

WG Key Activities

WG1: HBD-Related Clinical Studies – Endeavor Japan and SPIRIT Japan Trials

As described here, the ultimate goal of the HBD initiatives is to lead to more expeditious development and marketing in both countries for those new therapies that have been demonstrated to be reasonably safe and effective. One important step toward accomplishing this objective is the carrying out of global clinical trials enrolling patients worldwide using harmonized study protocols. In line with this HBD concept, the Endeavor Japan trial was initiated in the spring of 2005 for Medtronic's Endeavor® Zotarolimus-Eluting Coronary Stent. This Japanese study for a new DES utilized the identical study protocol that was used in the global Endeavor-II trial,¹⁰ which was an important supporting study for US marketing approval. Following the Endeavor Japan study, a second study, the SPIRIT Japan study, to evaluate the Abbott Vascular XIENCE V Everolimus-Eluting Coronary Stent, was conducted in the USA and Japan simultaneously under a global DES development strategy.

There were several discussion points in executing these clinical trials. One was whether these trials should be separated from a larger global trial of each. It was a dilemma. Future clinical trials might not happen if these trials failed, but if the trials were separated, Japan might lose opportunities to catch up with the advanced countries. Finally, as a learning step, the separated study design with an identical study was chosen.

Experience from these studies then demonstrated that the Japanese cardiology community was in fact ready to participate fully in global DES trials. Ten Japanese sites joined a global DES trial, the PLATINUM trial, and contributed very effectively.¹¹

WG2: INTERMACS and J-MACS

Referring to the INTERMACS program,¹² the J-MACS program¹³ was launched in 2010 for post-marketing follow-up for the 2 left ventricular assist devices approved in Japan in 2010, EVAHEART¹⁴ and DuraHeart.¹⁵ HBD WG2 fully supports this program.

WG3: Infrastructure for Clinical Studies in Japan

Through the clinical studies described here (the WGs 1- and 2-related studies), infrastructure for clinical studies in Japan was further developed. This might lead to more clinical studies from Japan such as the 2 post-market studies in cardiology, the TAXUS Japan Postmarket surveillance study¹⁶ and, the MIRACLE-ICD outcome measured in Japanese indication (MOMIJI) study.¹⁷

WG4: Collaborative Consultation and Review of Pre-Marketing Applications Pilot Program and Research Papers

Led by WG4, a direct consequence of HBD is the Medical Device Collaborative Consultation and Review of Pre-marketing Applications pilot program.^{18,19} This unique program involves the active engagement of the industry sponsor with both US FDA and MHLW/PMDA. A goal of the pilot program is to advance both the speed and quality of clinical/statistical consultations and the regulatory review process for potential earlier market access and improved public health benefit.

WG4 compared GCP between Japan and the USA. The outcome was published as a key accomplishment of WG4.²⁰ The research concluded that there were several administrative differences but no essential differences between the 2 sets of GCP.

HBD Future and Conclusions

The next HBD East Think-Tank Meeting, HBD East 2013 is currently planned for 9 and 10 July in Tokyo, Japan. As a dynamic program, HBD has recognized that it must be able to adapt as necessary to address new therapies and new scientific challenges. In addition to innovations in DES technology, such as the use of biodegradable materials, this meeting will address transcatheter aortic and mitral valve interventions. Moreover, it is envisioned that, over time and with appropriate support from its participating stakeholders, HBD could expand to include other medical devices such as orthopedic products. Finally, as the HBD program matures, other regulatory bodies in other countries could be involved.

Development of innovative medical devices is often a driver for evolution in medicine. Physicians, industry, and regulators all have an important role to play in ensuring that new medical devices provide safe and effective therapy. Physicians are primary contributors to medical device development through their participation in clinical trials. In clinical research activities for medical device development, collaboration among clinicians, the device industry, and regulators is an essential to making innovative therapy available to patients. Unlike GHTF, HBD provides physicians with an open platform on which to collaborate with industry and regulators. Contributions from each stakeholder group will be needed to ensure the future success of HBD, but with a focus on clinical and scientific challenges and new product innovation, HBD will continue to streamline the advance of new medical devices.

Disclaimer

This article represents the personal views of the authors and does not represent official FDA correspondence or guidance or official MHLW/PMDA correspondence or guidance. The HBD program is focused on collaborative efforts and demonstration projects that promote harmonization of clinical trial practices and medical device regulatory approval processes between the USA and Japan.

Acknowledgments

The authors appreciate all contributions to this manuscript as well as HBD WG1's activities from all stakeholders, including regulators of the US FDA.

Disclosures

Conflict of Interest: Gary Thompson is a full-time employee of Abbott Vascular. He receives travel expenses and salary, which are unrelated to research and exceed an annual total of 50,000 yen from the company.

References

1. Rao SV, Califf RM, Kramer JM, Peterson ED, Gross TP, Pepine CJ, et al. Postmarket evaluation of breakthrough technologies. *Am Heart J* 2008; **156**: 201–208.
2. Yusuf S, Hawken S, Ounpuu S, Dans T, Avezum A, Lanas F, et al. Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the INTERHEART study): Case-control study. *Lancet* 2004; **364**: 937–952.
3. Krucoff MW, Brindis RG, Hodgson PK, Mack MJ, Holmes DR Jr. Medical device innovation: Prospective solutions for an ecosystem in crisis: Adding a professional society perspective. *JACC Cardiovasc Interv* 2012; **5**: 790–796.
4. Maisel WH. Innovation at the Food and Drug Administration's Device Center. *JACC Cardiovasc Interv* 2012; **5**: 797–798.
5. Japanese Pharmaceuticals and Medical Device Agency. Incorporated administrative agency: Pharmaceuticals and Medical Devices Agency (PMDA) midterm plan. <http://www.pmda.go.jp/english/about/pdf/e-keikaku.pdf> (accessed January 7, 2013).
6. Japanese Pharmaceuticals and Medical Device Agency. Mid-term plan

- of the pharmaceuticals and medical devices agency. http://www.pmda.go.jp/english/about/pdf/Second_Mid-term_Plan.pdf (accessed January 7, 2013).
7. International Medical Device Regulators Forum. International medical device regulators forum homepage. <http://www.imdrf.org/index.asp> (accessed January 7, 2013).
 8. U.S. Food and Drug Administration. Japan-U.S. "Harmonization By Doing" HBD pilot program initiative. <http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/InternationalInformation/ucm053067.htm> (accessed January 7, 2013).
 9. Japanese Pharmaceuticals and Medical Device Agency. Japan-US synergies in global medical device innovation: Harmonization by Doing (HBD) at TCT. <http://www.pmda.go.jp/hbd/meeting/tct12.html> (accessed January 7, 2013).
 10. Fajadet J, Wijns W, Laarman GJ, Kuck KH, Ormiston J, Munzel T, et al. Randomized, double-blind, multicenter study of the Endeavor zotarolimus-eluting phosphorylcholine-encapsulated stent for treatment of native coronary artery lesions. Clinical and angiographic results of the ENDEAVOR II Trial. *Minerva Cardioangiol* 2007; **55**: 1–18.
 11. Stone GW, Teirstein PS, Meredith IT, Farah B, Dubois CL, Feldman RL, et al. A prospective, randomized, evaluation of a novel everolimus-eluting coronary stent: The PLATINUM (a Prospective, Randomized Multicenter Trial to Assess an Everolimus-Eluting Coronary Stent System [PROMUS Element] for the Treatment of Up to Two De Novo Coronary Artery Lesions) trial. *J Am Coll Cardiol* 2011; **57**: 1700–1708.
 12. Kirklín JK, Naftel DC, Stevenson LW, Kormos RL, Pagani FD, Miller MA, et al. INTERMACS database for durable devices for circulatory support: First annual report. *J Heart Lung Transplant* 2008; **27**: 1065–1072.
 13. Nakatani T, Sase K, Oshiyama H. 745 Initial Report of Japanese Registry for Mechanically Assisted Circulatory Support (J-MACS). *J Heart Lung Transplant* 2012; **31**(Suppl): S255.
 14. Yamazaki K. EVAHEART: Next-generation ventricular assist device. *Nippon Rinsho Jpn J Clin Med* 2007; **65**(Suppl 5): 594–600.
 15. Yoshitake I, El-Banayosy A, Yoda M, Hata M, Sezai A, Niino T, et al. First clinical application of the DuraHeart centrifugal ventricular assist device for a Japanese patient. *Artif Organs* 2009; **33**: 763–766.
 16. Nakamura M, Kotani J, Kozuma K, Uchida T, Iwabuchi M, Muramatsu T, et al. Effectiveness of paclitaxel-eluting stents in complex clinical patients. Insights from the TAXUS Japan Postmarket surveillance study. *Circ J* 2011; **75**: 2573–2580.
 17. Momomura S, Tsutsui H, Sugawara Y, Ito M, Mitsuhashi T, Fukamizu S, et al. Clinical efficacy of cardiac resynchronization therapy with an implantable defibrillator in a Japanese population: Results of the MIRACLE-ICD outcome measured in Japanese indication (MOMIJI) study. *Circ J* 2012; **76**: 1911–1919.
 18. U.S. Food and Drug Administration. Collaborative consultation and review of premarketing applications pilot program. <http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/InternationalInformation/ucm203552.htm> (accessed January 7, 2013).
 19. Japanese Ministry of Health, Labour, and Welfare. Collaborative consultation and review with the U.S. Food and Drug Administration (in Japanese). <http://www.ourei.mhlw.go.jp/ourei/doc/tsuchi/T090623I001.pdf> (accessed January 7, 2013).
 20. Harmonization-by-Doing Working Group 4. Comparing GCP requirements for medical device clinical trials in the US and Japan. *Regulatory Focus* 2010; **15**: 40–44.

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

小野 大樹¹⁾ 尾崎 哲夫²⁾ 池田 正行³⁾ 横井 英人¹⁾

香川大学医学部附属病院 医療情報部¹⁾ 株式会社NTTデータ²⁾
長崎大学 医歯薬学総合研究科 創薬科学³⁾

Text mining for analysis of the evaluation workflows about review reports of companion diagnostics and molecular target agents

Ono Hiroki¹⁾ Ozaki Tetsuo²⁾ Ikeda Masayuki³⁾ Yokoi Hideto¹⁾

Department of Medical Informatics, Kagawa University Hospital¹⁾ NTT DATA CORPORATION²⁾
Department of Pharmaceutical Medicine, Nagasaki University, School of Medicine³⁾

Owing to the recent development of the tailor-made medical treatments, the accumulated numbers of Biomarker patent applications, which is concerned with the pharmaceutical development, is on the increase. This tendency causes higher possibility of rapid growing numbers in both of the exploitation and application for the companion diagnostics and molecular targeted agents. However, the evaluation and the approval standards of companion diagnostics have not been defined. This research intended to construct the approval examination of companion diagnostics on the knowledge base. The research procedures were to perform text mining the review report on the website, moreover, classifying the stereotypes and patterns of the logical structure of the examination until obtaining its acknowledgement. After these surveys, the clarification of the criteria of authorization had been carried out. As an object, the review report of the molecular targeted agents, which was officially learned on the internet, was selected, in order to analyze sensitively and totalize review reports for drugs by Japanese analyzer engine "Nazuki". Based on the data collected through using its analyzing engine, some valuable morphemes were abstracted and these data were rated highly by specialists. On the one hand, the methods of this survey also gathered some finding cases of composed morphemes, which were insufficient for comparison between the efficacy of drugs and the adverse drug reactions. After this research, the extreme accuracy is required from the well-matched tuning to the characteristic of a review report, the analysis object of which is the analyzer engine "Nazuki".

Keywords: companion diagnostics, molecular target agents, review report, text mining

1. はじめに

近年のテーラーメイド医療(個別化医療)の発展により、ヒト疾患の診断を中心とするバイオマーカーの累計特許出願件数は増加傾向にある。図1¹⁾は診断に関連するバイオマーカー特許のうち、日本、米国、欧州の出願人国籍とする特許について累積出願件数を示したものである。1995年1月1日から2005年12月31日までに累積で3,890件であり、圧倒的な差はあるが日本、欧州は米国同様増加傾向にある。

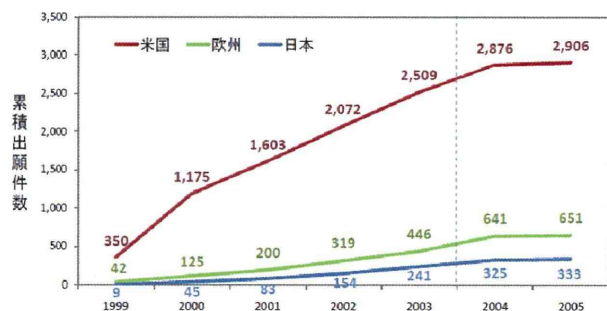


図1 ヒト疾患の診断を中心とするバイオマーカー
累積特許出願数(日・米・欧)

1995年1月1日から2005年12月31日まで累積で3,890件に達する

医薬関連のバイオマーカーの例を挙げると、乳癌に適應をもつトラスツズマブ(商品名:ハーセプチン)の標的分子「HER2」や結腸・直腸癌に適應をもつセツキシマブ(商品名:アービタックス)に関連する「KRAS遺伝子」などがあり、それぞれ患者選択や治療効果を予測するバイオマーカーとなっている。

このような中、今後はバイオマーカーをあらかじめ調べるための体外診断薬(以下、コンパニオン診断薬)とその診断薬が不可欠な分子標的医薬品の開発・申請が急増する可能性が高い。本邦においても、2011年4月成人T細胞白血病リンパ腫の治療薬である抗CCR4抗体とその標的であるCCR4を測定するコンパニオン診断薬が同時申請されている。

しかしながら、そのようなコンパニオン診断薬の位置づけや承認審査の基準は未だ明確ではない。

2. 目的

本研究では、コンパニオン診断薬の承認審査に関する知識ベースの構築を目指している。

今回は、コンパニオン診断薬を検討する前に、その対象となる抗癌剤を検討する必要があると考え、パイロットスタディとして抗癌剤の中でも分子標的薬の承認審査報告書を対象にテキストマイニングを行う。

3. 対象と方法

3.1 対象

2-E-2-3 一般口演/2-E-2:一般口演21

対象には、効能効果に偏りが少ない分子標的薬10種類を選定し、インターネット上に公開されている新薬承認審査報告書データベース²⁾から、各々の審査報告

書を抽出した。表1に抽出した分子標的薬の特徴を示す。

表1 今回対象とした10種類の分子標的薬

一般名	商品名	剤形	効能効果	標的	主な副作用(上位8つ)	重大な副作用(上位3つ)
セツキシマブ(遺伝子組換え)	アービタックス	注射液	EGFR陽性の治癒切除不能な進行・再発の結腸・直腸癌	EGFR	嘔吐(67.2%)、発疹(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(111In)イブリツモマブ チウキセタン(遺伝子組換え)注射液及びイトリウム(90Y)イブリツモマブチウキセタン(遺伝子組換え)注射液投与の前投与	CD20	発熱(64.3%)、悪寒(34.4%)、そう痒(21.7%)	アナフィラキシー様症状、肺障害、心障害
メシル酸イマチニブ	グレベック	錠剤	1. 慢性骨髄性白血病 2. KIT (CD117)陽性消化管間質腫瘍 3. F24フィラデルフィア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(42.8%)、血小板減少症(40.0%)	骨髄抑制、出血、消化管穿孔
ボルトゾミブ	ベルグイド	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(73.5%)、リンパ球減少(64.7%)、白血球減少	肺障害、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. 治癒切除不能な進行・再発の結腸・直腸癌 2. 扁平上皮癌を除く切除不能な進行・再発の非小細胞肺癌	VEGF	好中球減少症(16.8%)、白血球減少(16.5%)、高血圧(14.8%)	ショック、アナフィラキシー様症状、消化管穿孔、瘻孔
トラスツマブ(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癒切除不能な進行・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(43.8%)	心障害、アナフィラキシー様症状、間質性肺炎、肺障害
ソラフェニブシロニド	ネクサバル	錠剤	1. 根治切除不能又は転移性の腎細胞癌 2. 切除不能な肝細胞癌	VEGFR-TKI	リバーゼ上昇(58.6%)、手足症候群(55.2%)、アミラーゼ上昇(40.7%)	手足症候群、剥脱性皮膚炎、皮膚粘膜眼症候群(Stevens-Johnson症候群)
スニチニブリンゴ酸塩	スーテント	錠剤	1. イマチニブ抵抗性の消化管間質腫瘍 2. 根治切除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(82.7%)	骨髄抑制、感染症、高血圧
ゲムツマブオゴマイン(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(85.0%)、血小板減少(95.0%)白血球減少(92.5%)	infusion reaction、重篤な過敏症、血液障害(骨髄抑制等)
ゲフィチニブ	イレック	錠剤	手術不能又は再発非小細胞肺癌	EGFR-TKI	発疹(62.7%)、下痢(49.0%)、そう痒症(49.0%)	急性肺障害、間質性肺炎、重度の下痢、脱水

効能効果、主な副作用、重大な副作用は添付文書を参照した

3.2 方法

3.2.1 形態素解析ツール

形態素解析とは、文書を構成する文字列を単語に分割し、各単語に品詞や語形変化などの情報を与える処理のことである。近年、数多くの形態素解析ツールが出てきているが今回の研究では、審査報告書の形態素解析だけでなく、その内容の論理展開までを解析対象としているため、形態素解析ツールには、インターネット広告やマーケティングの領域において感性分析⁴⁾の手法で定評のある、日本語解析エンジン「なずき」³⁾(以下、「なずき」)を採用した。

3.2.2 感性分析

感性分析^{4,5)}は、主にマーケティングの分野において、ある商品に対する印象や評価が書かれているアンケートやブログ・ツイッターの分析などに使われている。

従来の手法では、意図を理解するには結局人間が文章を読む必要があったが、今回採用した「なずき」の感性分析によると、単語を照合するのではなく、人間の感性を81種類に分類し、それぞれの感性ごとに実際にはどのような文章表現が用いられるの

かというパターン辞書を使って意見を見つけ出すことが可能⁵⁾となる。具体例を示すと、「医薬品の効果が確認できたので良かった」は「医薬品に対する満足」、「医薬品の効果が確認できたら良かったのに」は「医薬品に対する要望」として検出することができる。

3.2.3 ベクトル空間モデル

ここでは、ベクトル空間モデル^{6,7)}について述べる。

今回対象とする審査報告書の集合をDとし、各審査報告書をd1, d2, ..., dj, ..., d10とおく。また、集合Dから「なずき」によって抽出した32,390個の単語をw1, w2, ..., wi, ..., w32390とする。次に、ある審査報告書djに現れる単語wiに対する重みを α_{ij} とおく。このときdjを次のようなベクトル

$$d_j = [\alpha_{1j} \ \alpha_{2j} \ \dots \ \alpha_{ij} \ \dots \ \alpha_{32390j}]$$

で表現し、これをある分子標的薬jの審査報告書のベクトルとする。

また対象の審査報告書の集合全体は、次のような32390×10行列D

$$D = [d_1 \ d_2 \ \dots \ d_{10}] = \begin{bmatrix} \alpha_{11} & \dots & \alpha_{110} \\ \vdots & \ddots & \vdots \\ \alpha_{323901} & \dots & \alpha_{3239010} \end{bmatrix}$$

で表現する事ができる。なお、索引語の重み付けの計算には、先行研究^{8,9)}でも採用されているTF*IDF法を用いることとし、重み α は

$$\alpha_{ij} = l_{ij} g_i$$

で表し、 l_{ij} と g_i は以下のように定義した。

$$l_{ij} = \log(1 + f_{ij})$$

$$g_i = \log\left(\frac{n}{n_i}\right)$$

ここで、 f_{ij} とは索引語 w_i の審査報告書 d_j における出現頻度である。また、 n は対象とする審査報告書の種類であり、 n_i は索引語 w_i が出現した審査報告書の数を示している。

また、審査報告書のベクトル間の類似度は、以下の式を用いてコサイン類似度で表現する。

$$\cos \theta = \frac{d_j \cdot d_k}{\|d_j\| \|d_k\|} = \frac{\sum_{i=1}^{32390} d_{ji} d_{ki}}{\sqrt{\sum_{i=1}^{32390} d_{ji}^2} \cdot \sqrt{\sum_{i=1}^{32390} d_{ki}^2}}$$

4. 結果

4.1 感性分析結果

まず、セツキシマブ(商品名:アービタックス)の審査報告書に対して行った「なずき」による感性分析の判定結果とその結果に対する専門家による評価を表2に示す。

ここで、専門家による評価の「肯定」は、セツキシマブに対する肯定的な評価を示し、「否定」はセツキシマブに対する否定的な評価を示している。次に、「肯定・否定」はセツキシマブに対して肯定的とも否定的とも言えない評価、「不適」はセツキシマブの評価に直接関係のない記述、「不明」はセツキシマブの評価に直接関係あるか不明な記述であったものをそれぞれ集計して示している。

表2 「なずき」の感性分析による判定と専門家による評価

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

結果を見ると、「なずき」による判定と専門家の評価が一致していたものは、肯定で39件(10%)、否定で55件(11%)となった。一方で、「肯定・否定」や「不適」と評価されたものが半数以上あり、その他にも「なずき」の判定では「肯定的」であっても、専門家の評価が「否定的」であったものや、その逆の事例も散見された。

今後、本研究の大目的である、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・修正をしていく必要性があると思われた。

4.2 単語の重み付けランキングと類似度

次に、各分子標的薬と関連が深いと思われる単語と分子標的薬間の類似度を示す。

なずきによる形態素解析後に、前述のTF*IDF法を用いて重み付けし、各分子標的薬の審査報告書のベクトルを算出し、それぞれの形態素を重み順に並べ替えることで、関連が強いと思われる単語を抽出した。表3に各分子標的薬毎の上位15位までの単語を示す。さらに、各分子標的薬同士の類似度を表4に示す。その結果を見ると、10種の分子標的薬間の中で一番類似度の数値が大きかったのは、「スーテントとネクサバル」の(0.14)であり、順に「アービタックスとアバスチン」の(0.118)、「ハーセプチンとリツキサン」の(0.101)と続いていた。

「スーテントとネクサバル」は共に、効能効果に「根治切除不能又は転移性の腎細胞癌」の適応を有しており、標的も「VEGFR-TKI」と同一であり、薬剤のプロファイルからも類似性が高い。次いで「アービタックスとアバスチン」においては、アービタックスは「EGFR陽性の治癒切除不能な進行・再発の結腸・直腸癌」、アバスチンは「治癒切除不能な進行・再発の結腸・直腸癌」の適応を有しており、癌の部位が結腸・直腸であることから、同様に関連性が高い薬剤であった。

表3 各分子標的薬の単語とその重み順

順位	アービタックス		アバステン		イレッサ		グリベック		スーデント	
	INDEX	重み	INDEX	重み	INDEX	重み	INDEX	重み	INDEX	重み
1	CPT-11	11.080	XELOX	9.872	CSS	7.220	ABL	5.906	SU012662	12.057
2	EGFR陽性	9.847	AVF2107G	9.783	MS法	6.898	GIST	5.675	不妊容	8.376
3	IMCL	9.098	AVF2192G	8.606	プラチナ系	6.655	BCR	5.059	癌薬物	7.980
4	CPT-11群	9.053	抗VEGシズマ抗体	8.494	褐色製剤	6.655	CML患者	5.059	スニチニブ	7.589
5	CPT-11併用群	8.660	FOLFOLX4	8.376	黄色製剤	6.524	CML	4.900	GIST	7.428
6	維持投与量250MG	8.136	LV群	7.753	褐色錠	6.524	GIST患者	4.900	報告書SU011248	7.412
7	CTG	8.120	血清中VEGシズマ濃度	7.753	500MG群	6.420	KITチロシンキナーゼ活性	4.788	PDGFR	7.131
8	初回投与量400MG	8.120	AVF0780G	7.673	AGP濃度	6.236	以下同様	4.788	リンゴ酸塩	7.117
9	抗セシジマ抗体	8.051	LOT	7.673	700MG	5.906	600MG群	4.528	腎細胞癌	7.053
10	CA	7.784	国際電気標準会議	7.589	HPLC	5.906	欧米人GIST患者	4.481	IFN-A群	6.780
11	CPT-11単独群	7.753	LV	7.580	幾何最小二乗	5.906	既承認申請資料	4.481	スニチニブリンゴ酸塩	6.780
12	試験番号IMCL	7.589	国内JO18157	7.412	225MG	5.722	効果持続期間	4.481	RET	6.524
13	ILD	7.412	第二群	7.010	国際共同	5.722	国内外臨床試験	4.481	カプセル剤	6.384
14	事象名	7.412	DG	6.655	無酸症	5.521	400MG群	4.441	GIST患者	6.359
15	CS	7.318	IFL	6.479	雄イヌ	5.521	中間集計	4.358	1日1回50MG	6.236
順位	ネクサバル		ハーセプチン		ベルケイド		マイロターグ		リツキサン	
	INDEX	重み	INDEX	重み	INDEX	重み	INDEX	重み	INDEX	重み
1	AUC0-12	7.753	転移性乳癌	8.914	JPN101	9.466	HP67.6	11.931	CD20	9.008
2	手足皮膚反応	7.589	維持量	8.376	多発性骨髄腫	9.269	CD33	11.785	非ホジキンリンパ腫	8.865
3	50MG錠	7.412	FSH法	7.589	プロテアソーム活性	8.120	カリクアマイシン誘導体	11.172	mantle細胞リンパ腫	7.980
4	相併用試験	6.780	負荷量	7.318	テキサメタン群	8.051	バート	10.486	癌腫性度	7.907
5	1回400MG	6.655	HERCEP	7.117	未公表	7.589	VOD	10.175	腫瘍性	7.589
6	200MG錠	6.655	AC療法	6.898	MM	6.780	殺細胞活性	9.682	癌性巨細胞性	7.220
7	MOTZER	6.384	PTX単独群	6.236	RPM18226	6.780	CMA-676	8.961	ラジ	6.780
8	プロセス	6.296	IHC法	6.197	血液中20	6.655	9MG	8.865	SMALL	6.655
9	リスク分類	6.236	初回負荷量	5.906	プロテアソーム	6.524	寛解後療法	8.765	375MG	6.524
10	腎細胞癌	6.162	HT	5.722	プロテアソーム活性阻害率	6.524	カリクアマイシン	8.660	試験番号IDEC	6.384
11	発症期間20	6.090	過剰発現	5.670	個人輸入症例	6.384	総カリクアマイシン誘導体	8.436	巨細胞性	6.264
12	PUGH分類	6.077	NEU遺伝子	5.521	ボルテゾミド	6.077	CRP例	8.376	巨細胞性	6.236
13	本薬400MG	5.977	RIII	5.521	起立性低血圧	6.077	寛解導入療法	8.120	1回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	重度1件	5.521	プロテアソーム阻害活性	5.906	非結合カリクアマイシン誘導体	7.907	リンパ腫	5.521

表4 分子標的薬間の類似度

類似度(COS θ)	アービタックス	アバステン	イレッサ	グリベック	スーデント	ネクサバル	ハーセプチン	ベルケイド	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	0.089	0.072	0.074	0.073	0.040
イレッサ	0.061	0.054	-	0.058	0.076	0.085	0.064	0.058	0.055	0.043
グリベック	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	0.076	0.062	-	0.140	0.047	0.080	0.059	0.027
ネクサバル	0.084	0.089	0.085	0.058	0.140	-	0.046	0.077	0.054	0.026
ハーセプチン	0.060	0.072	0.064	0.035	0.047	0.046	-	0.056	0.064	0.101
ベルケイド	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	-

5. まとめ

本研究では、コンパニオン診断薬の承認審査に関する知識ベースの構築に向けて、まずは、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できた。今後の課題は、まず「なずき」を審査報告書向けにチューニングすることである。具体的には、

- ・ 同義語の正規語への統一化の処理
- ・ 「なずき」の辞書に存在しない単語の登録
- ・ 感性分析に用いる判定パターンやルールの追加・修正

を中心に、進めていく必要があると思われる。

謝辞

本研究は厚生労働科学研究費補助金(医薬品・医療機器等レギュラトリーサイエンス総合研究事

業)の助成を受けたものである。

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. 政策研ニュース 2008; 26: 22-26.
- [2] JAPIC日本の新薬審査報告書DB. http://www.shinsahoukokusho.jp/dar_us/dar/search/usDarSearch.jsp.
- [3] 青江 順一, 結束 雅雪. 継続こそ力: 「なずき」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.
- [4] 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.
- [5] 株式会社NTTデータ. なずき 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tf*idf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.

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

小野 大樹^{*1} 池田 正行^{*3} 上村 幸司^{*2} 長井 美和^{*1} 横井 英人^{*2}

^{*1}香川大学 大学院 医学系研究科 ^{*2}香川大学医学部附属病院 医療情報部

^{*3}長崎大学 医歯薬学総合研究科 創薬科学

Trial of review points extraction from review reports about molecular target agents

Ono Hiroki^{*1} Ikeda Masayuki^{*3} Uemura Koji^{*2} Nagai Miwa^{*1} Yokoi Hideto^{*2}

^{*1}Graduate School of Medicine, Kagawa University

^{*2}Department of Medical Informatics, Kagawa University Hospital

^{*3}Graduate School of Biomedical Sciences, Nagasaki University

Owing to the recent development of the tailor-made medical treatments, the accumulated numbers of Biomarker patent application, is on the increase. This tendency causes the highly possibility of rapid increase in both of exploitation and the application for the companion diagnostic drugs and molecular-targeted agents. However, the evaluation and the approval standards of companion diagnostic drugs have not defined.

In this study, in order to clarify the approval standards of companion diagnostic drugs, the reports basis of approvals of molecular-targeted drugs, relevant to the companion diagnostic drugs, has been text mining. Moreover, the extraction of screening points (decision points) and the analysis of their contents had been also experimented. The subject as the review report of 10 types of molecular-targeted agents, which has been published on the internet, were analyzed by the text mining tool called "Text Mining Studio 4.01". The MEID dictionary and Life Science dictionary was added to the dictionary used in the morphological analysis, furthermore, some technical words that may need were also included for tuning the dictionary. In order to extract the points, firstly Pharmaceuticals and Medical Device Agency (PMDA) extracted the sentences as subjects, in addition, narrowed the decision points from its syntax analysis. The percentage of correct answers of morpheme was improved from 88% to 97%, comparing the results of morphological analysis, before and after dictionary tuning, with the 50 sentences extracted randomly from the review report of Avastin. Secondly, PMDA parsed the 189 subject sentences and distinguished them between Positive and Negative sentences. It achieved the following results such as 49 sentences were "Positive", 17 sentences were "Negative", 13 sentences were "Both Positive and negative" and 110 sentences were "Neither Positive nor negative". Although some medical experts had evaluated about their medical accuracy, there were multiple cases that their evaluations did not match the results. Therefore, it has been brought from this study that customizing of evaluation for syntax analysis and judgment are essential.

Keywords: molecular target agents, companion diagnostics, review report, text mining

1. はじめに

近年の分子生物学の発展により、疾患に関わるシグナルや遺伝子および遺伝子産物が徐々に解明され、ある疾患に特異的な分子を標的とした「分子標的薬」が数多く開発されている。

さらに、分子標的薬のターゲットとなるバイオマーカーの発現の有無を遺伝子レベルで検査する「コンパニオン診断薬(以下 CoDx)」の開発についても、分子標的薬と同時に行われるようになってきている。

本邦では、2012年3月に抗悪性腫瘍薬のモガムリズマブ(商品名:ポテリジオ点滴静注20mg)が「再発又は難治性のCCR4陽性の成人T細胞白血病リンパ腫」の適応を受け承認され、そのCoDx(商品名:ポテリジオテスト)もほぼ同時期に承認された。本事例は、同時承認としては国内初の事例である。

一方で、規制当局の動きも進みつつあり、2011年には、欧州・米国においてCoDxを伴う医薬品の開発に関するガイドラインやガイダンスが出されている¹⁻³⁾。本

邦でも、2012年4月よりPMDA(以下、機構)のコンパニオン診断薬プロジェクト⁴⁾でガイドラインの検討がはじまっている。

このような動きの中で今後は、分子標的薬のような医薬品とCoDxの同時開発・申請が急増する可能性が高く、規制当局の承認審査体制やその基準整備についても早急に進めていく必要がある。

2. 目的

本研究では、今後のCoDxに関連する承認審査の基準整備の一助とするため、CoDxと関連の深い分子標的薬の審査報告書の審査のポイントを、テキストマイニングツールを用いて抽出及び解析できるかを試みた。

3. 対象と方法

3.1 対象

対象には、効能効果に偏りが少ない分子標的薬10種類の審査報告書(表1)を選定した。

表1 対象とした10種類の分子標的薬

一般名	商品名	剤形	効能効果	標的	主な副作用(上位3つ)	重大な副作用(上位3つ)
セツキシマブ(遺伝子組換え)	アービタックス	注射液	EGFR陽性の治療不能な進行 再発の結腸・直腸癌	EGFR	嘔気(87.2%)、発疹(61.5%)、食欲不振(56.4%)	重度のinfusion reaction、重度の皮膚症状、間質性肺炎発症
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. インシウム(111In)イブリツモマブ チウキセタン(遺伝子組換え)注射液及びイットリウム(90Y)イブリツモマブチウキセタン(遺伝子組換え)注射液投与の前投与	CD20	発熱(64.3%)、悪寒(34.4%)、そう痒(21.7%)	アナフィラキシー様症状、肺障害、心障害
メシル酸イマチニブ	グリベック	錠口	1. 慢性骨髄性白血病 2. KIT (D0117) 陽性消化管間質腫瘍 3. F24フィラデルフィア染色体陽性急性リンパ性白血病	Bcr-Abl-TKI	嘔気(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
ボルテゾミブ	ベルゲイド	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(73.5%)、リンパ球減少(64.7%)、白血球数減少	肺障害、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. 治療不能な進行 再発の結腸・直腸癌 2. 扁平上皮癌を除く切除不能な進行 再発の非小細胞肺癌	VEGF	好中球減少症(18.8%)、白血球数減少(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、瘻孔
トラスツマブ(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、肺障害
ソラフェニブシル酸塩	ネクサバル	錠口	1. 根治不能又は転移性の腎細胞癌 2. 切除不能な肝細胞癌	VEGFR-TKI	リバーゼ上昇(58.6%)、手足症候群(55.2%)、アミラーゼ上昇(40.7%)	手足症候群、剥脱性皮膚炎、皮膚粘膜眼症候群(Stevens-Johnson症候群)
スニチニブリンゴ酸塩	スーテント	錠口	1. イマチニブ抵抗性の消化管間質腫瘍 2. 根治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%)、皮膚変色(82.7%)	骨髄抑制、感染症、高血圧
ゲムズマブオノガマイシン(遺伝子組換え)	マイロターゲ	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(95.0%)、血小板減少(95.0%)白血球減少(92.5%)	infusion reaction、重篤な過敏症、血液障害(骨髄抑制等)
グフィチニブ	イレッサ	錠口	手術不能又は再発非小細胞肺癌	EGFR-TKI	発疹(62.7%)、下痢(49.0%)、そう痒症(49.0%)	急性肺障害、間質性肺炎、重度の下痢、脱水

審査報告書とは、医薬品開発の経緯から規格・安定性試験、非臨床試験、臨床試験、審査の概要の順で記載され、医薬品を申請する企業と機構との審議の経過とその結果が記載された行政文書である。審査報告書を読み解く事で、審査時にポイントとなった内容が分かる。

文書ボリュームとしては、A4サイズで数十ページにまとめられているものから数百数十ページにまとめられているものまでである。なお、審査報告書は、承認された医薬品のみ公開され、インターネット上からもダウンロード可能⁵⁾である。

3.2 方法

3.2.1 テキストマイニングツール

テキストマイニングツールには、特許情報のテキストマイニング⁶⁾に用いられた実績のある“Text Mining Studio 4.01 (以下TMS)”を使用した。本ツールの特徴としては

- ・ 辞書のカスタマイズが容易
- ・ 各種パラメータの変更が可能
- ・ 豊富な分析機能
- ・ 原文参照機能
- ・ などが挙げられる。

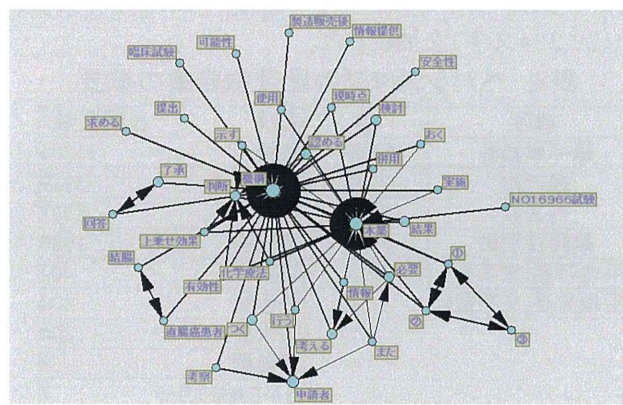


図1 TMSの分析機能の例

文書の形態素の共起関係をグラフ構造で表現している

3.2.2 形態素解析と辞書のカスタマイズ

形態素解析には、TMSに搭載されている「日本語解析エンジンTMStudio3.2」を用いた。

形態素解析に用いる辞書には、TMS既存の辞書に加えて、専門用語として「MEID辞書(電子化医学辞書)」と「ライフサイエンス辞書」を合わせた約22万語(重複した単語を除く)と、さらに審査報告書特有の用語として約100語を追加の用語として登録し、辞書のカスタマイズを行った。

3.3 審査ポイントの抽出と解析

前述のとおり、審査報告書とは、機構が行った審査の内容が記載された文書である。

その内容の多くは、各項目の申請資料に於いて、機構が当該医薬品の性質について情報や考察が不十分であると考えた点を、申請者に照会し、その回答を含めて、当該医薬品の性質が、日本国内で承認可能なものかを判断するというものである。つまり、審査報告書の審査のポイントは「機構の判断が述べられている文章(行)集合」と捉える事ができる。よって、審査報告書から「機構」が主語となる行の内容を検討することで、機構が当該医薬品に関して行った判断の経過を

知ることができる。

3.3.1 評判分析と頻度分析

評判分析とは、TMSの分析機能の1つである。予め定められた評価を与える語(デフォルトでは2200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、行の中から「肯定的:Positive」・「否定的:Negative」と解釈できる表現を抽出する機能である。この機能により「機構」が主語となる行の内容が、「肯定的:Positive」か「否定的:Negative」かを自動的に分類することができる。

さらに、その「肯定的:Positive」「否定的:Negative」といった分類毎に特異的な形態素を抽出するために、形態素の頻度分析を行い、その出現頻度や出現割合から、形態素毎に重み付けを行った。

4. 実験と結果

実験には、平成19年4月に承認されたベバシズマブ(商品名:アバスタチン)の審査報告書を用いた。文書のボリュームは、A4サイズで174ページ、文字数は201,205文字であった。詳細を表2に示す。次に、審査のポイントを抽出するために「機構」が主語となる189行(29,204文字)を抽出した。

表2 ベバシズマブの審査報告書の概要

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
総文字数	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

4.1 辞書による形態素解析精度の比較

ここでは、「TMSの既存の辞書(デフォルト辞書)」と「カスタマイズした辞書」を追加した場合による形態素解析の精度の結果を比較するため、「機構」が主語と

表5 形態素の重み付けランキング

順位	形態素	品詞	品詞詳細	positive		negative		出現割合 差*1000	形態素	品詞	品詞詳細	positive		negative		出現割合 差*1000
				出現頻度	出現割合	出現頻度	出現割合					出現頻度	出現割合	出現頻度	出現割合	
1	判断	名詞	一般	35	0.021	-0.010	11.88	悪血圧	名詞	一般		1	14	0.001	-0.010	-8.96
2	有効性	名詞	一般	16	0.010	-0.002	7.75	抗ベバシズマブ抗体	名詞	一般		0	9	0.000	-0.006	-6.15
3	副作用	名詞	一般	18	0.011	-0.003	7.61	毒える	名詞	一般		29	34	0.018	-0.023	-5.47
4	化学療法	名詞	一般	20	0.012	-0.003	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	0.000	7.35	高い	名詞	一般		3	9	0.002	-0.006	-4.31
6	了済	名詞	サ変接続	18	0.011	-0.004	6.93	影響	動詞	自立		1	7	0.001	-0.005	-4.17
7	未済	動詞	自立	16	0.010	-0.003	6.39	消化管穿孔	名詞	サ変接続		1	7	0.001	-0.005	-4.17
8	上乗せ効果	名詞	一般	18	0.011	-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	0.000	5.51	考察	名詞	サ変接続		7	12	0.004	-0.008	-3.91
10	不変	名詞	一般	71	0.044	-0.038	5.23	可能性	名詞	一般		5	10	0.003	-0.007	-3.77
11	認める	動詞	自立	18	0.011	-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	0.006	-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	-0.004	4.48	必要	名詞	サ変接続		15	18	0.009	-0.012	-3.11
14	治療切除不能	名詞	一般	7	0.004	0.000	4.29	有害事象	名詞	一般		4	8	0.002	-0.005	-3.02
15	妥当	名詞	一般	8	0.005	-0.001	4.22	薬投症	名詞	一般		3	7	0.002	-0.005	-2.95

左:上位15位、右:下位15位までそれぞれ抜粋した

5. 考察とまとめ

前回の研究では、テキストマイニングツールに、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

なる189行から、無作為に50行を選び、形態素解析結果の正答率を目視により比較した。その結果を、表3に示す。

表3 辞書の違いによる正答率の比較

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%

形態素解析にカスタマイズした辞書を用いることで、デフォルト辞書と比較して、正解率が11%向上した。

4.2 評判分析

次に、「機構」が主語となる189行を、TMSの評判分析によって、以下の4つに自動分類した。

- ・ Positive: 肯定的な行
- ・ Negative: 否定的な行
- ・ Positive & Negative: 肯定的・否定的両方の意味を含む行
- ・ Blank: 肯定的でも否定的でもない行

さらに、自動分類した結果と、予め専門家に上記の189行を同様の4種類に分類してもらった結果から、その一致数(一致率)を比較した。結果を表4に示す。

表4 評判判定と専門家の判定の一致数(一致率)

TMS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	結果	判定数
Positive	49(25%)	Positive	16(35%)
Negative	17(8%)	Negative	3(17%)
Positive & Negative	13(7%)	Positive & Negative	0
Blank	110(58%)	Blank	60(54%)
合計	189	合計	79(42%)

4.3 頻度分析

専門家によって「肯定的」「否定的」な行に分類された「機構」が主語となる189行の形態素解析の結果から、その形態素の出現頻度と出現割合を算出し、「肯定的」はプラス、「否定的」はマイナスとして、その差分で重み付けし、それぞれに特異的な形態素を抽出した。その結果をそれぞれランキング形式で並べ替えたものを表5に示す。

7)。

一方で、今回用いたTMSは、辞書の追加登録や、各種パラメータの変更が容易なので、テキストマイニングの対象に合わせて、ツールのカスタマイズが簡単に行える。さらに、様々な分析ツールが搭載されており、その結果も可視化されて表示され、感覚的にも分かりやすい。さらに、ある形態素が原文のどのような場面で使われているのかを確認するための原文検索機能も、分析を進めていく上で大いに役立った。

今回は、審査報告書の形態素解析に用いる辞書には、デフォルトの辞書とは別に、専門用語辞書として新たに「MEID辞書」と「ライフサイエンス辞書」を追加した。追加用語数は、重複を除き約22万語となったが、それでも審査報告書の解析には不十分であることがわかり、新たに追加候補となる用語の出現頻度を審査報告書から計算した上で、約100語の用語を追加登録している。それらの用語の例としては、「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)などである。

形態素解析を行う際は、専門用語集の辞書を追加するだけでは不十分な場合があり、解析対象とする文書をから、新たに用語を抽出して、対象に特化した辞書を作成することで、より精度の高い形態素解析が可能となると思われる。

また、前回の「なづき」の感性分析同様に、TMSの評判分析を用いて、審査報告書の行の内容が「肯定的: Positive」なのか「否定的: Negative」なのかを自動分類できるか実験を行ったが、その結果と専門家の判定結果との一致率は、低率であった。唯一、「肯定的でも否定的でもない行」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、あまり重要でない行を削ぎ落とすという点では役に立つかもしれない。今回の実験から、市販のテキストマイニングツールの分析機能を、審査報告書にそのまま適応しても、精度の高い結果が得られない可能性が高いと思われる。

今後の課題としては、行の判定に用いる形態素(評価語)の見直しや、構文内の形態素パターンを分析することによる、判定結果向上を検討して聞きたい。

続いて、頻度分析の結果を見てみると、前回の研究結果⁷⁾よりも、今回の結果の方がベバシズマブに特徴的と思われる用語が抽出されていた。特に、Negative

な形態素として、「高血圧」「消化管穿孔」「塞栓症」といったベバシズマブの特徴的な用語が上位に抽出されていた。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もある。

今回の実験では、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」の重みが、より一般化されるかも知れない。そして、それらの形態素を審査報告書の判定分析の評価語に用いたり、重み付けされた形態素を基に、構文の形態素パターンを分析したりする事で、より精度の高い医薬品の審査ポイントの抽出やその解析につながると思われる。ができるかも知れない。

6. 謝辞

本研究は、厚生労働科学研究費補助金(医薬品・医療機器等レギュラトリーサイエンス総合研究事業)の助成を受けたものである。

参考文献

- [1] EMA.Reflection paper on co-development of pharmacogenomic biomarkers and assays in the context of drug development (draft) . 24 June, 2010 http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2011/07/WC500108672.pdf
- [2] Guidance Documents (Medical Devices and Radiation-Emitting Products) Draft Guidance for Industry and Food and Drug Administration Staff - In Vitro Companion Diagnostic Devices <http://www.fda.gov/medicaldevices/deviceregulationandguidance/guidancedocuments/ucm262292.htm>
- [3] 嶋澤るみ子 池田正行 体外コンパニオン診断機器ガイダンス草案 翻訳 医薬品医療機器レギュラトリーサイエンス 2012;43(6):514-519.
- [4] PMDAのコンパニオン診断薬プロジェクト <http://www.pmda.go.jp/kijunsakusei/companion.html>
- [5] JAPIC日本の新薬審査報告書DB http://www.shinsahoukokusho.jp/dar_us/dar/search/usDarSearch.jsp
- [6] 豊田 裕貴,菰田 文男 特許情報のテキストマイニング 特許情報のテキストマイニング,2011,ミネルヴァ書房
- [7] 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 医療情報学連合大会論文集 2012;31:525-528

