

- ③ A4 15 ページ、形態素数（延べ数）は 2369 語、形態素数（重複除く）844 語であった。詳細は表 9 に示す。

C-3-1-1 形態素頻度解析

対象の 3 文書の形態素頻度解析を行った。その結果

- ① 留意事項文書では「コンパニオン診断薬等」「医薬品」「バイオマーカー」
- ② Q&A 文書では「コンパニオン診断薬等」「医薬品」「関連」
- ③ ガイダンス文書では「医薬品」「コンパニオン診断薬等」「係る」

の順で形態素の出現頻度が多かった。各文書の上位 20 位までの形態素を表 10 に示す。

C-3-1-2 特徴分析

対象の 3 文書における特徴的な形態素の抽出を試みた。

特徴語の抽出には、2 文字以上であり、かつ各文書で 2 回以上の出現頻度がある形態素で算出を行った。

また、今回対象となる文書では、形態素の出現頻度が少ないため、その点を考慮して算出方法には「Yates 補正 χ^2 乗値」を用いた。

結果は表 11 及び図 2 に示す。

C-3-1-3 評判分析

Q&A 文書における機構の「A：回答」部分に対して、TMS の評判分析を行い、以下の 4 つに自動判定した。

- positive：肯定的な回答
- negative：否定的な回答
- positive/negative：肯定的・否定的両方の意味を含む回答
- blank：肯定的・否定的どちらの意味も成さない回答

その判定結果と、機構の「回答」部分との判定一致数（一致率）を表 12 に示す。

C-3-2 実験 2 と結果

実験 2 では、厚労省法令 DB より「質疑応答集」のキーワードで検索した文書の中で、コンパニオン診断薬や医薬品に関連がある「質疑応答集」を平成 25 年制定分と平成 18 年制定分からそれぞれ 6 文書、計 12 文書を抽出し、実験の対象とした。

C-3-2-1 質問形式の自動判定

対象の質疑応答集の 12 文書から、質問文を抽出した。その結果、平成 25 年制定分の 6 文書からは 81 質問文、平成 18 年制定分の 6 文書からは 178 質問文が抽出された。文書によって、質問文数が異なっており、最も少ない文書の質問文数は 8 質問、最も多い文書の質問文数は 69 質問であった。

次に、これらの質問文が「開かれた質問 (Open Question)」であるか、それとも「閉ざされた質問 (Closed Question)」であるかを分類するため、前述の自動判定のツールを用いて、分類を試みた。その結果を表 13 に示す。

表の中の質問形式の「願望」とは、TMS による述語属性で「願望」と判定された質問文であり、「～して頂きたい」といった質問文が該当する。また、「その他」の項目には、質疑応答集自体の記載ミスや 1 つの質問文に複数の質問文がある場合、二者択一の質問文が含まれている。

C-3-2-2 評判分析

前項で判定した質問文の中から、Closed Question の質問文のみを抽出し、TMS による評判分析の判定と、目視で回答文が、「はい」「いいえ」に準じて回答になっているかを確認し、その一致数（率）を比較した。その結果を表 14・表 15 に示す。

D. 考察

D-1 初年度のマイニング結果について

D-1-1 感性分析結果

初年度に用いたマイニングツール「なずき」の感性分析は、主として商業活動に於けるフリーテキストからのナレッジベース構築に用いることを目的としている。具体的にはコールセンターへの顧客からの訴えをテキストに起こし、その内容を分析し、対象商品（サービス）などについて「どのような点」が「良かった」のか（肯定）、「どのような点」が「悪かった」のか（否定）という観点で集計するといった分析である。

フリーテキストから正規的な情報（明確な形の「属性」とその「属性値」）を抽出できれば、大量のテキストデータが発生する大手のコールセンターでは有効に作用するツールとなりうるであろう。しかし、今回、実際に同ツールの分析結果の一例を、原文をたどりながら検証した結果、「Aを評価する」を「Aに良い評価を与える」という意味に取っていた。実際に審査報告書での「Aを評価した」という記述は、「Aの評価を行った」という意味であったので、これは肯定でも否定でもない文として、専門家の評価としては「不適」とされた。このような例が散見された他、疾患名や症候名に対して、同ツールが「悪い」評価と判定する根拠としているきらいがあった。多くの有害事象は、このような疾患名や症候名で示されるので、その記述自体が否定的な記述と判定されたが、仮に有害事象が発生していても、それが同系薬・同効薬に較べて少なければ、同薬剤の安全プロファイルは肯定的な評価を受けるべきである。このような例を鑑みるに、審査報告書に於ける薬剤の

評価ロジックを、現時点の同ツールでの肯定・否定の検出アルゴリズムで十分に検出することは困難であろうと考えられた。

本研究の目的である、承認に至るまでの審査の論理展開の定型化・類型化を試みるためには、「なずき」の感性分類のパターン辞書やルールを審査報告書に特化した形に追加・修正が必要と考えられた。しかし、同システムは、辞書やルールについてユーザーサイドで変更ができないため、そのようなチューニングはかなり難しいと判断された。

D-1-2 単語の重み付けランキングと類似度

TF*IDF法を用いた単語の重み付けを使用して、各分子標的薬の類似度を検討した結果は、概ね効能効果の類似性と相関があるように思われた。このことは、文書内容のサマライズに有効であるという結果を多く持つ同手法が、一定の効果を示したと考えることが出来る。しかし審査報告書では、当該薬品名称がそもそもコード名で呼称されていることが多い他、その薬品の評価を考えるに大変重要な臨床評価も、企業が振り付けた試験名称（これもほとんどがコード名）で記載されている。このことにより、「〇〇試験は～のような試験である」というようなメタ情報がなければ、実際のマイニング結果から、これ以上薬に関する評価情報を抽出することは出来ない。この点も次年度以降、検討が必要である。

D-2 第二年度のマイニング結果について

初年度の研究では、アンケートやブログ・ツイッターの分析といった、マーケティングの分野で定評のあるテキストマイニングツール「なずき」を採用し分析を進めていたが、今回のように領域の異なる行政

文書を対象とした分析には不向きな面があった。

一方で、第二年度に用いた TMS は、辞書の追加登録や、各種パラメータの変更が容易であり、テキストマイニングの対象に合わせて、比較的簡単にツールのカスタマイズが行えると言った特徴を備えており、様々な分析ツールや結果も可視化表示される。さらに、原文検索機能も充実しており、これらの機能は分析を進めていく上で大いに役立った。

D-2-1 第二年度行った用語追加

第二年度の解析では、審査報告書の形態素解析の際に、専門用語に対応するため医学領域の用語を中心に扱っている「MEID 辞書」と「ライフサイエンス辞書」からおおよそ 22 万語を新たに追加登録し、形態素解析を行った。

その解析結果を見ると、追加登録した辞書には存在していない、薬事独特の用語や表現が使われていることが分かった。そのような薬事独特の用語については、対象となる審査報告書から出現頻度を算出し、約 100 語の薬事独特の専門用語を新たに追加登録した。

具体的には、「一次治療」「二次治療」「適合性書面調査」「多変量ロジスティック回帰分析」(出現頻度順)といった用語を追加した。

今回のように、専門性の高い文書を対象とした形態素解析を行う際は、専門用語辞書を追加するだけでは不十分であり、それに加えて、解析対象から、新たに用語を抽出し、より特化した専門用語辞書を作成して解析することで、形態素解析の精度はさらに高くなると思われた。

D-2-2 第二年度の審査報告書内の評価に関する分析

初年度の「なづき」の感性分析同様に、TMS の評判分析を用いて、審査報告書の内容が「肯定的」なのか「否定的」なのかを自動分類する実験を行ったが、専門家の判定と評判分析の判定一致率は低率であった。唯一、「肯定的でも否定的でも無い文」の自動判定と、専門家の判定との一致率が、50%を超えており、審査報告書の中で、あまり重要でない文を削ぎ落とすという点では、この結果は役に立つかも知れない。しかし、全体的な評判分析の結果を見ると、審査報告書を対象に、市販のテキストマイニングツールに付属されている、文判定分析機能をそのまま用いても、精度が高い結果を得られる可能性は低い。

判定分析の今後の課題として、形態素解析に用いる辞書同様、文判定分析に関するアルゴリズムについても、カスタマイズする必要があると思われた。

D-2-3 第二年度行った頻度分析

一方、頻度分析の結果をみると、初年度の研究結果と比較して、ベバシズマブの特徴的な用語が抽出されていた。

これは、解析対象を文書全体ではなく、「機構」を主語とする文でかつ専門家が「肯定的」「否定的」と分類した文に絞ったためとおもわれる。

形態素のランキング(表 8)をみると、特に Negative な形態素において、ベバシズマブに特徴的な用語が抽出されている。具体的には、「高血圧」「消化管穿孔」「塞栓症」といった用語である。これらの用語は、ベバシズマブの添付文書の「警告」欄に含まれている用語である。

第二年度の実験では、ベバシズマブのみを対象として、形態素を抽出したが、解析の対象を増やすことで、審査報告書における形態素の「肯定的」「否定的」の重み付けが、より一般化出来るかも知れない。そして、文判定分析のカスタマイズをする際に、これらの形態素を「肯定的」なのか「否定的」なのかを自動判定する時の評価語に加えることで、判定分析の精度より向上するかもしれない。

D-3 最終年度に行った研究への考察

最終年度は、まず対象文書に、CoDxに関連する行政文書を選定したが、CoDxの分野は比較的新しい分野と言う事もあり、公開されている行政文書数は3文書と限られており、テキストマイニングによる分析も限られた分析となった。そこで、既存の薬事行政文書をCoDxに限定せずに取材し、これらへのマイニング適用を通して、今後のテキストマイニングに資する知見の獲得・収集に努めた。

D-3-1 実験1の形態素頻度分析と特徴分析

実験1ではまず、CoDxに関連する3文書における形態素頻度分析を行った。当然の事ながら3文書において、「コンパニオン診断薬等」という形態素が上位であった。それ以外に関しては、各々の文書に応じた形態素が上位に示されている。特に、ガイドンス文書の形態素を見てみると、上位に「同等性試験」「臨床試験」「分析法バリデーション」といった特徴的な形態素が示されている。

さらに、TMSによる特徴分析により、対象の文書内において、形態素の出現分布を考慮した分析を行うと、頻度分析では上位に上がっていないが、文書の内容と関連が

深いと思われる特徴的な形態素が上位に示され、各々の文書の特徴をより示す形態素が抽出されている。以上のことから、文書の特徴を示す形態素を抽出するには、形態素の頻度による分析だけでなく、形態素の出現分布を考慮した分析が重要となる。

今後、CoDxに関連する文書数を増やすことで、より特徴的な形態素を特徴分析から抽出することができるようになると思われる。

D-3-2 実験1の評判分析

次に、Q&A文書の評判分析を行い、回答文の内容と自動判定の結果が一致するかを確認したが、回答の内容と自動判定の結果を一致させることは出来なかった。この理由としては、今回の回答文がTMSの評判判定のロジックに適合していない事が原因と考えられる。前述の通り、TMSの評判判定は、「評価を与える語」（デフォルトでは約3000種類の用語が登録されている）と「形態素間の係り受け関係」に基づいて判定されるが、今回対象とした回答文には、「形態素間の係り受け関係」が存在しない回答文が多数存在していた。具体的には、「差し支えない」「よい」「可能である」といった表現のみの回答文である。これらの回答文だけでも意味は十分通じるのだが、TMSの判定ロジックには、「この方法で差し支えない」というように、形態素間の係り受け関係がないと、判定することができないため、自動判定が困難となった。

今回のように回答文を自動判定させるには、対象となる行政文書に合わせて、独自の判定ロジックを作成する必要があると思われる。

D-3-3 実験2の質問と回答の検討

実験2では、文書を CoDx に関連する文書以外にも拡大し、質疑応答集として12文書を対象とした。質問文を、TMS で形態素解析し、「Open Question」と「Closed Question」に判定するロジックに当てはめてみたところ、目視での結果と判定ロジックの結果は、高率で一致した。

「判定不能」とされた質問文を見てみると、TMS による形態素解析の結果、「述語属性」が「疑問」と解析されていない例が複数存在している事がわかった。TMS の形態素解析のロジックを再度確認し修正することで、更に判定制度が向上する可能性がある。また、質問形式が「Open Question」か「Closed Question」か、平成25年と平成18年で比較してみたところ、特に差はなく、その割合も約半々程度であった。

次に、「Closed Question」に対応する回答文書に対して、「はい」「いいえ」に準じた回答がなされているかを確認したところ、約30—40%の割合で、「はい」「いいえ」に準じた回答がなされていた。ただ、その回答の多くは「差し支えない」「よい」「可能である」といった回答であった。

ここでも前実験と同様に対象の回答文に対して、評判分析を行ったが、一致率は低値であった。

今回は、厚労省法令 DB に公開されているコンパニオン診断薬に関連する行政文書を中心にテキストマイニングを試みた。対象となる文書が限られていたため、分析には限界があった。一方で、今回のような行政文書には、独特な表現や特有の言い回しが多用されており、市販のテキストマイニングツールをそのまま活用するには限界があることが分かった。

しかしながら、厚労省法令 DB には、毎年多くの行政文書データが蓄積され続けており、そのような文書データを2次利用して活用することは、非常に意義のあることである。現在のテキストマイニングツールをベースに、行政文書に適した辞書の構築やパラメータの調整、新たな判定ロジックの開発を進めることで、より正確で精度の高い行政文書の分析が可能となると思われた。

E. 結論

本研究は、CoDx の承認審査に関する知識ベースの構築に向けて、テキストマイニングの技術を適用した。

初年度は抗癌剤特に分子標的薬の審査報告書を対象にテキストマイニングを試みた。審査報告書の形態素解析やベクトル空間モデル・TF*IDF 法による重み付けによって、各分子標的薬の特徴となりうる単語を抽出した。

第二年度は、従来の辞書に存在しない単語の登録を行い、また審査報告書に於ける評価の分類処理のために、評判分析に対して、判定パターンやルールの追加・修正を行い、その精度を検討したが、一定の性能向上が見いだされた。

最終年度は、審査報告書以外の行政文書として、質疑応答集に対してテキストマイニングを行った。そこで質問形式と回答形式についての分類について自動判別を形態素解析や評判分析の機能を用いて行ったが、高率での自動化は難しかった。

ただ、今回の研究では、係り受けの分析結果を十分に用いていなかったため、この結果を用いることにより、正答率の向上は

期待された。

また、薬事行政文書が持つ、定型的に見えつつ多分に高度な科学技術知識に基づいた記載については、通常のテキストマイニングツールではその特徴を十分に抽出することが難しいことが判明した。

F. 健康危険情報

特になし。

G. 研究発表

1. 学会発表 (論文あり)

1) 小野 大樹, 尾崎 哲夫, 池田 正行, 横井 英人, コンパニオン診断薬の承認審査に関するナレッジベースの構築に向けたテキストマイニング技術活用の検討, 医療情報学, 31(Suppl.), 525-8, 2011

2) 小野 大樹, 池田 正行, 上村 幸司, 長井 美和, 横井 英人, 分子標的薬の承認審査報告書における審査のポイント抽出とその解析の試み, 医療情報学, 32(Suppl.), 276-9, 2012

H. 知的財産権の出願・登録状況

1. 特許取得

なし

2. 実用新案登録

なし

3. その他

なし

参考文献

[1] 鳥山裕司. 医薬関連バイオマーカーの特許出願動向にみる日本の課題. 政策研究ニユース 2008 ; 26 : 22-26.

[2] 厚生労働省法令等データベースサービス.

<http://www.hourei.mhlw.go.jp/hourei/>.

[3] JAPIC 日本の新薬審査報告書DB. http://www.shinsahoukokusho.jp/dar_us/dar/search/usDarSearch.jsp.

[3] Masao Fuketa, Yuki Kadoya, El-Sayed Atlam, et al. A Method of Extracting and Evaluating Good and Bad Reputations for Natural Language Expressions. Information Technology & Decision Making 2005 ; Vol. 4, No. 2 : 177-196.

[4] 青江順一, 結束雅雪. 継続こそ力: 「なぞき」開発物語(特論-4, <特集>イノベーションが生まれたルーツを探る). 品質 2007 ; 37 (3) : 246-251.

[5] 株式会社 NTT データ. なぞきエモーションアナライザ Ver. 1.4 使用説明書.

[6] Salton G, Wong A, Yang C S. A Vector Space Model for Automatic Indexing. CACM 1975 ; 18 : 613-620.

[7] 北研二, 津田和彦, 獅々堀正幹. ベクトル空間モデルに基づく文書検索. 情報検索アルゴリズム. 共立出版, 2002 : 50-64.

[8] 鈴木隆弘, 小野大樹, 横井英人, 井宮淳, 高林克日己. 退院サマリーのテキストマイニングにおけるエントロピー法と $t f \times i d f$ 法の比較. 医療情報学 2005 ; 25(3) : 173-180.

[9] 小野大樹, 高林克日己, 鈴木隆弘, 横井英人, 井宮淳, 里村洋一. テキストマイニングによる退院サマリー自動分類の試み.

医療情報学 2004 ; 24 : 35-44.

[10] 豊田裕貴, 菰田文男. 特許情報のテキストマイニング, ミネルヴァ書房, 2011

[11] 株式会社 NTT データ数理システム. Text Mining Studio バージョン 4.2 技術資料, 2013

一般名	商品名	剤形	効能効果	機序	主な副作用(上位3つ)	重篤な副作用(上位3つ)
セッキシマブ(遺伝子組換え)	アービタックス	注射液	EGFR陽性の治療切替不能な進行・再発の結腸・直腸癌	EGFR	嘔吐(87.2%), 発熱(61.5%), 食欲不振(56.4%)	重度のinfusion reaction, 重度の皮膚症状, 間質性肺炎
リツキシマブ(遺伝子組換え)	リツキサン	注射液	1 CD20陽性のB細胞性非ホジキンリンパ腫 2 インジウム(111m)イリットマブ・チウキセタン(遺伝子組換え)注射液及びイリットリウム(90m)イリットマブチウキセタン(遺伝子組換え)注射液投与の前投与	CD20	発熱(64.3%), 悪寒(54.4%), そう痒(21.7%)	アナフィラキシー様症状, 脳障害, 心障害
ゲムテキマブ	グレベック	錠剤	1 慢性骨髄性白血病 2 KIT(GD17)陽性消化管間質腫瘍 3 PD24フィラフィラ染色体陽性急性リンパ性白血病	Bcr-Abl-TKI	嘔吐(45.7%), 好中球減少症(42.9%), 血小板減少症(40.0%)	骨髄抑制, 出血, 消化管穿孔
ボルネゾミブ	ベルケイト	注射液	再発又は難治性の多発性骨髄腫	プロテアソーム	貧血(79.8%), リンパ球減少(64.7%), 白血球減少	脳障害, 心障害, 末梢性ニューロパシー
ペバシズマブ(遺伝子組換え)	アバステン	注射液	1 治療切替不能な進行・再発の結腸・直腸癌 2 標準治療を除外し治療不能な進行・再発の非小細胞肺癌	VEGF	好中球減少症(18.8%), 白血球減少(18.5%), 高血圧(14.6%)	ショック, アナフィラキシー様症状, 消化管穿孔, 腹膜炎
トラスツズマブ(遺伝子組換え)	ハーセプチン	注射液	1 HER2過剰発現が確認された転移性乳癌 2 HER2過剰発現が確認された乳癌における術後補助化学療法 3 HER2過剰発現が確認された治療切替不能な進行・再発の胃癌	HER2	乳癌:悪寒(4.5%), 頭痛(3.6%), 発熱(3.5%) 胃癌:悪心(63.3%), 好中球減少症(53.4%), 嘔吐(43.9%)	心障害, アナフィラキシー様症状, 間質性肺炎, 脳障害
ソタメニフィシル酸塩	ネクスバル	錠剤	1 根治切替不能又は転移性の腎細胞癌 2 根治不能な肝細胞癌	VEGFR-TKI	カパーゼ上昇(58.6%), 手足麻痺(35.2%), アミラーゼ上昇(40.7%)	手足麻痺, 劇症性高尿酸, 皮膚粘膜炎(Stevens-Johnson症候群)
スニチニブリン酸塩	スーデント	錠剤	1 イマチニブ抵抗性の消化管間質腫瘍 2 根治切替不能又は転移性の腎細胞癌	VEGFR-TKI	血小板減少(91.4%), 白血球減少(85.2%), 皮膚発赤(87.2%)	骨髄抑制, 感染症, 高血圧
ゲムツズマブオプアマイシン(遺伝子組換え)	マイロターグ	注射液	再発又は難治性のCD33陽性の急性骨髄性白血病	CD33	発熱(86.0%), 血小板減少(86.0%)白血球減少(82.5%)	infusion reaction, 重篤な過敏症, 血液障害(骨髄抑制等)
ゲムテキマブ	イレック	錠剤	手術不能又は再発非小細胞肺癌	EGFR-TKI	発熱(62.7%), 下痢(49.0%), そう痒(49.0%)	急性肺障害, 間質性肺炎, 重度の下痢, 脱水

表 1 今回対象とした 10 種類の分子標的薬

なずきによる判定		専門家による評価					
判定結果	判定数	肯定	否定	肯定・否定	不適	不明	合計
肯定	410	39(10%)	15(4%)	12(3%)	344(84%)	0	410
否定	508	55(11%)	65(13%)	225(44%)	160(31%)	3(1%)	508

表 2 「なずき」の感性分析による判定と専門家による評価

順位	アービタックス		アバステン		イレック		グリーブック		スーデント	
	INDEX	実効	INDEX	実効	INDEX	実効	INDEX	実効	INDEX	実効
1	CPT-11	11,080	XELOX	9,972	GSS	7,220	ABL	5,906	SU012662	12,657
2	EGFR阻害	9,847	AVF2107G	9,783	MSS法	6,888	GIST	5,675	不効薬	8,376
3	IMCL	9,008	AVF2192G	8,606	プラチナ系	6,695	BCR	5,059	抗癌薬	7,880
4	CPT-11群	9,003	抗ペニシズマブ抗体	8,494	抗体製剤	6,655	CML患者	5,059	スニチニブ	7,589
5	CPT-11併用群	8,880	FOXFOX4	8,378	黄色製剤	6,524	CML	4,800	GIST	7,428
6	抗HER2薬250MG	8,188	LVF群	7,753	抗HER2	6,524	GIST患者	4,800	検査費SU012648	7,412
7	CTG	8,129	血液検査/シズマブ濃度	7,753	EGFR阻害	6,460	KITチロシンキナーゼ阻害	4,788	PDGFR	7,191
8	抗HER2薬400MG	8,129	AVF2192G	7,673	ACE阻害	6,328	副作用	4,753	副作用	7,117
9	抗ペニシズマブ抗体	8,061	LOT	7,673	700MG	5,906	500MG群	4,828	腎臓病	7,093
10	CA	7,784	国際電気標準会議	7,589	HPLC	5,906	成人GIST患者	4,481	IFN-α群	6,780
11	CPT-11単独群	7,753	LV	7,589	副作用小2度	5,906	既承認申請資料	4,481	スニチニブリン酸塩	6,780
12	抗癌薬IMCL	7,589	国際1018157	7,412	225MG	5,722	効果検証項目	4,481	RET	6,524
13	ILD	7,412	第二群	7,010	経腸剤	5,722	副作用検証項目	4,481	カプセル剤	6,384
14	腫瘍名	7,412	CG	6,959	経腸剤	5,521	400MG群	4,441	GIST患者	6,259
15	CS	7,316	IFL	6,478	緑内障	5,521	中間集計	4,359	1日1回50MG	6,236
順位	ネクスバル		ハーセプチン		ベルケイト		マイロターグ		リツキサン	
	INDEX	実効	INDEX	実効	INDEX	実効	INDEX	実効	INDEX	実効
1	AUCO-12	7,753	転移性乳癌	8,014	JPN101	8,466	HPS7.6	11,931	CD20	9,008
2	手足皮膚反応	7,589	抗HER2	8,378	多発性骨髄腫	8,269	CD33	11,785	非ホジキンリンパ腫	8,885
3	50MG群	7,412	PI3K阻害	7,389	プロテアソーム阻害	8,120	抗CD33抗体	11,172	マクロ球阻害リンパ腫	7,880
4	副作用軽減	9,739	副作用	7,318	ネキトクタン群	8,061	パット	10,489	副作用	7,692
5	1日400MG	8,655	HER2阻害	7,117	末次薬	7,589	NOD	10,175	副作用	7,589
6	200MG群	8,855	AC療法	6,888	MM	6,780	副作用発生	9,662	副作用発生項目	7,230
7	MOTZER	9,384	PTX単独群	6,338	RPM8226	6,780	CMA-676	8,851	ラニー	6,780
8	プロテソ	8,269	IMC法	6,197	血液CD20	6,655	9MG	8,855	SMALL	6,655
9	PI3K阻害	8,238	副作用発生	5,908	プロテアソーム	8,524	産後療法	8,785	375MG	8,524
10	腎臓病	9,162	HER2	5,722	プロテアソーム活性阻害薬	8,524	抗CD33抗体	8,880	抗癌薬IMCL	6,384
11	副作用CD20	8,009	過剰発現	5,979	輸入輸入症例	8,064	抗CD33抗体	8,436	副作用	6,254
12	IL2H分注	8,007	HER2阻害	5,521	ホルテアソム	6,072	CR2阻害	8,376	副作用	6,236
13	本薬400MG	9,977	IMC	5,521	転移性乳癌	6,072	副作用発生	8,120	1日375MG	5,906
14	AUCNORM	5,722	オープン試験	5,521	CHT	5,906	抗CD33	7,907	CHOP療法	5,521
15	コホートあたり	5,722	腫瘍1件	5,521	プロテアソーム阻害薬	5,906	非社会カクアマイシン阻害	7,907	リンパ腫	5,521

表 3 各分子標的薬の単語とその重み順

類似度(COSθ)	アービタックス	アバステン	イレック	グリーブック	スーデント	ネクスバル	ハーセプチン	ベルケイト	マイロターグ	リツキサン
アービタックス	-	0.118	0.061	0.039	0.096	0.084	0.060	0.070	0.061	0.045
アバステン	0.118	-	0.054	0.034	0.086	0.069	0.072	0.074	0.073	0.040
イレック	0.061	0.054	-	0.068	0.078	0.085	0.068	0.058	0.055	0.043
グリーブック	0.039	0.034	0.058	-	0.062	0.058	0.035	0.032	0.032	0.026
スーデント	0.096	0.086	0.078	0.062	-	0.140	0.047	0.080	0.059	0.027
ネクスバル	0.084	0.069	0.065	0.058	0.140	-	0.045	0.077	0.054	0.026
ハーセプチン	0.060	0.072	0.064	0.035	0.047	0.045	-	0.056	0.064	0.101
ベルケイト	0.070	0.074	0.058	0.032	0.080	0.077	0.056	-	0.074	0.040
マイロターグ	0.061	0.073	0.055	0.032	0.059	0.054	0.064	0.074	-	0.059
リツキサン	0.045	0.040	0.043	0.026	0.027	0.026	0.101	0.040	0.059	-

表 4 分子標的薬間の類似度

項目	値	品詞	出現回数
延べ単語数	40081	名詞	31205
総行数	2172	記号	3919
総文数	8960	動詞	2537
単語種別数	7466	接続詞	1486
平均行長(文字数)	63.4	副詞	402
平均文長(文字数)	15.4	形容詞	300
		連体詞	204
		感動詞	8
		分かち書きエラー	7
		接頭詞	6

表 5 ベバシズマブの審査報告書の概要

デフォルト辞書			カスタマイズ辞書		
評価	形態素数	割合	評価	形態素数	割合
正解	1573	88%	正解	1731	97%
誤り	192	11%	誤り	34	2%

表 6 辞書の違いによる正答率の比較

TMSIによる判定		専門家による判定と一致した数(一致率)
結果	判定数	
Positive	49(25%)	16(35%)
Negative	17(8%)	3(17%)
Positive & Negative	13(7%)	0
Blank	110(58%)	60(54%)
合計	189	79(42%)

表 7 評判判定と専門家の判定の一致数 (一致率)

形態素	品詞	品詞詳細	positive		negative		出現割合 ※#1000	形態素	品詞	品詞詳細	positive		negative		出現割合 ※#1000
			出現回数	出現割合	出現回数	出現割合					出現回数	出現割合	出現回数	出現割合	
1 判断	名詞	一般	35	0.021	14	-0.010	11.88	高血圧	名詞	一般	1	14	0.001	-0.010	-8.96
2 有効性	名詞	一般	18	0.010	3	-0.002	7.75	抗べバシズマブ抗体	名詞	一般	0	9	0.000	-0.006	-6.15
3 任用	名詞	一般	18	0.011	5	-0.003	7.61	消える	名詞	一般	29	34	0.018	-0.023	-5.47
4 化学療法	名詞	一般	20	0.012	7	-0.005	7.47	危険因子	名詞	一般	0	8	0.000	-0.005	-5.47
5 二次治療以降	名詞	一般	12	0.007	0	0.000	7.35	高い	名詞	一般	3	9	0.002	-0.006	-4.31
6 了	名詞	サ変接続	18	0.011	6	-0.004	6.93	影響	動詞	自立	1	7	0.001	-0.005	-4.17
7 示す	動詞	自立	16	0.010	5	-0.003	6.39	消化管穿孔	名詞	サ変接続	1	7	0.001	-0.005	-4.17
8 上気支炎	名詞	一般	18	0.011	8	-0.005	5.56	申請者	名詞	一般	28	31	0.017	-0.021	-4.03
9 進行	名詞	サ変接続	9	0.005	0	0.000	5.51	考察	名詞	サ変接続	7	12	0.004	-0.008	-3.91
10 必要	名詞	一般	71	0.044	56	-0.038	5.29	可能性	名詞	一般	5	10	0.003	-0.007	-3.77
11 求める	動詞	自立	18	0.011	9	-0.006	4.88	求める	動詞	自立	4	9	0.002	-0.006	-3.70
12 一次治療例	名詞	一般	10	0.005	2	-0.001	4.76	今後	名詞	一般	6	10	0.004	-0.007	-3.16
13 回答	名詞	サ変接続	14	0.009	6	-0.004	4.48	必要	名詞	サ変接続	15	18	0.009	-0.012	-3.11
14 治療効果不詳	名詞	一般	7	0.004	0	0.000	4.29	治療効果	名詞	一般	4	8	0.002	-0.005	-3.02
15 妥当	名詞	一般	8	0.005	1	-0.001	4.22	炎症症	名詞	一般	3	7	0.002	-0.005	-2.95

表 8 形態素の重み付けランキング (左: 上位 15 位、右: 下位 15 位までそれぞれ抜粋)

項目	値	品詞	出現回数
総行数	46	名詞	720
総文数	62	動詞	65
平均文長(文字数)	45.3	形容詞	1
延べ単語数	868	副詞	15
単語種別数	316	連体詞	11
		接続詞	34
		格助詞	1
		記号	21

項目	値	品詞	出現回数
総行数	43	名詞	420
平均行長(文字数)	41.7	動詞	54
総文数	63	形容詞	12
平均文長(文字数)	28.5	副詞	14
延べ単語数	541	連体詞	3
単語種別数	271	接続詞	19
		記号	19

項目	値	品詞	出現回数
総行数	164	名詞	1872
総文数	223	動詞	186
平均文長(文字数)	38.1	形容詞	31
延べ単語数	2369	副詞	52
単語種別数	844	連体詞	27
		接続詞	121
		記号	80

表9 対象の3文書の基本情報（上：留意事項文書 中：Q&A文書 下：ガイドンス文書）

形態素	品詞	品詞詳細	頻度
医薬品	名詞	一般	20
コンパニオン診断薬等	名詞	一般	19
係る	動詞	自立	13
平成	名詞	固有名詞	13
記載	名詞	サ変接続	10
1	名詞	数	9
2	名詞	数	9
旨	名詞	一般	9
治験	名詞	一般	9
届け出る	名詞	一般	9
医療機器	名詞	一般	8
関連	名詞	サ変接続	8
承認申請	名詞	一般	8
また	名詞	一般	7
行う	動詞	自立	7
取り扱う	名詞	一般	7
対応	名詞	サ変接続	7
必要	名詞	一般	7
開発	名詞	サ変接続	6
開発状況	名詞	一般	6

形態素	品詞	品詞詳細	頻度
コンパニオン診断薬等	名詞	一般	18
医薬品	名詞	一般	10
関連	名詞	サ変接続	7
必要	名詞	一般	7
PMDA	名詞	固有名詞 組織	6
課長通知	名詞	一般	6
記	名詞	一般	5
行う	動詞	自立	5
範囲	名詞	一般	5
1	名詞	数	4
診断薬等	名詞	一般	4
相談	名詞	サ変接続	4
当該医薬品	名詞	一般	4
同時期	名詞	一般	4
望ましい	形容詞	自立	4
目的	名詞	一般	4
用いる	動詞	自立	4
ある	動詞	自立	3
医療機器	名詞	一般	3
可能	名詞	形容動詞 語幹	3

形態素	品詞	品詞詳細	頻度
コンパニオン診断薬等	名詞	一般	44
医薬品	名詞	一般	33
2	名詞	数	25
バイオマーカー	名詞	一般	21
関連	名詞	サ変接続	20
示す	動詞	自立	20
用いる	動詞	自立	20
適切	名詞	形容動詞 語幹	19
1	名詞	数	18
3	名詞	数	18
実施	名詞	サ変接続	18
同等性試験	名詞	一般	18
必要	名詞	一般	15
臨床試験	名詞	一般	15
評価	名詞	サ変接続	14
また	名詞	一般	13
分析法バリデーション	名詞	一般	13
本ガイダンス	名詞	一般	12
検体	名詞	一般	11
試験	名詞	一般	11

表 10 対象の3文書の形態素頻度解析（上位20位まで）
（上：留意事項文書 中：Q&A文書 下：ガイダンス文書）

	留意事項の文書	指標値	Q&Aの文書	指標値	ガイダンスの文書	指標値
1	記載	39.316	治療薬	20.122	バイオマーカー	15.566
2	表示等	37.944	課長通知	18.578	同等性試験	11.417
3	治験	34.414	効能	16.84	示す	9.19
4	届け出る	27.368	同時期	12.908	本ガイダンス	7.884
5	当該コンパニオン診断薬等	26.182	効果	12.143	評価	7.298
6	治験依頼者	23.854	照会	12.143	臨床試験	6.912
7	係る	23.497	ある	11.22	検体	6.713
8	企業	19.603	なす	11.22	分析法バリデーション	6.713
9	名称	19.603	製造販売業者	11.22	試験	5.546
10	医療機器	17.047	添付文書	11.22	バイオマーカー陰性例	4.966
11	課長	16.854	診断薬等	9.069	前向き	4.966
12	平成	16.321	判断	9.069	体外診断薬	4.966
13	申請	16.257	質疑応答集	7.253	臨床的カットオフ値	4.387
14	12	13.375	PMDA	6.579	実施	4.327
15	厚生労働省医薬食品局審査管理課医療機器審査管理室長通知	13.375	やむを得ない	5.771	無作為化比較試験	3.811
16	対応	13.13	可否	5.771	バイオマーカー陽性例	3.238
17	開発状況	10.105	箇所	5.771	困難	3.238
18	供給	9.923	疑義	5.771	精度	3.238
19	治験届出	9.923	使用上	5.771	適切	3.2
20	治験用	9.923	事情	5.771	データ	2.67

表 11 対象の3文書の特徴分析（上位20位まで抜粋）

判定結果	TMS判定	回答一致率
positive	6	0%
negative	0	0%
positive and negative	1	0%
not positive and not negative	2	0%
合計	9	0%

表 12 Q&A 文書の回答部分の自動判定一致率

質問形式		目視による確認					合計	自動判定一致率
		open	closed	願望 (open)	願望(closed)	その他		
自動判定	Open	35	0	0	0	1	36	97%
	Closed	0	38	0	1	0	39	97%
	判定不能	0	0	4	0	2	6	***
合計		35	38	4	1	3	81	***
目視一致率		100%	100%	0%	100%	***		

質問形式		目視による確認					合計	自動判定一致率
		open	Closed	願望 (open)	願望(closed)	その他		
自動判定	Open	55	1	1	0	12	69	80%
	Closed	1	74	0	0	4	79	94%
	判定不能	13	2	11	0	4	30	***
合計		69	77	12	0	20	178	***
目視一致率		80%	96%	8%	***	***		

表 13 上段 平成 2 5 年の質疑応答集 6 文書 下段 平成 1 8 年の質疑応答集 6 文書

- ・質問形式の「願望」とは、TMS による述語属性が「願望」と判定された質問文であり、「～して頂きたい」といった質問が該当する。
- ・「その他」には、質疑応答集の記載ミスや 1 つの質問文に複数の質問文がある場合、二者択一の質問文が含まれる。

Closed Question に対する回答	「はい」「いいえ」に準ずる回答	それ以外	合計
平成 2 5 年質疑応答集	15 (39%)	23 (61%)	38
平成 1 8 年質疑応答集	27 (35%)	50 (65%)	77

表 14 対象の質疑応答集での” Closed Question” に対応する回答文の割合

Closed Question に対する回答	「はい」「いいえ」に 準ずる回答文	評判分析 一致数(率)
平成25年質疑応答集	15	3(8%)
平成18年質疑応答集	27	9(11%)

表15 対象の質疑応答集での「はい」「いいえ」に準ずる回答文に対する評判分析との一致数

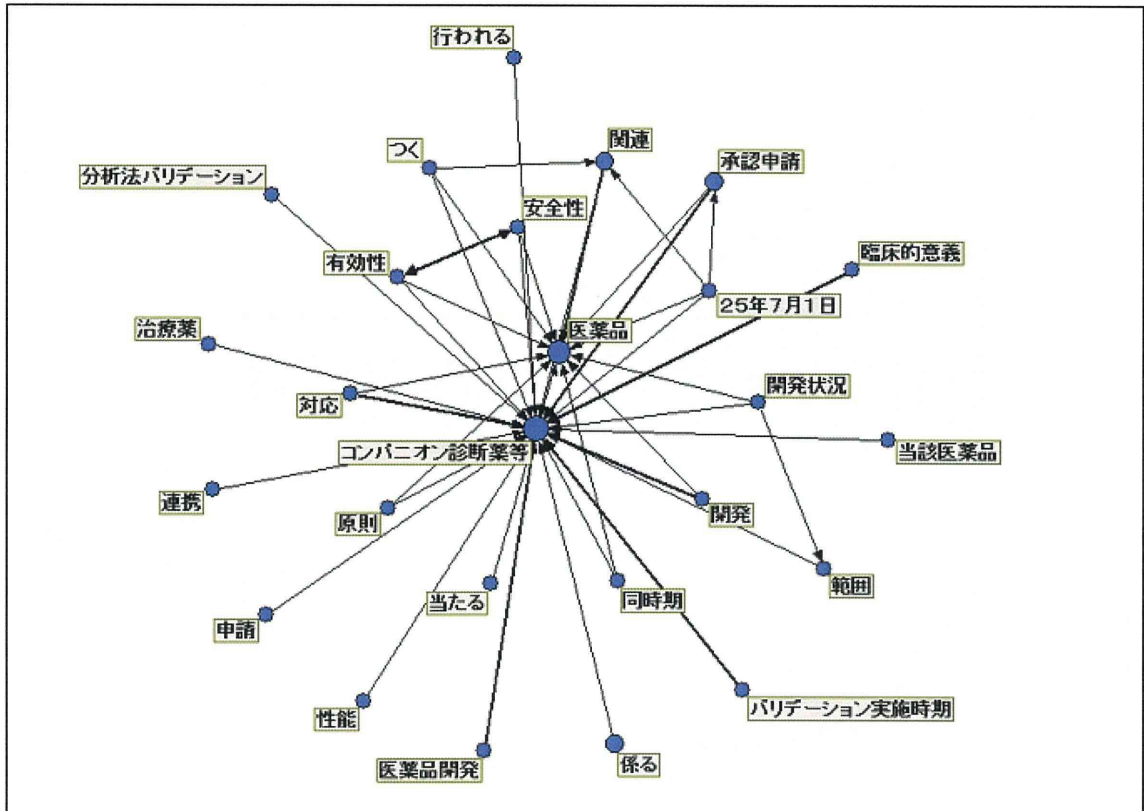


図1 「コンパニオン診断薬等」の形態素に注目したグラフ構造の例

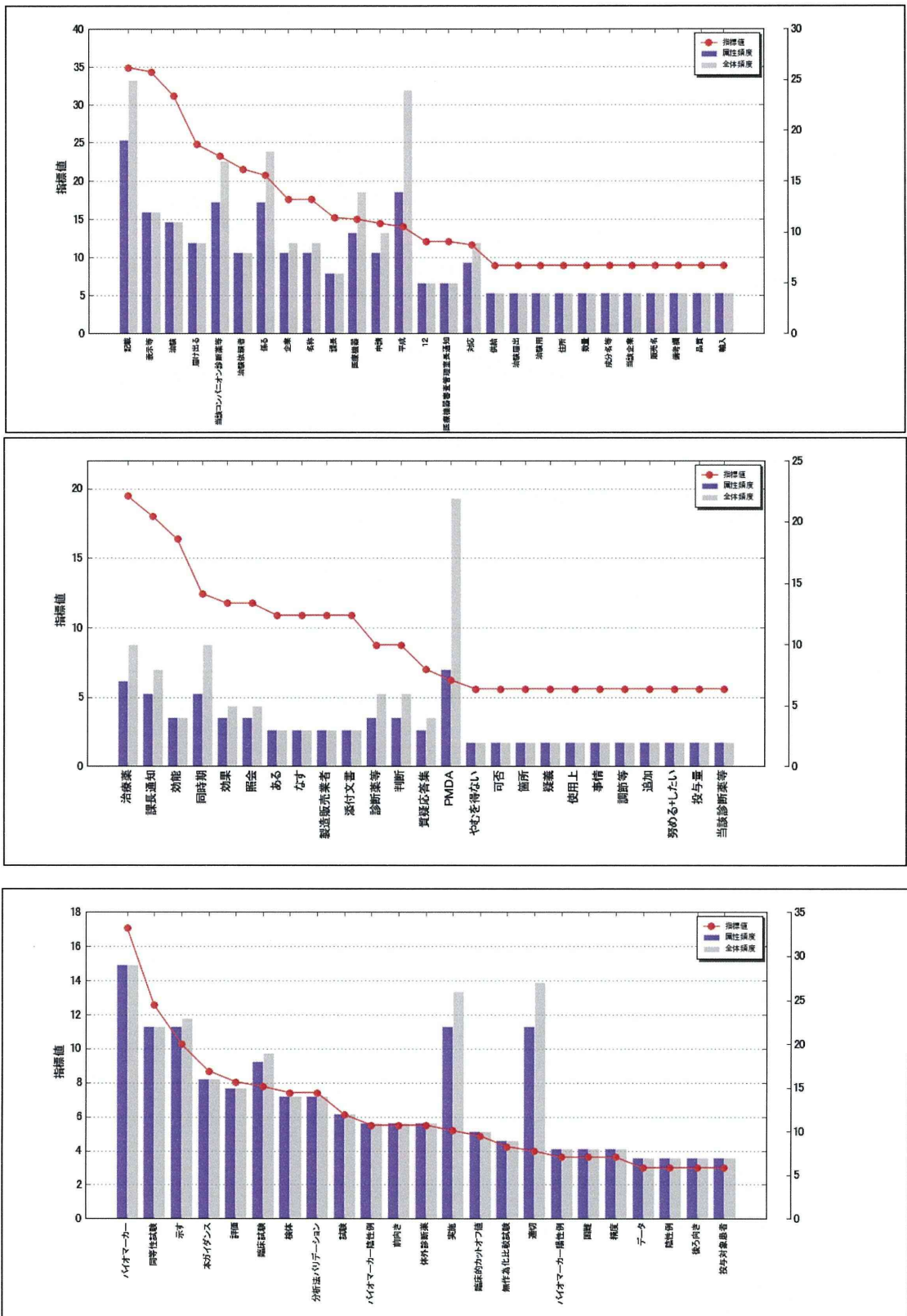


図2 対象の3文書の特徴分析（上位20位まで）
 （上：注意事項文書 中：Q&A 文書 下：ガイダンス文書）

Ⅱ. 研究成果の刊行に関する一覧表

研究成果の刊行に関する一覧表

雑誌

発表者氏名	論文タイトル名	発表誌名	巻号	ページ	出版年
Shimazawa R, Ikeda M	Are there any differences in the regulations of personalized medicine among the USA, EU and Japan?	Br J Clin Pharmacol	75	1365-7	2012
嶋澤るみ子 池田正行	産業界及びFDAスタッフ向けガイダンス草案 体外コンパニオン診断機器ガイダンス草案 (和訳資料)	医薬品医療機器レギュラトリーサイエンス	43	514-9	2012
嶋澤るみ子 池田正行	臨床開発と患者選択に関連したゲノム薬理学バイオマーカーに伴う方法論的問題への考察書 草案 (1)	医薬品医療機器レギュラトリーサイエンス	43	1010-6	2012
嶋澤るみ子 池田正行	臨床開発と患者選択に関連したゲノム薬理学バイオマーカーに伴う方法論的問題への考察書 草案 (2)	医薬品医療機器レギュラトリーサイエンス	43	1097-105	2012
Shimazawa R, Ikeda M	Approval gap of pharmacogenomic biomarkers and in vitro companion diagnostics between the United States and Japan	J Clin Pharm Ther	39	210-4	2014
Shimazawa R, Ikeda M	Differences in pharmacogenomic biomarker information in package inserts from the United States, the United Kingdom, and Japan	J Clin Pharm Ther	38	468-75	2013
Uchida T, Ikeno F, Ikeda K, Suzuki Y, Todaka K, Yokoi H, Thompson G, Krucoff M, Saito S	Global Cardiovascular Device Innovation: Japan-USA Synergies.	Circ J	77	1722-1726	2013
小野 大樹, 尾崎哲夫, 池田 正行, 横井 英人,	コンパニオン診断薬の承認審査に関するナレッジベースの構築に向けたテキストマイニング技術活用の検討	医療情報学	31(Suppl.)	525-8	2011
小野 大樹, 池田正行, 上村 幸司, 長井 美和, 横井 英人	分子標的薬の承認審査報告書における審査のポイント抽出とその解析の試み	医療情報学	32(Suppl.)	276-9	2012

Ⅲ. 研究成果の刊行物・別刷

Letter to the Editors

Are there any differences in the regulations of personalized medicine among the USA, EU and Japan?

Rumiko Shimazawa & Masayuki Ikeda

Graduate School of Biomedical Sciences, Nagasaki University, Nagasaki, Japan

In their cautious review, Shah and Shah [1] emphasized differences in regulations of personalized medicine (PM) among the three major authorities, the US Food and Drug Administration (FDA), the European Medicines Agency (EMA) and the Pharmaceuticals and Medical Devices Agency (PMDA) in Japan. Specific points regarding the differences, however, were not raised for the drugs they selected for discussion in their review. Given that the International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) brings together the regulatory authorities and pharmaceutical industries of the USA, Europe and Japan, scientific and technical aspects of drug registration should be harmonized. To identify differences, if any, in regulations of PM, we investigated approvals of PM drugs in the three regions.

As a typical example of PM, we focused on PM [2] drugs whose pharmacogenomic biomarker is required on the label. We also studied ivacaftor and pertuzumab, which were omitted in the list [2] simply because they were approved after publication of the list. The US, European and Japanese approval data on these drugs were obtained from Drugs@FDA (<http://www.accessdata.fda.gov/Scripts/cder/DrugsatFDA/>), European public assessment reports (http://www.ema.europa.eu/ema/index.jsp?curl=pages/medicines/landing/epar_search.jsp&mid=WC0b01ac058001d125) and the PMDA website (<http://www.info.pmda.go.jp/approvalSrch/PharmacySrchInIt/>), respectively. We defined submission/approval delay as the difference between the date of submission/approval in the USA and that in the EU or in Japan.

Of 17 FDA-approved drugs and 18 indications whose biomarker is labelled as required, 13 drugs and 14 indications were approved in the EU, whereas 12 drugs and 12

indications were approved in Japan (Table 1). The median submission delay from the submission in the USA was 0 months in the EU and 21 months in Japan. The median approval delay from the approval in the USA was 6 months in the EU and 28 months in Japan.

One would expect that labels would not differ significantly among countries, given that regulatory authorities evaluate the same scientific data. Both biological and non-biological factors, however, can affect regulatory decisions. For example, a much lower incidence of cystic fibrosis [3] and melanoma [4] in Japan compared with the West could discourage the makers of ivacaftor and vemurafenib to file an application to the PMDA. Denileukin diftitox and tositumomab, which were approved for lymphoma by the FDA in 1999 and 2003, respectively, remain unavailable in both the EU and Japan, probably because better treatment modalities are available now.

The approval delay in Japan was observed in other therapeutic areas [5]. The present study shows that three-quarters of the approval delay consisted of delays in submission. The approval delay without submission delay in the EU indicates that the reviews took longer for the EMA than for the FDA. The cross-sectional design of our study makes causal inference of these delays difficult.

Our results show some similarities and differences in the approvals of PM drugs among the three regions of the ICH. Further studies are needed to investigate differences in postmarketing regulations of PM drugs, because such regulations are important for risk-benefit assessment of PM and are greatly affected by local factors, such as health policies, culture and financial settings.

Competing Interests

There are no competing interests to declare.