品の整備の遅れは顕著であり、現在添付文書に小児の用法・用量が明確に記載されていないものが全体の 60～70%を占めているといわれており、小児適応が明記されていない薬剤で薬物療法を実施せざるを得ないことが多くある。このような状況下において、小児医療人工知能（小児 AI）開発のためには、小児医薬品の適正使用および安全対策推進のための情報を得ることのできるデータベースの整備が喫緊の課題である。

小児医薬品の適正使用および安全対策推進のための情報は、添付文書や電子カルテや症例報告などの医療文書に自然言語処理技術を適用することで抽出可能である。しかし、昨年度に我々が実施した調査では、添付文書のような形式的な文書であっても、表記のゆれや曖昧性が存在することが明らかになっている。実際の医療現場で記載される医療文書につい

ても同様であり、自然言語処理を適用する際の障壁となっている。

そのため、今年度は、添付文書や医療文書における表記のゆれや曖昧性の実態調査およびその解消（以下、正規化）のために、医薬品の一般名と臨床現場などで用いられる表現を対応付けた医薬品辞書「百薬辞書」から、小児領域で使用実態のある医薬品を抽出した。そして、各医薬品に表現のバリエーションがどれくらい存在するかを調査した。

B．研究方法

奈良先端科学技術大学院大学 ソーシャル・コンピューティング研究室では、研究室で収集した医療文書や、クラウドソーシングによる一般ユーザへのアンケートを通して、大量の医薬品表現を収集している。そして、収集した各医薬品表現（以下、出現形）に対応するよみがな、一般名、メーカー名等のデータへの紐付けを行い、医薬品辞書「百薬辞書」を構築している。2020 年 3 月現在、百薬辞書には 38,428 語の出現形が収録されている。「百薬辞書」を用いることで、臨床現場などで用いられ

る医薬品表現を一般名へ変換することが可能である。

今年度の分担研究では、小児領域で使用実態のある医薬品を「百薬辞書」から抽出した。本報告書では、これを「百薬辞書：小児版」と呼ぶ。次に、「百薬辞書：小児版」において、各医薬品にどれくらいの表現のバリエーションがあるかを調査した。具体的には、厚生労働科学研究（小児 AI 研究）医薬品使用実態調査に記載されている医薬品の一般名をもとに「百薬辞書：小児版」の一般名を検索し、医薬品表現を抽出した。なお、医薬品使用実態調査の抽出期間は 2018 年 1 月～2018 年 12 月、抽出年齢は 15 歳以下である。

対象とした医薬品は、1: 外用剤を除く抗生物質（抗菌剤）112 薬品（一般名の重複を除くと 97 薬品）、2: アレルギー薬（内服、注射）33 薬品、3: アレルギー薬（外用）20 薬品であった。

（倫理面への配慮）

本研究では、個人情報と紐付かないデータを扱っており、倫理面への配慮は必要としなかった。

C . 研究結果

小児領域で使用実態のある医薬品の一般名を「百薬辞書」で検索した結果、3,682 語の出現形が一致（部分一致も含む）した。この結果である「百薬辞書：小児版」に対し、検索に用いた一般名をフラグとして付与した。また、研究室で収集した電子カルテにおける頻度をもとに「高」「中」「低」「レア」「出現なし」の頻度レベルも付与した。

次に、各医薬品にどれくらい表現のバリエーションがあるかをまとめた。表 1 に表現のバリエーションが多い 10 医薬品を示す。ここで、「一般名」は医薬品のカテゴリを示す数字（1: 外用剤を除く抗生物質（抗菌剤）、2: アレルギー薬（内服、注射）、3: アレルギー薬（外用））と医薬品使用実態調査に記載されている一般名からなる。表 1 において、1:レボフロキサシン、2:モンテルカストナトリウム、2:オロ

パタジン塩酸塩の 3 つについては、一般名に対応する表現数が 100 を超えており、50 を超えるものも 14 存在した。例えば、1:レボフロキサシンの出現形としては、「レボフロキサシン」「Lev」「LVFX」（頻度レベル：高）、「クラビット」（頻度レベル：中）、「Levo」（頻度レベル：低）、「レボフルロキサシン」「レボフラキサシン」（頻度レベル：レア）などがあった。2:モンテルカストナトリウムの出現形としては、「モンテルカスト」（頻度レベル：中）、「モンテカルスト」（頻度レベル：低）、「キプレス」（頻度レベル：レア）などがあった。2:オロパタジン塩酸塩の出現形としては、「オロパタジン」（頻度レベル：低）、「アレロック」（頻度レベル：レア）などがあった。一方、一般名に対応する表現数が 1 のものは、「1:エノキサシン」、「1:セフォペラゾン」、「1:リボスタマイシン」の 3 つのみであった。全医薬品の表現数の平均は 24.04、分散は 22.94 であった。カテゴリ別の結果では、1: 外用剤を除く抗生物質（抗菌剤）の表現数の平均と分散はそれぞれ 20.85 と 19.90、2: アレルギー薬（内服、注射）の表現数の平均と分散はそれぞれ 36.55 と 30.75、3: アレルギー薬（外用）の表現数の平均と分散はそれぞれ 18.7 と 13.19 であった。

表 1 . 表現のバリエーションが多い一般名（上位 10 薬品）

一般名	表現数
1:レボフロキサシン	151
2:モンテルカストナトリウム	138
2:オロパタジン塩酸塩	103
2:フェキソフェナジン	87
1:クラリスロマイシン	71
2:プレドニゾロン	70
2:ケトチフェンフマル酸塩	67
3:0.1% タクロリムス軟膏	67
2:エバスチン	60
2:セチリジン塩酸塩	59

D . 考察

医薬品使用実態調査の結果をもとに「百薬辞書」から抽出した「百薬辞書：小児版」を調

査したところ、小児領域で使用実態のある医薬品には様々な表現が対応することが明らかになった。「百薬辞書：小児版」で医療の現場で実際に用いられる医薬品表現を一般名に変換することで、小児医薬品の適正使用および安全対策推進のための情報を効率的に取得・集約が可能になると期待される。

E．結論

小児 AI 開発のためには、添付文書や医療文書から小児医薬品の適正使用および安全対策推進のための情報を取得し、データベースを整備することが重要である。自然言語処理技術を適用することで、添付文書や医療文書からの情報抽出が可能であるが、網羅性を高めるためには、多様な医薬品表現を正規化して処理するための辞書の整備が必須である。そのため、医薬品の一般名と臨床現場などで用いられる表現を対応付けた「百薬辞書」から、小児領域で使用実態のある医薬品を抽出した「百薬辞書：小児版」を整備した。さらに、各医薬品の一般名に対応する表現のバリエーションを調査し、対応する表現数が 100 を超えるものもあることを報告した。

F．健康危険情報

該当なし

G．研究発表

1. 論文発表 なし
2. 学会発表 なし
(発表誌名巻号・頁・発行年等も記入)
なし

H．知的財産権の出願・登録状況

(予定を含む。)

1. 特許取得 なし
2. 実用新案登録 なし
3. その他 なし

百薬辞書：小児版
- 薬品別一覧 -

項目名	説明
一般名	「厚生労働科学研究（小児 AI 研究）医薬品使用実態調査」に記載されている一般名
表記ゆれ個数	一般名に該当する「百薬辞書」に記載されている医薬品の個数

1. 外用剤を除く抗生物質（抗菌剤）

一般名	表記ゆれ個数	一般名	表記ゆれ個数
アジスロマイシン	48	ジベカシン	10
アズトレオナム	5	ジョサマイシン	11
アセチルスピラマイシン	6	ストレプトマイシン	9
アミカシン	24	スペクチノマイシン	6
アモキシシリン・クラブラン酸(14:1) 1)アモキシシリン・クラブラン酸2:1)	16	スルタミシリン	10
アモキシシリン	41	スルバクタム・セフォペラゾン(1:1)	35
アルベカシン	26	スルファメトキサゾール・トリメトプリム	27
アンピシリン	10	セファクロル	24
アンピシリン・クロキサシリン	14	セファゾリン	41
アンピシリン・スルバクタム(2:1)	54	セファレキシン	30
イセパマイシン	14	セフィキシム	10
イソニアジド	14	セフェピム	25
イソニアジドメタンスルホン酸ナトリウム	7	セフォゾプラン	14
イミペネム・シラスタチン(1:1)	34	セフォタキシム	13
エタンブトール	25	セフォチアム	37
エチオナミド	4	セフォチアムヘキセチル	5
エノキサシン	1	セフォペラゾン	1
エリスロマイシン	15	セフカペンピボキシル	40
エンピオマイシン	5	セフジトレンピボキシル	24
オフロキサシン	40	セフジニル	36
ガチフロキサシン	7	セフトジジム	31
カナマイシン	15	セフチブテン	6
キヌプリスチン・ダルホプリスチン	4	セフテラムピボキシル	13
クラリスロマイシン	71	セフトリアキソン	45
クリンダマイシン	11	セフピロム	12
クロラムフェニコール	16	セフメタゾール	34
ゲンタマイシン	37	セフメノキシム	11
コリスチン	14	セフロキサジン	5
サイクロセリン	4	セフロキシムアキセチル	5
シプロフロキサシン	32	タゾバクタム・ピペラシリン(1:8)	49
		ダブトマイシン	6
		テイコプラニン	21
		テトラサイクリン	19
		テビペネムピボキシル	4

デメチルクロルテトラサイクリン	14
ドキシサイクリン	8
トスフロキサシン	39
トブラマイシン	12
ドリペネム	14
ナリジクス酸	10
ノルフロキサシン	29
バカンピシリン	5
パズフロキサシン	15
パニペネム・ベタミプロン(1:1)	12
バンコマイシン	38
ピアペネム	7
ピペラシリン	41
ピラジナミド	13
ファロペネム	10
ブルリフロキサシン	6
フロモキシフ	10
ベンジルペニシリン	22
ベンジルペニシリンベンザチン	7
ホスホマイシン	46
ポリミキシン B	15
ミノサイクリン	44
メロペネム	51
モキシフロキサシン	14
ラタモキシフ	7
リネゾリド	20
リファンピシン	17
リボスタマイシン	1
リンコマイシン	18
レボフロキサシン	151
ロキシスロマイシン	12
ロメフロキサシン	11

2. アレルギー薬(内服・注射)

一般名	表記ゆれ個数
d-クロルフェニラミンマレイン酸塩	21
アゼラスチン塩酸塩	20
エバスチン	60
エピナスチン塩酸塩	53
オキサトミド	33
オロパタジン塩酸塩	103
クレマスチンフマル酸塩	23
ケトチフェンフマル酸塩	67
ジフェンヒドラミン塩酸塩	11

シプロヘプタジン塩酸塩水和物	12
スプラタストシル酸塩	18
セチリジン塩酸塩	59
デキサメサゾン	59
デスロラタジン	4
トラニラスト	29
ヒドロキシジンパモ酸塩	14
ヒドロコルチゾンコハク酸エステルナトリウム	15
ピラスチン	4
フェキソフェナジン	87
プランルカスト水和物	42
プレドニゾン	70
ベタメタゾン	20
ベタメタゾン/d-クロルフェニラミンマレイン酸塩	19
ベボタスチンベジル酸塩	37
ベミロラストカリウム	32
メクタジン	27
メチルプレドニゾン	32
メチルプレドニゾンコハク酸エステルナトリウム	28
モンテルカストナトリウム	138
ラマトロパン	8
ルバタジン	4
レボセチリジン塩酸塩	8
2:ロラタジン	49

3. アレルギー薬(外用)

一般名	表記ゆれ個数
0.025%フルオシノロンアセトニド	12
0.025%ベクロメタゾンプロピオン酸エステル	27
0.05%クロベタゾールプロピオン酸エステル	31
0.05%ジフルプレドナート	14
0.05%ジフロラゾン酢酸エステル	18
0.05%酪酸プロピオン酸ベタメタゾン	18
0.1%アルクロメタゾンプロピオン酸エステル	8
0.1%クロベタゾン酪酸エステル	17
0.1%ヒドロコルチゾン酪酸エステル	9
0.1%プロピオン酸デキサメタゾン	18
0.1%モメタゾンフランカルボン酸エ	19

ステル	
0.1%吉草酸ジフルコルトロン	15
0.1%酪酸プロピオン酸ヒドロコルチゾン	18
0.1%タクロリムス軟膏	67
0.12%デキサメタゾン吉草酸エステル	10
0.12%ベタメタゾン吉草酸エステル	22
0.12%ベタメタゾン吉草酸エステル/ ゲンタマイシン硫酸塩	23
0.3%吉草酸酢酸プレドニゾロン	17
フルオシノロンアセトニド/フラジオマイシン硫酸塩	6
メチルプレドニゾロン/フラジオマイシン硫酸塩	5