ORIGINAL ARTICLE

Natural language processing-based assessment of consistency in summaries of product characteristics of generic antimicrobials

Rumiko Shimazawa[1] | Yoshinobu Kano[2] | Masayuki Ikeda[3] (ID)

[1]Department of Clinical Pharmacology, Tokai University School of Medicine, Isehara, Japan

[2]Faculty of Informatics, Shizuoka University, Hamamatsu, Shizuoka, Japan

[3]Department of Medical Informatics, Kagawa University Hospital, Kagawa, Japan

Correspondence
Masayuki Ikeda, Department of Medical Informatics, Kagawa University Hospital, Miki-cho Ikenobe, Kagawa, Japan.
Email: massie.ikeda@gmail.com

Abstract

To investigate consistency in summaries of product characteristics (SmPCs) of generic antimicrobials, we used natural language processing (NLP) to analyze and compare large amounts of text quantifying consistency between original and generic SmPCs. We manually compared each section of generic and original SmPCs for antimicrobials listed in the electronic Medicines Compendium in the United Kingdom, focusing on omissions and additions of clinically significant information (CSI). Independently, we quantified differences between the original and generic SmPCs using Kachako, a fully automatic NLP platform. Among the 137 antimicrobials listed in the electronic Medicines Compendium, we identified 193 pairs of original and generic antimicrobial SmPCs for the 48 antimicrobials for which generic SmPCs existed. Of these 193 pairs, 157 (81%) were consistent and 36 were inconsistent with the original SmPC. When the cut-off value of RATE (the index of similarity between two SmPCs) was set at 0.860, our NLP system effectively discriminated consistent generic SmPCs with a specificity of 100% and a sensitivity of 61%. We observed CSI omissions but not additions in the SmPC subsection related to pharmacokinetic properties. CSI additions but not omissions were found in the subsections dealing with therapeutic indications and fertility, pregnancy and lactation. Despite regulatory guidance, we observed substantial inconsistencies in the information in the United Kingdom SmPCs for antimicrobials. NLP technology proved to be a useful tool for checking large numbers of SmPCs for consistency.

KEYWORDS
Antibiotic resistance, drug information, drug regulation, general medicine, medication safety

1 | INTRODUCTION

The summary of product characteristics (SmPC)[1,2]—a controlled, standardized format for information about medicines in European Union (EU) member countries that is also called drug labeling—is a

fundamental resource for promoting the correct use of medicines.[3] SmPCs should regularly be reviewed and updated as new information emerges, because misleading information in SmPCs can result in avoidable adverse events, unnecessary treatment, and failure to treat. The majority of health care professionals believe that SmPCs include sufficient information to make rational decisions when prescribing or dispensing medicines. Previous studies have shown,

however, that drug labeling and SmPCs have omitted core clinical pharmacology information in the United States (USA)[4] and the EU.[5] Information about drug interactions in SmPCs has often been found to be incomplete and outdated[6] or even absent.[7]

According to the Quality Review of Documents guidance,[8] all relevant aspects of the content of generic SmPCs should be consistent with the SmPCs of reference medicinal products—the so-called "brand-name" medicines. This is because inconsistent information regarding medicines containing the same active ingredient contributes to confusion and poor prescription decisions. Although similar regulatory requirements are applied to generic medicines in the USA, inconsistency has been reported among bioequivalent medicines, both there[9] and in other countries.[7,10,11] To the best of our knowledge, only two original studies[9,10] have reported labeling inconsistencies regarding the same drug, authorized by the same regulatory agency. Storflor et al[10] investigated 71 generic labels of the 17 top-selling medicines in Norway, and Duke et al[9] reported that 77.9% of generic manufacturers produced labels that differed from those of the corresponding brand-name medicines in the USA. Evaluating the SmPCs of medicines marketed in the USA, the United Kingdom (UK), and Germany, Pfistermeister et al[7] found inconsistencies in labels for the same medicines among different regulatory bodies. Doogue and Thynne[11] reported inconsistencies in generic drug labels in Australia.

Evaluating inconsistencies among SmPCs is, however, time consuming, particularly because SmPCs in the UK contain a higher proportion of safety information than do those in the USA and Japan.[12] An automated, reproducible mechanism using natural language processing (NLP)[13] could expedite the evaluation of SmPCs on a large-scale basis.[9,13]

To quantify consistency between generic and original SmPCs, we used Kachako, an automated system for NLP.[14,15] Because they have a social impact that extends beyond individual patients,[16] we aimed to investigate consistency among the UK SmPCs of generic antimicrobials. The long-term risk of antimicrobial resistance is a major threat to global public health, leading to increased health-care costs, prolonged hospital stays, treatment failure, and excess mortality.[17,18] Substantial differences in safety information for the same drug have been reported among different regulatory bodies;[7] however, we are unaware of any studies on inconsistencies among drug labels for the same antimicrobial with marketing authorization through the same authority.

To determine inconsistencies between original and generic SmPCs, in this study, we examined omissions or additions of clinically significant information (CSI) in generic SmPCs, compared with original SmPCs. We focused on UK SmPCs for antimicrobials to demonstrate that NLP helps to quantify consistency among SmPCs.

2 | MATERIALS AND METHODS

2.1 | Data sources

The unit of analysis for this study was the SmPC, taken as a whole. It provides details of a medicine and its use for health care professionals.

An SmPC is a legal document approved as part of the marketing authorization of each medicine, but it is not a list of information on a specific group of medicines. One SmPC contains information on only one medicine. Each medicine, whether it is the original or a generic medicine, has its own SmPC. An SmPC consists of six sections: section 1 (name of the medicinal product), section 2 (qualitative and quantitative composition), section 3 (pharmaceutical form), section 4 (clinical particulars), section 5 (pharmacological properties), and section 6 (pharmaceutical properties).[2] SmPCs are updated as long as the product is on the market, as additional findings emerge.

In January 2016, we retrieved UK SmPCs and patient information leaflets about therapeutic antimicrobials for systemic use through the electronic Medicines Compendium[19] and we obtained US Structured Product Labels (US labels)[20,21] for the corresponding antimicrobials. We classified the antimicrobials as antibacterials, antimycotics, antimycobacterials, or antivirals according to the Anatomical Therapeutic Chemical Classification System.[22]

2.2 | Manual data analysis

Among the antimicrobials, we designated that with the oldest date of first marketing authorization as the original. The original SmPC was not always that of the brand-name antimicrobial, because some of these had been retired from the market. For precise quantification of the consistency between generic and original SmPCs and to identify the differences with high sensitivity, we selected pairs of SmPCs for original and generic antimicrobials that had the same dosage form and strength.

To create a gold standard for the NLP analysis,[13] we manually reviewed and compared these generic and original SmPCs according to the sections in the SmPCs.[1] Of the six SmPC sections, we focused on sections 4 (clinical particulars) and 5 (pharmacological properties), because the contents of these two sections have serious effects on the safe and effective use of the medicines, whereas the other sections do not or should not differ between the original and the generic. We classified concordance between the two types of SmPCs into the following five categories: (1) identical; (2) the same (i.e. different only in format, not content); (3) similar (i.e. with clinically nonsignificant differences in content); (4) CSI omissions (i.e. the presence of omissions of CSI included in the original SmPC); and (5) CSI additions (i.e. the presence of additions of CSI absent from the original SmPC). When the content of the generic SmPC was identical, the same or similar, we classified that SmPC as consistent with the original SmPC. Otherwise (i.e. when CSI omissions or additions were found in the generic SmPC), we classified it as inconsistent.

We took the criteria for clinical significance and its consistency to be whether omissions or additions of information affected the safety or efficacy in a regulatory context (Table 1). The criteria consisted of clearly objective conditions, allowing no room for individual judgment. Specifically, when the information was present in the original SmPC but absent from the generic SmPC and we found comparable information in the US label[21] for the same active pharmaceutical ingredient, we designated the status as CSI omission.

TABLE 1 Omission and addition of clinically significant and clinically non-significant information

	Original SmPC	Generic SmPC	US Label
CSI Omission	Present	Absent	Present
CSI Addition	Absent	Present	Present
CnSI Omission	Present	Absent	Absent
CnSI Addition	Absent	Present	Absent

CSI, clinically significant information; CnSI, clinically non-significant information; SmPC, summary of product characteristics.

When the information was absent from the original SmPC but present in the generic SmPC and we found comparable information in the corresponding US label, we designated the status as CSI addition (Table S1). When the comparable information was absent from the corresponding US label, we designated the difference in the presence of information between the original and generic SmPCs as clinically non-significant. (Table S2). The detailed procedures regarding this designation are described in Appendix A1.

2.3 | Evaluating the performance of our NLP system

For an independent analytical comparison with the gold standard (ie, our manual review), we quantified consistency among the generic SmPCs using our NLP system based on Kachako. We did this to evaluate the performance of the NLP system to support our semantic analysis[14,15]. Kachako was designed to thoroughly automate any procedure using services for unstructured information processing. Using Kachako, we calculated RATE, an index of similarity between two documents. When one SmPC was identical to the other, RATE was 1; when the two SmPCs were completely different, RATE was 0.

Given a pair of documents to compare, we calculated the RATE value by counting the number of the same tokens, normalized by the total number of tokens. Tokens of non-content words were discarded using parts of speech. The number of documents available for this study, <1000, which is extremely small and insufficient to create meaningful vectors (e.g. by word2vec/doc2vec). Moreover, these documents could contain different technical words, because they covered different domains of medicines; this could be detrimental when the data size was not large. Our measure, RATE, is robust for small sample size, and does not require training like supervised machine learning methods, as we show in our results.

To confirm the performance of RATE in recognizing differences between two regulatory documents for the same antimicrobial with the same dosage form and strength, we compared RATE between the original SmPC and corresponding patient information leaflet. We also compared RATE between the original SmPC and the corresponding US label.

2.4 | Statistical analysis

We calculated descriptive statistics for the outcome measures. The data are presented as means and standard deviations. To evaluate the performance of RATE, we used a receiver operating characteristic curve[23] and the area under the curve to quantify how well RATE performed in determining consistency between the original and generic SmPCs. We set an optimal cut-off value for the receiver operating characteristic analysis to maximize Youden's index[23], which is maximum = sensitivity + specificity−1. We used JMP software from SAS Institute Japan Ltd. (Tokyo, Japan), as appropriate.

3 | RESULTS

We found generic SmPCs available for 48 (35%) active pharmaceutical ingredients of antimicrobials among the 137 listed in the electronic Medicines Compendium. For the other 89, we found no generic available. Among the 48 with generic SmPCs, we identified 193 pairs of original and generic SmPCs for analysis (Table 2). After reviewing the data, we constructed a table with a specific set of CSI for a pair of SmPCs. When we checked whether the generic and original SmPCs had each type of CSI (Table S3), we found that the content, number, and section of CSI omissions varied widely among the generic SmPCs. All of the SmPCs had omissions, additions or neither of these; no SmPCs had both omissions and additions (Table S4). According to these features of CSI omission and addition, we defined the presence of any CSI omission or addition as an inconsistency. We treated the inconsistency in the same way whether there were single or multiple omissions or additions.

The manual data analysis identified 157 (81%) of the 193 pairs of SmPCs as consistent (identical, the same, or similar) with the original SmPC and 36 (19%) as inconsistent (CSI omissions or additions). As shown in Table 2, the 36 pairs of SmPCs comprised 33 antibacterial and three antimycotic SmPCs.

We quantified consistency among the SmPCs using RATE. RATE verified the result of the manual review of the 193 pairs of original and generic SmPCs. RATE was very effective in confirming the manually defined grade of concordance (identical, the same, similar, or CSI omissions or additions) between the generic and original SmPCs (Table 3). The 99% confidence interval (CI) for RATE indicated a clear disparity between the consistent (0.829-0.869) and inconsistent (0.612-0.702) generic SmPCs. Among the

TABLE 2 Generic summaries of product characteristics (SmPCs), their consistency and classification

Number of SmPCs (Number of APIs)	Generic SmPCs (APIs with generic SmPCs available)	Consistent generic SmPCs	Inconsistent generic SmPCs
Total in Antimicrobials	193 (48)	157	36
Antibacterials	145 (32)	112	33
Antimycotics	14 (2)	11	3
Antimycobacterials	3 (2)	3	0
Antivirals	31 (12)	31	0

APIs, active pharmaceutical ingredients.

PRP ASPET BRITISH PHARMACOLOGICAL SOCIETY

TABLE 3 RATE[a] stratified according to grade of concordance between a generic summary of product characteristics (SmPC) and that of the reference medicinal product

Grade of concordance[b]	Number of generic SmPCs	Mean	SD	99%CI[c]	Range
Consistent	157	0.849	0.096	0.829-0.869	0.583-0.962
Identical	64	0.923	0.021	0.916-0.930	0.875-0.955
Same	16	0.907	0.051	0.869-0.944	0.788-0.962
Similar	77	0.775	0.085	0.750-0.801	0.583-0.920
Inconsistent	36	0.657	0.104	0.612-0.702	0.505-0.858
CSI omissions	23	0.666	0.096	0.613-0.718	0.505-0.793
CSI additions	13	0.647	0.117	0.548-0.745	0.509-0.858
Total	193	0.810	0.124	0.787-0.833	0.505-0.962

[a]RATE is an index of similarity between a generic SmPC and that of the reference medicinal product

[b]Grade of concordance: (1) identical; (2) the same (different only in format, not content); (3) similar (with clinically non-significant differences in content); (4) CSI omissions (omissions of clinically significant information that is present in the reference medicinal product); (5) CSI additions (additions of clinically significant information that is absent from the reference medicinal product). When the contents of the generic SmPC were considered identical, the same or similar, the SmPC was classified as consistent with that of the reference medicinal product. Otherwise, the generic SmPC was classified as inconsistent.

[c]CI, confidence interval.

consistent generic SmPCs, the mean of RATE was significantly lower for similar SmPCs than it was for generic SmPCs that were identical or the same, with no overlap between 99% Cis for these point estimates. The sensitivity cut-off for including all of the consistent generic SmPCs was 0.583, the minimum of RATE for consistent SmPCs. The specificity cut-off for excluding any of the inconsistent SmPCs was 0.858, the maximum of RATE for inconsistent SmPCs (Table 3).

Figure 1 shows the receiver operating characteristic curve for RATE for consistency of the generic SmPCs. The area under the curve for RATE was 0.903 (standard error = 0.02). The optimal cut-off was 0.761 (Table 4A). Because of the small sample size, the 95% CI was wide: 29–100, for sensitivity and 31–100 for specificity. The cut-off value for RATE of 0.860 effectively discriminated consistent generic SmPCs and excluded inconsistent ones (specificity of 100% and sensitivity of 61%, Table 4B). We could not calculate the 95% CI for sensitivity and specificity at the cut-off value of 0.860, because the number was 0 for the inconsistent SmPCs with RATE of 0.860 or higher.

We also confirmed the performance of RATE in identifying obviously different documents for the same medicine. Among the 137 active pharmaceutical ingredients of antimicrobials in the electronic Medicines Compendium, we found 32 active pharmaceutical ingredients with both original and generic SmPCs where the corresponding UK patient information leaflet and US label for the same antimicrobial had the same dosage form and strength. We calculated the resultant 64 RATEs between the SmPCs and the UK
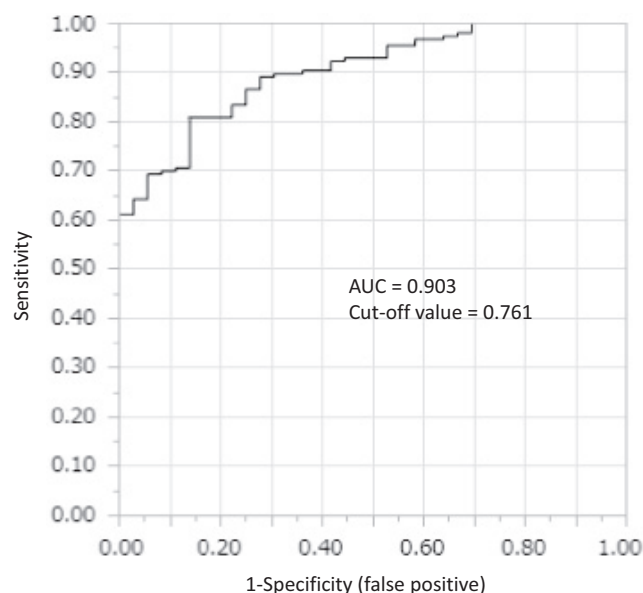

FIGURE 1 Receiver operating characteristic (ROC) curve for RATE (index of similarity between two documents) for consistency of generic summaries of product characteristics (SmPCs). ROC was performed for RATE for 193 generic SmPCs analyzed for similarity to the original SmPC. The area under the curve (AUC) for RATE was 0.903 (standard error = 0.02). The cut-off value to maximize Youden's index (maximum = sensitivity + specificity−1) was 0.761

TABLE 4 Sensitivity and specificity for generic summaries of product characteristics (SmPCs) consistent with the original using the RATE[a] value of 0.860 (A) or 0.761 (B) as the cut-off point

		Consistency		
		Consistent	Inconsistent	
(A)				
RATE[a]	0.761<	127 (Identical 64, Same 16, Similar 47)	5 (CSI Omission 3, CSI Addition 2)	158
	0.761>	30 (Identical 0, Same 0, Similar 30)	31 (CSI Omission 20, CSI Addition 11)	35
		157	36	193
		Sensitivity = 81% (95% CI ~29%)	Specificity = 86% (95% CI ~31%)	
(B)				
RATE[a]	0.860<	96 (Identical 64, Same 14, Similar 18)	0	96
	0.860>	61 (Identical 0, Same 2, Similar 59)	36 (CSI Omission 23, CSI Addition 13)	97
		157	36	193
		Sensitivity = 61%	Specificity = 100%	

[a]RATE is an index of similarity between a generic SmPC and the SmPC of the original.

patient information leaflets and between the SmPCs and the US labels. For most pairs, RATE between the SmPC of the original product and the corresponding patient information leaflet was much lower than that between the SmPC and the corresponding US

label. For all pairs, RATE between the original SmPC and the corresponding generic SmPC was higher than that between the SmPC and the corresponding US label as well as RATE between the original SmPC and the corresponding PIL (Figure 2).

With regard to the type of inconsistency between the original and generic SmPCs, the frequency of CSI omissions and additions varied depending on the specific antimicrobial (Table 5). We found CSI omissions in 20 generic antibacterial and three generic antimycotic SmPCs. CSI additions were observed in 13 generic antibacterial SmPCs but in no generic antimycotic SmPCs. We identified generic SmPCs with CSI omissions most frequently for amoxicillin. The percentage of generic SmPCs that were inconsistent with the original SmPC varied according to the medicine; for example, this percentage was 100% (3/3) for minocycline, 80% (8/10) for amoxicillin, and 10% (1/10) for clarithromycin. Both types of inconsistency—CSI omissions and additions—were found in the generic SmPCs only for vancomycin (Table 5). Of the three generic vancomycin SmPCs, two had omissions, whereas the other had additions. For other antimicrobials, none had SmPCs with both omissions and additions. For example, of the 10 generic amoxicillin SmPCs, eight had omissions and the other two had no
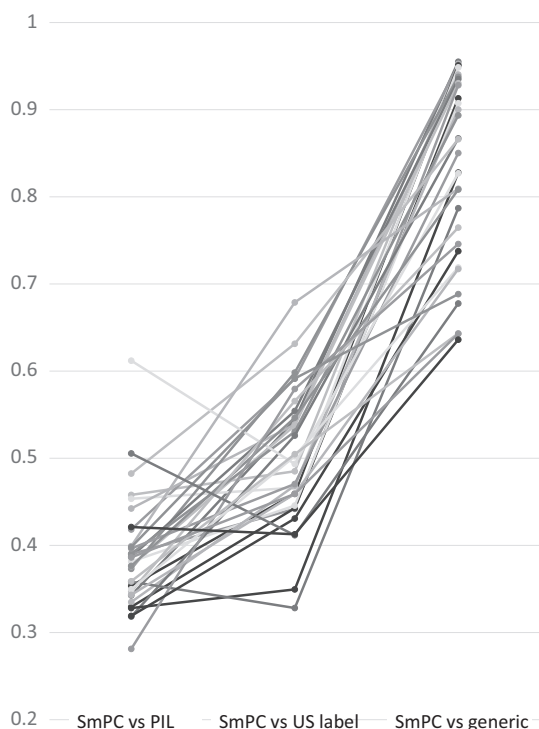
inconsistencies, whereas all of the four generic phenoxymethylpenicillin SmPCs had additions and none had omissions.

Table 6 shows the frequency of inconsistencies in subsections of SmPC sections 4 (clinical particulars) and 5 (pharmacological properties). CSI omissions were found most frequently in subsection 5.2 (pharmacokinetic properties, 85%) followed by subsection 4.2 (posology and method of administration, 65%). We identified CSI additions most frequently in subsection 5.1 (pharmacodynamic properties, 69%) followed by subsections 4.8 (undesirable effects, 46%) and 4.4 (special warnings and precautions for use, 46%). We observed CSI omissions but not additions in subsection 5.2 (pharmacokinetic properties). CSI additions but not omissions were found in subsection 4.1 (therapeutic indications) for the four generic phenoxymethylpenicillin SmPCs and in subsection 4.6 (fertility, pregnancy, and lactation) for the two generic clindamycin SmPCs. The phenoxymethylpenicillin SmPCs offered advice about antimicrobial resistance. The clindamycin SmPCs provided detailed descriptions of the potential side effects on pregnancy. In other subsections, we found both omissions and additions of CSI. There were no CSI omissions or additions in subsections 4.3 (contraindications), 4.7 (effects on ability to drive and use machines), or 4.9 (overdose).

FIGURE 2 Direct comparison of RATE (index of similarity between two documents) with the same active pharmaceutical ingredient in the same dosage form (n = 32). For most pairs, RATE between the summary of product characteristics (SmPC) of the original product and the corresponding patient information leaflet (PIL) was much lower than that between the SmPC and the corresponding US Structured Product Label (US label). For all pairs, RATE between the original SmPC and the corresponding generic SmPC was higher than that between the SmPC and the corresponding US label as well as RATE between the original SmPC and the corresponding PIL

TABLE 5 Consistency in generic antimicrobial summaries of product characteristics (SmPCs)

Antimicrobials (No. of generic SmPCs)	Inconsistent (% in available generic SmPCs)	
	CSI omissions[a]	CSI additions[b]
Total of Antibacterials	20	13
Amoxicillin (10)	8 (80%)	0
Ampicillin (1)	1 (100%)	0
Azithromycin (2)	0	2 (100%)
Ceftriaxone (2)	2 (100%)	0
Cefuroxime (5)	1 (20%)	0
Clarithromycin (10)	1 (10%)	0
Clindamycin (2)	0	2 (100%)
Erythromycin (12)	2 (17%)	0
Gentamicin (3)	0	1 (33%)
Lymecycline (1)	0	1 (100%)
Minocycline (3)	3 (100%)	0
Phenoxymethylpenicillin (4)	0	4 (100%)
Pivmecillinam (1)	0	1 (100%)
Trimethoprim (3)	0	1 (33%)
Vancomycin (4)	2 (50%)	1 (25%)
Total of Antimycotics	3	0
Fluconazole (10)	2 (20%)	0
Itraconazole (4)	1 (25%)	0

[a]Generic SmPCs with omissions of clinically significant information that is present in the original SmPC.
[b]Generic SmPCs with additions of clinically significant information that is absent from the original SmPC.

TABLE 6 Frequency of subsections with inconsistencies in generic summaries of product characteristics (SmPCs)

SmPC Section	Number of SmPCs	
	CSI omissions[a] (%) n = 20	CSI additions[b] (%) n = 13
4. Clinical particulars		
4.1 Therapeutic indications	0	4 (31%)
4.2 Posology and method of administration	13 (65%)	3 (23%)
4.3 Contraindications	0	0
4.4 Special warnings and precautions for use	8 (40%)	6 (46%)
4.5 Interaction with other medicinal products and other forms of interaction	7 (35%)	4 (31%)
4.6 Fertility, pregnancy and lactation	0	2 (15%)
4.7 Effects on ability to drive and use machines	0	0
4.8 Undesirable effects	6 (30%)	6 (46%)
4.9 Overdose	0	0
5. Pharmacological properties		
5.1 Pharmacodynamic properties	10 (50%)	9 (69%)
5.2 Pharmacokinetic properties	17 (85%)	0

Data represent percentages of generic SmPCs inconsistent with the original in the corresponding SmPC section.

[a]Omissions of clinically significant information that is present in the original SmPC.

[b]Additions of clinically significant information that is absent from the original SmPC.

4 | DISCUSSION

Despite regulatory requirements[8], we found a substantial number of generic antimicrobial SmPCs that were inconsistent with the corresponding original SmPCs. The present study demonstrated that our NLP system was able to quantify consistency among generic and original SmPCs. RATE showed sufficient power, with no overlap of the 99% CIs, in discriminating generic SmPCs that were consistent with the original SmPCs from those that were inconsistent with the original SmPCs. For the RATE cut-off of 0.860, the specificity was 100% and the sensitivity was 61% for detecting consistent generic SmPCs. This means that we may focus exclusively on SmPCs with RATE values less than 0.860 when identifying inconsistencies among generic SmPCs. In this study, with RATE set at 0.860 or above, we were able to exclude 96 of 193 generic SmPCs and focus on the remaining 97 to review their content and manually investigate clinically significant inconsistencies. With regard to using RATE to compare SmPCs, patient information leaflets and US labels, the clear distinction between the RATE of generic SmPCs and that of other documents indicated that RATE has sufficient power to discriminate different regulatory documents for the same product.

We found CSI omissions and additions between generic antimicrobial SmPCs and the SmPCs for the corresponding originals. CSI omissions were more common than were CSI additions, and both were observed in generic antibacterial and antimycotic SmPCs. We could not determine why CSI present in original SmPCs was omitted from generic SmPCs; however, it is easy to imagine that generic companies seek to make their SmPCs more "concise" by avoiding partial overlap of the sections in an SmPC, which are not mutually exclusive or collectively exhaustive. When an adverse event is described in one section, the event might not be touched on or, in contrast, might be explained more extensively in another section. We had not expected CSI additions, because generic manufacturers are not responsible for updating SmPCs.[8]

The distribution of CSI omissions and additions in the SmPC subsections showed both similarities and differences between the two types of inconsistency. The high proportion of CSI omissions without additions in subsection 5.2 (pharmacokinetic properties) indicated the removal of redundancy in the content of the corresponding original SmPC. In subsection 4.1 (therapeutic indications) for phenoxymethylpenicillin SmPCs and subsection 4.6 (fertility, pregnancy and lactation) for clindamycin SmPCs, we observed CSI additions but not omissions. The content of the additional information in those generic SmPCs related to clinically relevant action intended to reduce adverse events, such as antimicrobial resistance and foetal toxicity.

The subsections in which both additions and omissions of CSI were found indicated diversity: 4.2 (posology and method of administration), 4.4 (special warnings and precautions for use), 4.5 (interaction with other medicinal products and other forms of interaction), 4.8 (undesirable effects), and 5.1 (pharmacodynamic properties). In these subsections, it is difficult to adjust the content to optimize the risk-benefit balance of the medicines. In contrast, subsections 4.3 (contraindications), 4.7 (effects on ability to drive and use machines), and 4.9 (overdose), in which neither omissions nor additions were observed, allow little room for inconsistency in generic SmPCs.

Overall, the inconsistencies revealed in this study may result in important information being overlooked, complications in clinical practice, and increased risk of prescription errors and adverse events. We had supposed that any inconsistencies in the SmPCs of generic antimicrobials would be minimal because these inconsistencies could lead to antimicrobial misuse and public health risks, such as antimicrobial resistance.[24] Storflor et al[10] found that generic labels for 13 of the 17 top-selling medicines in Norway had discrepancies, mainly in the information on side effects. Duke et al[9] reported that 68% of multi-manufacturer medicines in the USA had discrepancies inverse drug reactions in safety labeling and that 77.9% of generic manufacturers produced labels that differed from those of the brand-name medicine. The higher percentages of inconsistencies found in generic labels in these two previous studies compared with those found in our study may have resulted from differences in several factors, including the criteria for identifying inconsistencies, the labeling sections of focus and therapeutic areas of interest.

Along with previous studies, the present study has identified substantial inconsistencies among generic SmPCs. These findings point to a challenging issue: harmonization across generic medicines. A number of reasons, such as limited technical, human, and financial resources, may explain these inconsistencies. Legal requirements for generic manufacturers to update their SmPCs as new data become available are impractical: Such companies are unlikely to have the resources of brand-name pharmaceutical companies for conducting post-marketing surveys and data collection. Given these challenges, the existing scheme for updating SmPCs has to undergo fundamental change to achieve harmonization across generic medicines. Such change could be supported by a system capable of monitoring inconsistencies among generic SmPCs on an ongoing basis.[9] Implementing structured, standardized electronic SmPCs will also help to reduce inconsistencies and improve prescribing decisions.[25]

We recognize several limitations of the present study. First, the cross-sectional design did not allow us to identify why, despite regulatory requirements, inconsistencies exist in generic SmPCs. Further research is required to clarify what produces these inconsistencies. However, we assume that multiple factors, such as time after marketing authorization, sales quantity, and post-marketing data, are involved. Second, despite the manual review we undertook to exclude insignificant differences, such as formatting, the inconsistencies found in this study may not necessarily be relevant to clinical practice in the real world. These inconsistencies were defined in a regulatory context and may not affect health care professionals, who would not need generic SmPCs if they learned the essential information for the safe and effective use of the brand-name medicine before the generics came out. In the real world, however, this is not always the case, because no one can perfectly remember all of the information in the SmPCs and because new and important information often comes out even after many generics are available. Third, we restricted our analysis to UK SmPCs of antimicrobials for systemic use. The wide range of the 95% CIs for sensitivity and specificity indicate the need for further studies using a larger number of SmPCs to evaluate the performance of RATE. More issues could have arisen had we extended our analysis: Inconsistencies might vary depending on the countries, regulatory bodies, and therapeutic areas involved.

In conclusion, we demonstrated that our NLP system, based on the Kachako platform, helped to quantify consistency among SmPCs and extracted inconsistencies between generic and original SmPCs. Inconsistencies among SmPCs for the same drug authorized by the same authority indicate that the existing regulatory scheme does not work effectively in terms of achieving consistency across generic SmPCs. However, NLP can address the challenge of checking large numbers of regulatory documents for consistency. Further research on rapidly comparing, correcting, and updating SmPCs should contribute to harmonization among generic SmPCs and, ultimately, to the production of a centralized online drug information and safety resource.

DISCLOSURE

The authors report no conflict of interest in performing this study.

ORCID

Masayuki Ikeda http://orcid.org/0000-0001-5157-1362

REFERENCES

1. Datapharm Communications Limited. Glossary The electronic Medicines Compendium. https://www.medicines.org.uk/emc/glossary/. Accessed May 22, 2018.
2. European Medicines Agency. Summary of product characteristics (SmPC). What is it and what does it contain? http://www.ema.europa.eu/docs/en_GB/document_library/Presentation/2013/01/WC500137035.pdf. Accessed May 22, 2018.
3. van Dijk L, Monteiro SP, Vervloet M, de Bie J, Raynor DK. Study on the package leaflets and the summaries of product characteristics of medicinal products for human use: PIL-S study. http://ec.europa.eu/health/files/committee/75meeting/pil_s.pdf. Accessed 22 May 2018.
4. Spyker DA, Harvey ED, Harvey BE, et al. Assessment and reporting of clinical pharmacology information in drug labeling. *Clin Pharmacol Ther*. 2000;67:196-200.
5. Arguello B, Fernandez-Llimos F. Clinical pharmacology information in summaries of product characteristics and package inserts. *Clin Pharmacol Ther*. 2007;82:566-571.
6. Bergk V, Haefeli WE, Gasse C, Brenner H, Martin-Facklam M. Information deficits in the summary of product characteristics preclude an optimal management of drug interactions: a comparison with evidence from the literature. *Eur J Clin Pharmacol*. 2005;61:327-335.
7. Pfistermeister B, Sass A, Criegee-Rieck M, Burkle T, Fromm MF, Maas R. Inconsistencies and misleading information in officially approved prescribing information from three major drug markets. *Clin Pharmacol Ther*. 2014;96:616-624.
8. European Medicines Agency. QRD general principles regarding the SmPC information for a generic/hybrid/biosimilar product. EMA/627621/2011. http://www.ema.europa.eu/docs/en_GB/document_library/Regulatory_and_procedural_guideline/2012/05/WC500127589.pdf. Accessed 22 May 2018.
9. Duke J, Friedlin J, Li X. Consistency in the safety labeling of bioequivalent medications. *Pharmacoepidemiol Drug Saf*. 2013;22:294-301.
10. Storflor JG, Pettersen LC, Slordal L, Spigset O. Drug package inserts–varying information for the same medicines. *Tidsskr Nor Laegeforen*. 2013;133:955-959.

11. Doogue MP, Thynne TR. Product information for generic drugs: old, unloved and sometimes unsafe. *Intern Med J*. 2013;43:1346-1348.

12. Shimazawa R, Ikeda M. Safety information in drug labeling: a comparison of the USA, the UK, and Japan. *Pharmacoepidemiol Drug Saf*. 2013;22:306-318.

13. Velupillai S, Mowery D, South BR, Kvist M, Dalianis H. Recent advances in clinical natural language processing in support of semantic analysis. *Yearb Med Inform*. 2015;10:183-193.

14. Kano Y. Kachako: a hybrid-cloud unstructured information platform for full automation of service composition, scalable deployment and evaluation. In: Ghose A, Zhu H, Yu Q, Delis A, Sheng QZ, Perrin O, Wang J, Wang Y, eds. *Service-Oriented Computing - ICSOC 2012 Workshops*. Lecture Notes in Computer Science, Vol. 7759. Berlin: Springer; 2013:72-84.

15. Kano Y. Kachko. http://kachako.org/ Accessed 22 May 2018.

16. Shehab N, Patel PR, Srinivasan A, Budnitz DS. Emergency department visits for antibiotic-associated adverse events. *Clin Infect Dis*. 2008;47:735-743.

17. Gualano MR, Gili R, Scaioli G, Bert F, Siliquini R. General population's knowledge and attitudes about antibiotics: a systematic review and meta-analysis. *Pharmacoepidemiol Drug Saf*. 2015;24:2-10.

18. European centre for disease prevention and control. Summary of the latest data on antibiotic resistance in the European Union. http://ecdc.europa.eu/en/publications/Documents/antibiotic-resistance-in-EU-summary.pdf. Accessed 22 May 2018.

19. Datapharm Communications Limited. The electronic Medicines Compendium. https://www.medicines.org.uk/emc/. Accessed 22 May 2018.

20. National Library of Medicine. DailyMed. http://dailymed.nlm.nih.gov/dailymed/index.cfm. Accessed 22 May 2018.

21. US Food and Drug Administration. Drugs@FDA. https://www.accessdata.fda.gov/scripts/cder/daf/. Accessed: May 22, 2018.

22. WHO Collaborating Centre for Drug Statistics Methodology. ATC/DDD Index. http://www.whocc.no/atc_ddd_index/. Accessed 22 May 2018.

23. Ruopp MD, Perkins NJ, Whitcomb BW, Schisterman EF. Youden Index and optimal cut-point estimated from observations affected by a lower limit of detection. *Biom J*. 2008;50:419-430.

24. World Health Organization. The evolving threat of antimicrobial resistance: options for action. http://apps.who.int/iris/bitstream/10665/44812/1/9789241503181_eng.pdf. Accessed 22 May 2018.

25. Maxwell S, Eichler HG, Bucsics A, Haefeli WE, Gustafsson LL. e-SPC - delivering drug information in the 21st century: developing new approaches to deliver drug information to prescribers. *Br J Clin Pharmacol*. 2012;73:12-15.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

APPENDIX A1

In this section, we describe the methods we used to designate the omission and addition of clinically significant information (CSI) and of clinically non-significant information. We did this according to the presence or absence of this information in the original summary of product characteristics (SmPC), the generic SmPC, and the corresponding US label, and we used the UK SmPC in the absence of a corresponding US label.

To exclude individual bias and ensure objectivity in the judgment of CSI, we focused on the commonality of the information among the original SmPC, the generic SmPC, and the US label. When the information was present in the original SmPC but absent in the generic SmPC and we found comparable information in the US label for the same active pharmaceutical ingredient, we designated the status as CSI omission. When the information was absent from the original SmPC but present in the generic SmPC and we found comparable information in the corresponding US label, we designated the status as CSI addition. Below are examples of omission and addition of CSI.

An example of an omission of CSI is the generic SmPCs for erythromycin. In the Erythrocin (brand name of erythromycin) SmPC, there was a specific warning against antimicrobial resistance. A similar warning was found in the US label, but the generic erythromycin SmPCs had no such warning.

The SmPC for Erythrocin. (Section 4.5: Interaction with other medicinal products and other forms of interaction) included the following warning:

> Anti-bacterial agents: Erythromycin antagonizes the action of clindamycin, lincomycin, and chloramphenicol.

The US label for erythromycin included the following warning:

> Interactions with Other Antibiotics: Antagonism exists in vitro between erythromycin and clindamycin, lincomycin, and chloramphenicol.

An example of an addition of CSI is the generic SmPC for azithromycin. In the Zithromax (brand name of azithromycin) SmPC, there was no specific warning against antimicrobial resistance, but a generic azithromycin SmPC included the following text:

4.4 | Special warnings and precautions for use

The following should be considered before prescribing azithromycin:

Azithromycin tablets are not suitable for treatment of severe infections where a high concentration of the antibiotic in the blood is rapidly needed. Azithromycin is not the first choice for the empiric treatment of infections in areas where the prevalence of resistant isolates is 10% or more (see section 5.1). In areas with a high incidence of erythromycin A resistance, it is especially important to take into consideration the evolution of the pattern of susceptibility to azithromycin and other antibiotics. As for other macrolides, high resistance rates of *Streptococcus pneumoniae* (>30 %) have been reported for azithromycin in some European countries (see section 5.1). This should be taken into account when treating infections caused by *S. pneumoniae*.

The US label for azithromycin also warns against antimicrobial resistance:

1.3 | Limitations of use

Azithromycin should not be used in patients with pneumonia who are judged to be inappropriate for oral therapy because of moderate to severe illness or risk factors such as any of the following:

1. patients with cystic fibrosis,
2. patients with nosocomial infections,
3. patients with known or suspected bacteremia,
4. patients requiring hospitalization,
5. elderly or debilitated patients, or
6. patients with significant underlying health problems that may compromise their ability to respond to their illness (including immunodeficiency or functional asplenia).

1.4 | Usage

To reduce the development of drug-resistant bacteria and maintain the effectiveness of ZITHROMAX (azithromycin) and other antibacterial drugs, ZITHROMAX (azithromycin) should be used only to treat infections that are proven or strongly suspected to be caused by susceptible bacteria. When culture and susceptibility information are available, they should be considered in selecting or modifying antibacterial therapy. In the absence of such data, local epidemiology and susceptibility patterns may contribute to the empiric selection of therapy.

5.8 | Development of drug-resistant bacteria

Prescribing ZITHROMAX in the absence of a proven or strongly suspected bacterial infection is unlikely to provide benefit to the patient and increases the risk of the development of drug-resistant bacteria.