

# 闘病ブログ記事に対する 投与薬剤の奏功情報アノテーション

高山 隼矢, 荒瀬 由紀 (阪大),  
梶原 智之 (愛媛大), Chenhui Chu (京大)

コーパス公開中 : [https://yukiar.github.io/pro\\_jp/](https://yukiar.github.io/pro_jp/)

# 医薬品奏功情報

- 奏功情報＝投与した医薬品について
  - 狙った効果があった
  - 狙った効果がでなかった
  - 副作用があった
  - 副作用がなかった



患者のQoL向上、治療計画策定、医薬品開発において  
貴重なデータ

# これまでの情報源

Pmda 医薬品副作用報告

治験のレポート

医師の診療カルテ

- スケールするのが難しい
- 副作用報告では詳細な情報を収集するのは困難
  - 患者プロフィール
  - どんな状況で発症したか？

# 患者SNSの普及

- [PatientsLikeMe.com](https://PatientsLikeMe.com)
- [TOBYO.jp](https://TOBYO.jp)

闘病体験を共有する  
**TOBYO**

TOBYO図書室  
TOBYO LIBRARY

TOBYOレファレンス  
TOBYO REFERENCE

Myデスク  
MY DESK

TOBYO 辞典

はじめての方へ 開発ブログ

会員登録 ヘルプ ログイン サイトマップ

ようこそゲストさん (会員登録・ログイン)

最大の闘病サイトライブラリー + 世界初の闘病専門検索  
**患者の叡智**

Wisdom of Patients

最新情報 TOBYO収録の闘病記・ブログ件数は6万件に達しました。

闘病記・ブログ収録数

1,504疾患  
63,686サイト  
2022年02月08日現在

d2のお知らせ

患者さんの生の声を聞く

d2  
TOBYO

患者ブログ調査プラットフォーム

詳しくはこちら▶

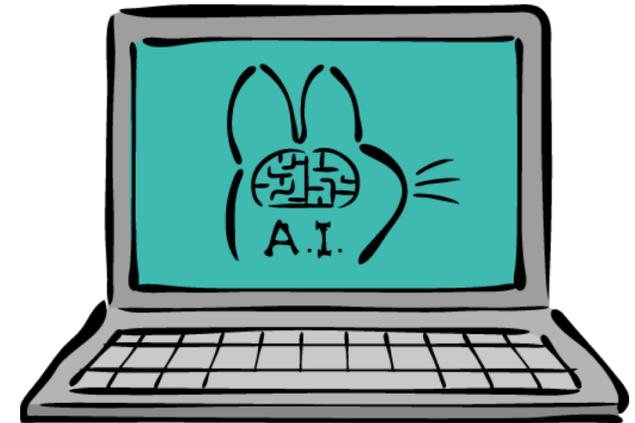
TOBYO 図書室  
闘病記を探す  
詳細はこちら▶  
病名、性別、年代など、いろいろな方向から闘病記を探せます

TOBYO レファレンス  
闘病情報を調べる  
詳細はこちら▶  
みんなで集めた闘病記や闘病情報のノートから情報を検索

TOBYO Myデスク  
情報を整理する  
詳細はこちら▶  
自分用の闘病記の本棚や、闘病情報をまとめるノートが作れます (会員専用)

# 闘病ブログから医薬品奏功情報の自動収集

- 医薬品副作用報告に比べると信頼度は落ちるが、大規模なデータ収集が可能
- 患者プロフィールや副作用発生時の詳細なデータを取得できる
- 患者自身による主観的な症状記述



# 必要な道具

1. 機械学習モデルを訓練するためのアノテーションコーパス (Arase et al. LREC2020)
2. 機械学習モデル (Ohashi et al. ACL2020)
3. 稀なケースを上手く取り扱うモデル (Ohashi et al. ACL2021)

# 必要な道具

1. 機械学習モデルを訓練するためのアノテーションコーパス (Arase et al. LREC2020)
2. 機械学習モデル (Ohashi et al. ACL2020)
3. 稀なケースを上手く取り扱うモデル (Ohashi et al. ACL2021)



# アノテーションコーパス構築

- TOBYO.JPから医薬品名を含むブログ記事を収集
  - 対象疾患：がん、IPF
- 奏功情報を記述したスパンを人手で特定、ラベル付け
  - MedDRAコード
  - ICD10コード
  - 症状に関与したと考えられる医薬品
  - 奏功ラベル



# アノテーション例

遂に梅雨明けたようですね 🍷

今年は梅雨を感じないまま夏が来た感じがします 🐟

今日は、4週間ぶりのがんセンターへの通院です。

娘は午後横浜にバイトの面接に行くと言っていたので

私は娘と待ち合わせして横浜で合流するので電車で病院にきました。

(薬剤性発疹, L270, タルセバ, 副作用) (頭部発疹, R21, タルセバ, 副作用)

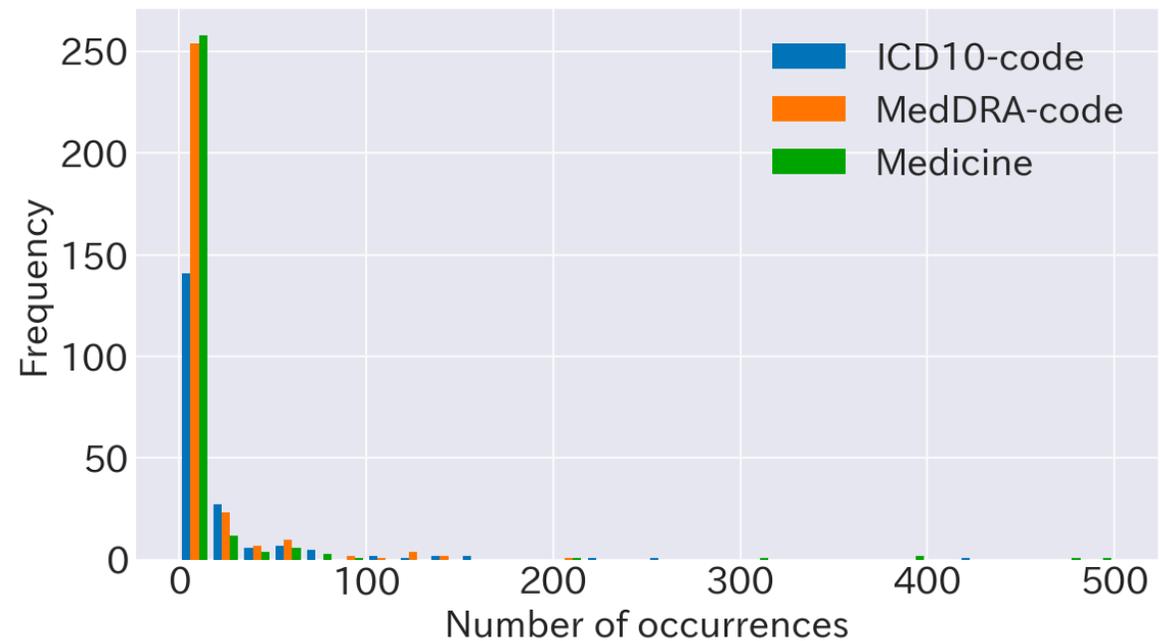
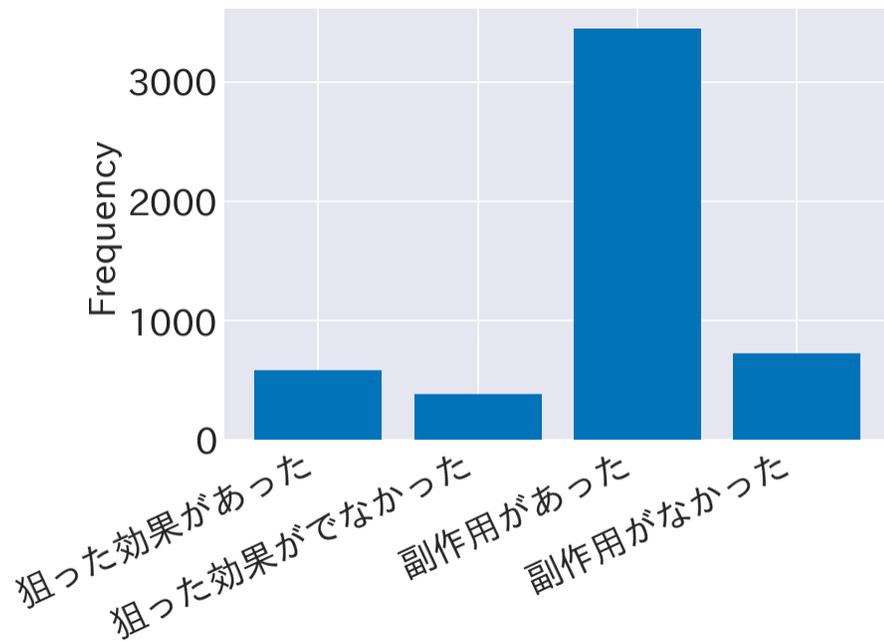
この1か月タルセバくんの副作用発疹が強く出て、頭にかかなりのブツブツが 😞

(顔もかなり出てきています)

<https://ameblo.jp/princessebiko/entry-12294170362.html>

# アノテーション結果

- がんおよびIPFの患者ブログ記事1,019件にアノテーション
- 292種の薬剤について5,146件のタグ付け



# データ公開

Webで公開中 : [https://yukiar.github.io/pro\\_jp/](https://yukiar.github.io/pro_jp/)



PRO annotations on Weblogs

## PRO annotations on Weblogs

医薬品の開発において、患者報告アウトカム (Patient Reported Outcomes; PRO) と呼ばれる、患者自身による主観的な評価が活用されています。PRO の有力な情報源として、闘病ブログ等、患者自身が作成した Web コンテンツの活用が期待されています。

本データセットは、医薬品の奏功情報を文脈付きでアノテーションした初めてのデータセットです。患者が作成したブログ記事をクロールし、そこで報告されている医薬品の奏功情報についてアノテーションしました。具体的には、奏功情報を記述するスパンを特定し、関連する薬剤名、症状を表す標準コード (ICD-10, MedDRA)、効果の種類を示すラベルをアノテーションしています。

本データセットについての詳細は下記の文献をご参照ください。  
Yuki Arase, Tomoyuki Kajiwara, Chenhui Chu. 2020. Annotation of Adverse Drug Reactions in Patients' Weblogs. in Proceedings of the 12th Edition of its Language Resources and Evaluation Conference (LREC 2020), May 2020.

\* 本データセットを使用した成果を出版する際は、上記の文献を引用してください。

## データセット

URL とアノテーション情報のみからなるメタデータを公開します。URL を用いて記事をクロールし、全文を復元してください。  
[PRO\\_weblog.json](#)