

患者が執筆した闘病記の医学的な利用可能性に関する研究

研究分担者 荒牧英治 京都大学デザイン学ユニット 特定准教授
森田瑞樹 東京大学知の構造化センター 特任研究員

研究要旨

患者主体の患者レジストリでは、患者によって書かれた文章の収集が容易になることが想定される。患者またはその関係者が記述し Web 上で公開している手記、文章、ブログ（以下、闘病記）はたいへん多く存在するが、従来データとして使われていた検査データ、診療記録、看護記録とくらべて、正確性に劣る恐れもある。そこで本研究では、患者が記述した文章の特徴を解析した。167 サイトの闘病記を収集し、そこに記述された症状表現を調査し、その量、また、医学用語として扱われている語彙とどの程度一致するかを調査した。この調査は、医師の記述した文章（病歴報告）においても行い、両者の比較も行った。調査の結果、闘病記における症状表現の数は医師文章よりも少なく 2/5 程度であった。しかし、それが医学用語と一致する割合は約 60%であり、これは医師文章の割合（46%）よりも多かった。特に、不安、倦怠感、不眠症などの QOL に関する情報が多かった。

A. 研究目的

最近では医療者が記述するデータだけでなく、患者が自ら発する文章から臨床情報を抽出する試みが注目を集め始めている。患者の記述する文章というと、ブログ、ソーシャルメディアからメールまで膨大な種類と量になり、医療者の文章量を遥かに凌駕する。

このデータの可能性は以下の 2 点にあると考える。

- 膨大な量： いずれも患者自身がデータを記述するため膨大なデータ量が得られる。
- 質的な異なり： 従来、看護記録に記録されるにとどまっていた精神面や、QOL に関する記述が豊富に得られる。

このような有用性もあるものの、医学的な正確性と観点では疑問が残る。例えば、何による痛みか勘違いして記述する錯誤や、薬を飲み忘れていたのを恥ずかしくて隠して飲んだように記述する可能性もある（正確性の問題）、また、正確に記述したつもりでも、医学的な用語との乖離が大きく、既存の集計とマッチしない場合がある（用語の差異の問題）。例えば、ある胃癌の闘病記には、「指先がピリピリ痺れています。」と言及されているが、副作用集計という観点からは、「末梢神経障害」として集計されるのが適切である。このように、ある種の変換が必要であれば、非常にコストがかかることになる。

ここで前者の正確性の問題は真の症状との対比調査が必要で困難であるが、後者の用語の差異の問題は患者文章を大量に収集することで検討可能である。

そこで、本研究では、患者が執筆したデータとして、167 サイトの闘病記を材料とし、そこに記述された症状表現を調査し、その量、また、医学用語として扱われている語彙とどの程度一致するかを調査する。この調査は、医師の記述した文章（病歴報告）においても行い、両者の比較を行った。

B. 研究方法

まず、日本人の死因の主要な疾病の統計から、闘病が長く続く、癌、認知症、鬱を対象とした。ここで闘病記の検索サイトにて検索を行い、上位に得られたものより順に収集した。この際、単テーマ、多テーマの区別なく行った。この結果、167 の闘病記が得られ、各闘病記から、罹患時を起点として 5 記事（ブログの場合はエントリ）をランダムに抽出した。これは 24715 文に相当し、1 記事あたり、29.6 文を含んでいる。これと比較するための模擬病歴は、ワークショップ (<http://mednlp.jp/medistj-ja/>) で配布されているデータを用いた。

これらに対して、2 名の医療者（検査技師、治験レポート集計者）が人手でこれを精査し、患者の症状が読み取れる箇所についてマークアップを行った。これは約 4 ヶ月を要した。

表 1：結果

	闘病記	病歴報告（模擬病歴報告）
症状文リソース	9.0% (=2247 文 /24715 文)	2.4% (=83 文/3366 文)
利用可能リソース	39.7% (=1299 語 /3269 語)	30.0% (=799 語 /2655 語)
自動変換リソース	19.6% (=643 語 /3269 語)	16.0% (=426 語 /2655 語)
手動変換リソース	40.6% (=1327 語 /3269 語)	53.8% (=1430 語 /2655 語)

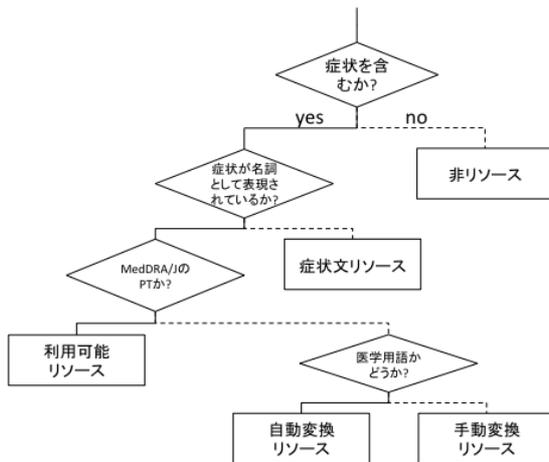


図 1：リソースの分類

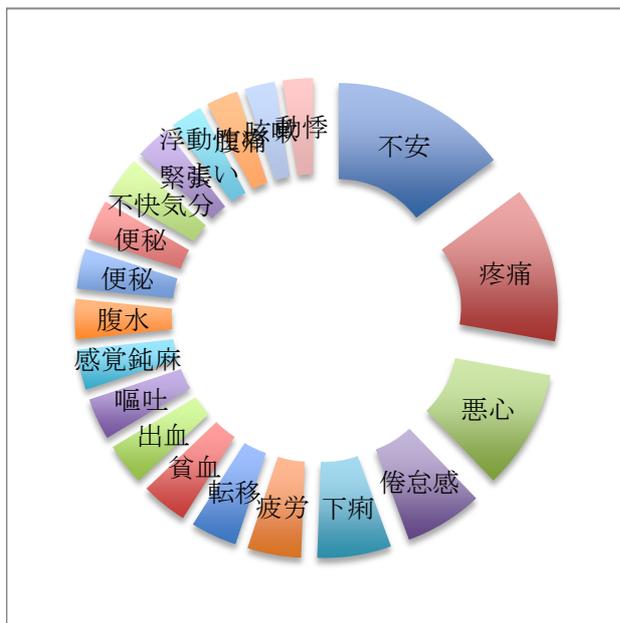


図 2：主要な症状の頻度
(MedDRA/J の PT 単位で集計)

C. 研究結果

ここで、闘病記に出現する副作用レポートとしての利用の観点から分類してみる。

- **副作用用語**：もっとも簡単なものは、副作用

レポートで用いられる語「悪性新生物」がそのまま副作用報告の用語である MedDRA/J の PT とマッチする場合である。これを**利用可能リソース**と呼ぶ。

- **医学用語**：次に利用が容易なのは、「胃悪性新生物」など PT とは異なるレベル（より細かい、またはより荒い）の用語が使用される場合である。この場合は、他の医学リソース、例えば MedDRA/J の Lower Level Term (以降、LLT と呼ぶ) や ICD など、何らかの医療用語にマッチすると考えられ、UMLS など既存のリソースを用いて変換が可能である。これを**自動変換リソース**と呼ぶ。
- **一般用語**：さらに、「ガン（カタカナ表記は正式な用語でない）」など、医療用語として用いられていない用語が用いられる場合がある。この場合は、変換リソースを新たに構築する必要がある。これを**手動変換リソース**と呼ぶ。
- **フレーズ／文**：最後に、もっとも利用が難しいのは、例えば「少ししか食べられない感じです（食欲不振）」など、名詞として表現されず、句や文として副作用症状が表現される場合である。これを**症状文リソース**と呼ぶ。

このような関係を図 1 に示す。また各リソースの割合を表 1 に示す。図 2 に主要な症状の頻度を示す。

D. 考察

調査の結果、闘病記に含まれる症状表現は医師文章よりも少なく、2/5 程度である。しかし、それが医学用語と一致する割合は約 60% であり、これは医師文章の割合（46%）よりも多い。

特に、不安、倦怠感、不眠症などの QOL に関する情報が多く、時系列の整合性など、いくつかの問題点に留意して執筆すれば膨大なデータとして活用できる可能性があると思われる。

作業にあたった医療者のアンケートによると、闘病記を副作用レポートなど医学的に有益なものにするには、以下が重要であること示された。

- **時系列について**：過去を回想しながら記載している闘病記は、過去と現在の記載が混合するため、それぞれの時期が不明瞭にならないように記載する。

- **原因を明確に**：感情的な表現が主立っている闘病記の場合は、その感情や症状の原因が記載されていない場合が見られる（ただ「痛い」「辛い」などと綴ってある）ため、感情や症状の原因となる事象、治療なども併せて記載すると、記載内容が副作用であるか否かの判断にもつながると思われる。
- **副作用について**：副作用レポートにおいては、各副作用の転帰について必ず記録することになっているが、闘病記では転帰に関する記載は少なかった。副作用が発現した時の状況は多くの闘病記で詳細に記載されているので、その後回復したのか、症状が続いているのか等の様子も併せて記載するとよいかと思われる。

E. 結論

患者主導でデータを集計する際に、医学用語と患者の用語の違いが危惧される。しかし、本研究により、一定の精度で変換可能である可能性が示唆された。これに加え、記述指針を事前に説明するなどの配慮を行えば、さらにこの可能性は高まるとと思われる。

F. 研究発表

1. 論文発表

該当なし。

2. 学会発表

該当なし。

G. 知的財産権の出願・登録状況

1. 特許取得

該当なし。

2. 実用新案登録

該当なし。

3. その他

該当なし。