



Examination of the Regulatory Frameworks Applicable to Biologic Drugs (Including Stem Cells and Their Progeny) in Europe, the U.S., and Australia: Part II—A Method of Software Documentary Analysis

NINA ILIC,^{a,b} SNEZANA SAVIC,^a EVAN SIEGEL,^c KERRY ATKINSON,^{d,e} LJILJANA TASIC^a

Key Words. Cellular therapy • Clinical translation • Clinical trials • Ethics

ABSTRACT

A wide range of regulatory standards applicable to production and use of tissues, cells, and other biologics (or biologicals), as advanced therapies, indicates considerable interest in the regulation of these products. The objective of this study was to analyze and compare high-tier documents within the Australian, European, and U.S. biologic drug regulatory environments using qualitative methodology. Eighteen high-tier documents from the European Medicines Agency (EMA), U.S. Food and Drug Administration (FDA), and Therapeutic Goods Administration (TGA) regulatory frameworks were subject to automated text analysis. Selected documents were consistent with the legal requirements for manufacturing and use of biologic drugs in humans and fall into six different categories. Concepts, themes, and their co-occurrence were identified and compared. The most frequent concepts in TGA, FDA, and EMA frameworks were “biological,” “product,” and “medicinal,” respectively. This was consistent with the previous manual terminology search. Good Manufacturing Practice documents, across frameworks, identified “quality” and “appropriate” as main concepts, whereas in Good Clinical Practice (GCP) documents it was “clinical,” followed by “trial,” “subjects,” “sponsor,” and “data.” GCP documents displayed considerably higher concordance between different regulatory frameworks, as demonstrated by a smaller number of concepts, similar size, and similar distance between them. Although high-tier documents often use different terminology, they share concepts and themes. This paper may be a modest contribution to the recognition of similarities and differences between analyzed regulatory documents. It may also fill the literature gap and provide some foundation for future comparative research of biologic drug regulations on a global level. *STEM CELLS TRANSLATIONAL MEDICINE* 2012;1:909–920

INTRODUCTION

Although therapeutics derived from biological sources have been subject to regulatory oversight for some time (i.e., monoclonal antibodies), the biologic products used in transplantation procedures (i.e., cells and tissues) have historically been exempted from such oversight [1]. The unique source of the “active ingredients” renders biologic drugs difficult to assess by traditional regulatory systems, which provide support to conventional pharmaceutical manufacturing processes [2]. New considerations have led the existing regulatory agencies of the developed world to propound new regulatory approaches for the sector of biologic drugs or biopharmaceuticals that “has become one of the most research-intensive sectors with a great potential for delivering innovative human medicines in the future” [3]. Regulation of cell and tissue products

is closely linked not only to sensitive areas of public health policy and funding but also to the development of novel disciplines such as regulatory science [4, 5]. Biologic drug regulations are widely developed by mature regulatory agencies; for example, the Australian framework is administered by the Therapeutic Goods Administration (TGA), the European Union’s by the European Medicines Agency (EMA), and that of the U.S. by the Food and Drug Administration (FDA) [6].

The intention of this research is to compare the content and terminology in high-tier regulatory documents related to biologic drugs in Australia [7–10], Europe [11, 12], and the U.S. [13–16], where these products are also referred to as biologicals, advanced therapy medicinal products, and biologics, respectively. It was hypothesized that there are similarities between them although the terminology and focus of

^aFaculty of Pharmacy, University of Belgrade, Belgrade, Serbia; ^bMater Health Services, Brisbane, Queensland, Australia;

^cGround Zero Pharmaceuticals, Irvine, California, USA; ^dMater Medical Research Institute, Brisbane, Queensland, Australia; ^eSchool of Medicine, University of Queensland, Brisbane, Queensland, Australia

Correspondence: Nina Ilic, B.Sc., M.Sc., M.B.A., Raymond Terrace, South Brisbane 4101, Queensland, Australia. Telephone: 61-7-3163-2779; Fax: 61-7-3163-2701; E-Mail: Nina.Ilic@mater.org.au

Received April 10, 2012; accepted for publication August 30, 2012; first published online in *SCTM EXPRESS* December 4, 2012.

©AlphaMed Press
1066-5099/2012/\$20.00/0

<http://dx.doi.org/10.5966/sctm.2012-0038>

regulatory frameworks differ to some extent. It was also anticipated that high-tier regulatory documents could be taken as a central representation of the regulatory frameworks from which they were drawn. The overall aim of the study was to establish a correlation between manual [17] and software-based documentary analysis, to assess usability of the software analysis, and to ascertain the use of a qualitative methodology in the highly specialized area of regulatory science. Additionally, the aim of this approach was to contribute to a generalization rather than a simple statement of sameness or difference between the documents, not only for the sake of future development of biologic drugs but for further development of associated regulations in less developed regional and national regulatory sets of principles (i.e., Southeast Asian or Eastern European countries).

An inherent complexity, the historical background, and the differences between regulatory frameworks applicable to biologic drug development and application are reflected in the terminology they use [17]. The computer-aided qualitative data analysis (CAQDAS) software [18] Leximancer v3.1 (Leximancer Pty. Ltd., Jindalee, QLD, Australia, <https://www.leximancer.com>), was used in this research [19]. The documents required the ability to analyze them while applying other skills, such as recognizing certain textual forms or genres, and the ability to judge, interpret, or contextualize [20, 21]. Concepts and themes were identified (conceptual analysis) in groups of documents, and the co-occurrence patterns between them (relational analysis) were explored across the entire data corpus [22]. It is anticipated that during the future harmonization of regulations and guidances these similarities can potentially be used as a feasible source of knowledge and expertise.

Theory of Terminology

The general theory of terminology arises from the work of Wüster (1898–1977), and it is based on the significance of concepts and their delineation from each other [23]. The nature of concepts, conceptual relations, relationships between terms and concepts, and the designation of terms from concepts are the main features of the general theory of terminology [24]. The domain of concepts is seen as independent from the domain of terms, and terminology delineation continuously flows from concepts working their way to terms [25]. Although the general theory of terminology has been criticized by other authors—such as its naivety in presenting the concept as a cornerstone of any terminology [26] or being overly reliant upon semantics [27]—it provides a relevant start point for this research [23]. First, it enhances the features of the software used in documentary analysis. Second, the descriptive structure of concepts in this theory (i.e., assigning a bundle of characteristics to a concept) fulfills the need of the research to compare but not analyze specific characteristics. Finally, terminology or specialized language is more than a technical or particular instance of general language in the modern society in which we live [25]. In today's society with its emphasis on science and technology, the way specialized concepts are named, structured, described, and understood has put terminology in the spotlight. The information in scientific and technical texts is encoded in terms or specialized knowledge units, which can be regarded as access points to more complex knowledge structures. As such, they only mark the tip of the iceberg [25]. Complex knowledge beneath the surface can be considered a concept or a collection of specialized knowledge [25]. The specialist knowledge may be inherent to a scientific

discipline represented by an entity or a group of experts (i.e., a regulatory agency). This specialized knowledge may be represented, along with its specific characteristics, within the publications (i.e., scientific journals) or documents (i.e., high-tier regulatory documents). Hence, the general terminology theory's attitude toward a concept makes a valid approach for this particular analysis. It is important to state that any attempt toward semantic analysis, analysis of the rules of syntax, or cognitive aspects of the language and terminology are beyond the scope of this study.

Sampling Considerations

Most intellectual work involved in sampling and selection concerns establishing an appropriate relationship between the sample and the wider universe. There were both empirical and theoretical elements to this study. In an empirical sense this research involved the groups, countries/regions, policies, considerations, and a broad spectrum of activities related to cell and tissue therapies or biologic drug development, manufacturing, and application. In a theoretical sense this research entailed some of the theories emerging from the social sciences and applied them to the complex world of regulatory science. While doing so, the study conveys complexity on numerous levels relating to regional regulations (in terms of the scope, biologic drug groups, and related terminology) [28] and complexity of the research methodology [17]. The sampling was not inherently about empirical representation, as it tends to be in some quantitative forms of research [21]. The type of sampling used in this research was theoretical or purposive sampling for high-tier regulatory documents that are, for the purposes of this research, taken as a central representation of the regulatory frameworks from which they were drawn.

Software Considerations

Content analysis is a research tool used for determining the presence of words or concepts in collections of textual documents [29]. It is used for breaking down the material into manageable categories and relations that can be quantified and analyzed. These extracted measurements can be used to make valid inferences about the material contained within the text [29]. Leximancer software is increasingly used instead of hand coding, and it has advantages in some research domains [30]. It has been used in research to analyze opinion polls [31], to assess enterprise risks in a specific industry [32, 33], to evaluate marine operations accidents [34] and academic performance characteristics [35], and to track changes in journals [36]. It has also been applied in health research into practitioner training and perceptions of common practices [37, 38], patient safety and incident reports [39], patient management and perceptions of health care [40], and assessing the change of health practices [41].

As in this study, Leximancer has been applied to concept mapping in different contexts [42, 43] and for the purpose of identifying themes and patterns [44, 45]. Although it has been applied extensively, the use of Leximancer software in this study was somewhat unusual. To our knowledge, this the first study to apply Leximancer software to assess high-tier, or any other, regulatory documents pertinent to biopharmaceuticals, biologics, or any other medicinal products.

Table 1. List of analyzed documents

Group of documents	Document title	Document version	Year of issue/ revised	Current Document
Cohort 1				
EMA				
Document Code				
EMA DOC1	Directive 2001/83/EC	n/a	2001	Yes ^a
EMA DOC2	Regulation 726/2004	n/a	2004	Yes ^a
EMA DOC3	Regulation 1394/2007	n/a	2007	Yes
FDA ^b				
Document Code				
FDA DOC1	21CFR312	n/a	2011	Yes
FDA DOC2	21CFR600 Biological Products: General	n/a	2011	Yes
FDA DOC3	21CFR1271 Human Cells, Tissues, and Cellular and Tissue-based Products	n/a	2011	Yes
TGA				
Document Code				
TGA DOC1	Australian Regulatory Guidelines for Biologicals Part 1	1.0	2011	Yes
TGA DOC2	Australian Regulatory Guidelines for Biologicals Part 2	1.0	2011	Yes
TGA DOC3	Australian Regulatory Guidelines for Biologicals Part 3	1.0	2011	Yes
Cohort 2				
Good Manufacturing Practice (1)				
Document Code				
GMP DOC1	Directive 2003/94/EC (EMA)	n/a	2003	Yes
GMP DOC2	Australian Code of Good Manufacturing Practice—Human Blood and Tissues (TGA)	n/a	2000	Yes ^c
GMP DOC3	Australian Code of Good Manufacturing Practice for Medicinal Products (TGA)	n/a	2002	Yes
Good Manufacturing Practice (2)				
Document Code				
GMP DOC4	21CFR211 Current Good Manufacturing Practice for Finished Pharmaceuticals (FDA)	n/a	2011	Yes
GMP DOC5	Guide to Good Manufacturing Practice for Medicinal Products—Part I (PIC/S)	n/a	2009	Yes
GMP DOC6	Guide to Good Manufacturing Practice for Medicinal Products—Part II (PIC/S)	n/a	2009	Yes
Good Clinical Practice				
Document Code				
GCP DOC1	Directive 2003/20/EC (EMA)	n/a	2001	Yes
GCP DOC2	Guidance for Industry E6 Good Clinical Practice (FDA)	n/a	1996 ^d	Yes
GCP DOC3	Note for Guidance on Good Clinical Practice (CPMP/ICH/135/95) Annotated (TGA)	n/a	2000	Yes

^aSubsequently amended.

^bDisclaimer: Major, specific, and updated FDA guidances relating to key subsets of biologics were not included in cohort 1 and 2 documents. Although the basic FDA regulations have not been changed much since the expansion of biologic drugs, including vaccines, the guidances have really become more like regulations in the interim. Due to the research sampling criteria, guidances (including the key manufacturing guidances for biologics that the FDA really does enforce in a way similar to that for regulations) have not been analyzed here. These documents can be found at the following link: <http://www.fda.gov/BiologicsBloodVaccines/GuidanceComplianceRegulatoryInformation/default.htm>.

^cIn the process of being amended.

^dDocument was exempt from the time frame.

Abbreviations: DOC, document; API, Active Pharmaceutical Ingredients; CFR, Code of Federal Regulations; EMA, European Medicines Agency; FDA, Food and Drug Administration; GCP, Good Clinical Practice; GMP, Good Manufacturing Practice; n/a, not applicable; PIC/S, Pharmaceutical Inspection Convention/Cooperation Scheme; TGA, Therapeutic Goods Administration.

MATERIALS AND METHODS

Materials

The study design, including document selection criteria and qualitative research considerations, has been already explained in detail [17]. An extensive list of selected documents is presented in Table 1.

Methods

The CAQDAS software [18] Leximancer v3.1 was used to identify concepts and themes across a range of documents and provide correlation analysis to some extent [19]. Using documents or texts as a source of data required the ability to analyze them while applying other skills, such as recognizing certain textual forms or genres, and the ability to judge, interpret, or contextualize [20, 21]. In the process of moving the documents from converting sets of

data into meaningful findings it was important to decide the intent for deriving and presenting data. The process involved obtaining documents from the official websites [7–16] and analyzing them in a comparative manner using the software. Concepts and themes were identified (conceptual analysis) in groups of documents, and the co-occurrence patterns between them (relational analysis) were explored across the entire data corpus.

RESULTS

Documents presented in Table 1 (cohorts 1 and 2) were analyzed by Leximancer v3.1 software content analysis, and the data are presented in Tables 2–4 and Figures 1–4. After processing groups of documents, a range of data was generated—concepts,

Table 2. Frequently occurring concepts in the software documentary analysis (cohorts 1 and 2)

Concepts in order of frequency	1	2	3	4	5
Documents in individual regulatory frameworks					
TGA	Biological	Human	Section	Trial	
FDA	Product	Section	Used	HCT	CFR
EMA	Medicinal	Marketing	Commission	Directive	Public
Documents across regulatory frameworks					
GMP (1)	Quality	Materials	Product	Contamination	Manufacture
GMP (2)	Appropriate	Drug	Records	API	
GCP	Clinical	Trial	Subjects	Sponsor	Data

Abbreviations: API, Active Pharmaceutical Ingredients; CFR, Code of Federal Regulations; EMA, European Medicines Agency; FDA, Food and Drug Administration; GCP, Good Clinical Practice; GMP, Good Manufacturing Practice; HCT, human cell and tissues; TGA, Therapeutic Goods Administration.

Table 3. Thematic summary in individual regulatory frameworks and across regulatory areas (cohorts 1 and 2)

TGA		FDA		EMA		GMP (1)		GMP (2)		GCP	
Theme	Connectivity (%) ^a	Theme	Connectivity (%)	Theme	Connectivity (%)	Theme	Connectivity (%)	Theme	Connectivity (%)	Theme	Connectivity (%)
Biological	100	Section	100	Medicinal	100	Quality	100	Appropriate	100	Sponsor	100
Class	46	Donor	82	Marketing	88	Product	91	Drug	50	Subjects	66
Use	44	Product	53	Referred	59	Materials	73	Use	45	Trial	56
Biologicals	42	HCT	49	Community	29	Products	63	Batch	40	Written	34
ARTG	34	Including	46	Directive	28	Used	46	Used	39	Clinical	28
Human	20	Manufacture	39	Market	15	Manufacture	36	Manufacture	28	Products	18
Medical	15	Products	21	Agency	12	Validation	26	Written	25	Product	15
Section	10	Requirements	19	Appropriate	10	System	26	Established	23	Adverse	03
Trial	02	FDA	11	Human	10	Contamination	19	API	22		
Page	01	Drug	05	Therapy	09	Area	18	Maintained	09		
				Commission	09	Containers	08	Apis	08		
						Filling	06	Personnel	04		
						Manufacture	03	Data	03		

In the full software interface, a list of concepts ranked by their frequency is displayed below the map.

^aConnectivity is a measure of co-occurrence between the concepts and the degree of association between them.

Abbreviations: API, Active Pharmaceutical Ingredients; ARTG, Australian Register of Therapeutic Goods; FDA, Food and Drug Administration; HCT; human cell and tissues.

Table 4. Frequency and connectivity of themes identified in individual regulatory frameworks and across regulatory areas (cohorts 1 and 2)

Themes connectivity ^a		Themes identified by the software							Themes overall frequency
Minimum (%)	Maximum (%)	TGA documents	FDA documents	EMA documents	GMP (1) documents	GMP (2) documents	GCP documents		
91	100	Biological	Section	Medicinal	Quality Product	Appropriate	Sponsor	High (9/65)	
81	90		Donor	Marketing					
51	80		Product	Referred	Materials Products		Subjects Trial	Medium (41/65)	
11	50	Class	HCT	Community	Used	Drug	Written		
		Use	Including	Directive	Manufacture	Use	Clinical		
		Biologicals	Manufacture	Market	Validation	Batch	Products		
		ARTG	Products	Agency	System	Used	Product		
		Human	Requirements		Contamination	Manufacture	Adverse		
		Medical	FDA		Area	Written			
≤10						Established			
		Section	Drug	Appropriate	Containers	Maintained		Low (15/65)	
		Trial		Human	Filling	Apis			
		Page		Therapy	Manufacture	Personnel			
				Commission		Data			

In the full software interface, a list of concepts ranked by their frequency is displayed below the map.

^aConnectivity is a measure of co-occurrence between the concepts and the degree of association between them.

Abbreviations: API, Active Pharmaceutical Ingredients; ARTG, Australian Register of Therapeutic Goods; EMA, European Medicines Agency; FDA, Food and Drug Administration; GCP, Good Clinical Practice; GMP, Good Manufacturing Practice; HCT, human cell and tissues; TGA, Therapeutic Goods Administration.

concept clouds, concept maps, themes, and thematic summaries. The data presented in the study appear in the form of the following figures and tables:

- Primary concept maps (Fig. 1) and concept clouds (Fig. 2)
- Centrality of each concept (Figs. 2 and 4)
- Concept co-occurrence (Fig. 3)
- Central themes (Fig. 4)
- Document summary (supplemental online data 1)
- Thematic summary (supplemental online data 2)
- Main concepts discussed in the document set (Table 2)

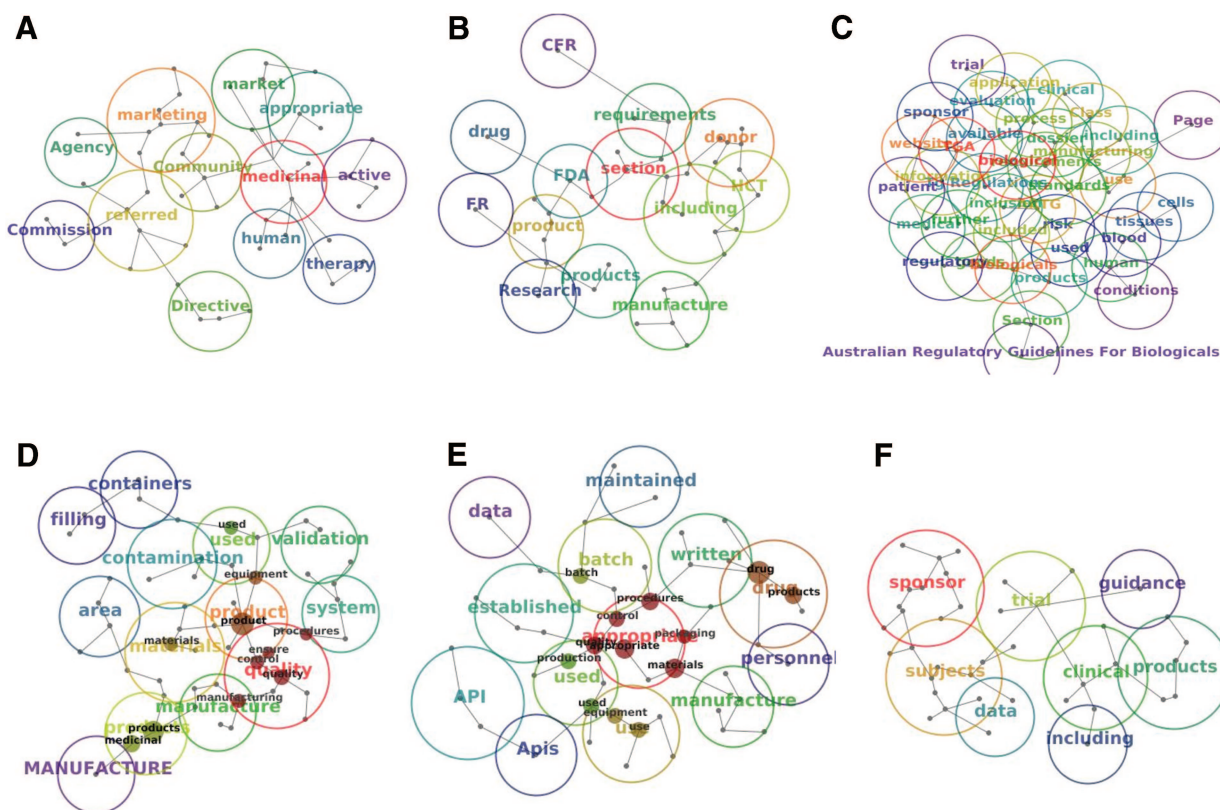


Figure 1. Concept maps. (A–C): Concepts identified in individual regulatory frameworks’ documents: European Medicines Agency documents (A), FDA documents (B), and TGA documents (C). (D–F): Concepts identified in documents across regulatory frameworks: Good Manufacturing Practice (GMP) (1) documents (D), GMP (2) documents (E), and Good Clinical Practice documents (F). For a list of the specific documents in each group, see Table 1. Abbreviations: API, Active Pharmaceutical Ingredients; CFR, Code of Federal Regulations; FDA, Food and Drug Administration; FR, Federal Regulations; TGA, Therapeutic Goods Administration.

- Similarity in contexts in which the concepts occur such as thematic groups (themes) and relative frequencies of the concepts (Table 3)
- Connectivity of themes (Table 4)

The concept maps appeared different with regard to their complexity, number of concepts, or size of the circles around them. Each map displayed important sources of information about the analyzed text, as demonstrated in Figures 1A–1F. Although difficult to read (because of the number of concepts identified by the software), the TGA concept map is presented in Figure 1C. It is an example of a set of documents discussing numerous topics, which are represented by the vastness of overlapping concepts.

The first Good Manufacturing Practice (GMP) concept cloud (Fig. 2A) appeared wider and busier, with the spread of numerous concepts surrounding the centrally positioned “product.” It encompassed analysis of diverse GMP documents from the EMA and the TGA regulatory frameworks, which is reflected in a diverse and broad range of concepts. The second GMP concept cloud (Fig. 2B) revolves around the central concept “appropriate” closely encircled with “procedures,” “processing,” “quality,” and “control.” Its wide circle reaches to “intended” and “use” on one side and “batch” and “records” on the other side. Overall, the second GMP concept cloud (Fig. 2B) appeared relatively narrow compared with the first GMP concept cloud (Fig. 2A). It covered analysis of two medicinal product GMP documents and the GMP code of the finished pharmaceutical prod-

ucts reflecting its coherence. The Good Clinical Practice (GCP) concept cloud (Figure 2C) was the most consistent, with a fairly narrow range of concepts surrounding “sponsor” and “trial” occupying the central positions. This clearly reflects the focus of all GCP documents revolving around one central concept. On the contrary, the EMA concept cloud (Fig. 2D) and TGA concept cloud (Fig. 2E) revolved around their vertical and horizontal axes, respectively, thus reflecting the diversity and wider spread of the concepts.

The concept clouds in Figure 2 are a useful illustration of differences between the documents’ content. For instance, the content of the documents in the EMA (Fig. 2D) and the TGA can be compared visually (Fig. 2E). The EMA concept cloud appears rather focused (‘narrower’ than the TGA cloud) with the central or primary concepts “medicinal” and “products” and with a number of secondary concepts, such as “marketing,” “authorization,” “holder,” and “referred” (Fig. 2D). The TGA concept cloud features the central concepts “biological” and “class,” with a wider spread of other concepts around them. These include “human” and “products” on one side and “information” and “application” on the other side of the TGA concept cloud (Fig. 2E). A similar observation is applicable in the comparison of GMP and GCP documents across different regulatory frameworks.

Different colors indicated different thematic groups (themes) surrounded by circles. They form around highly connected concepts, which are parents of thematic clusters [19, 22].

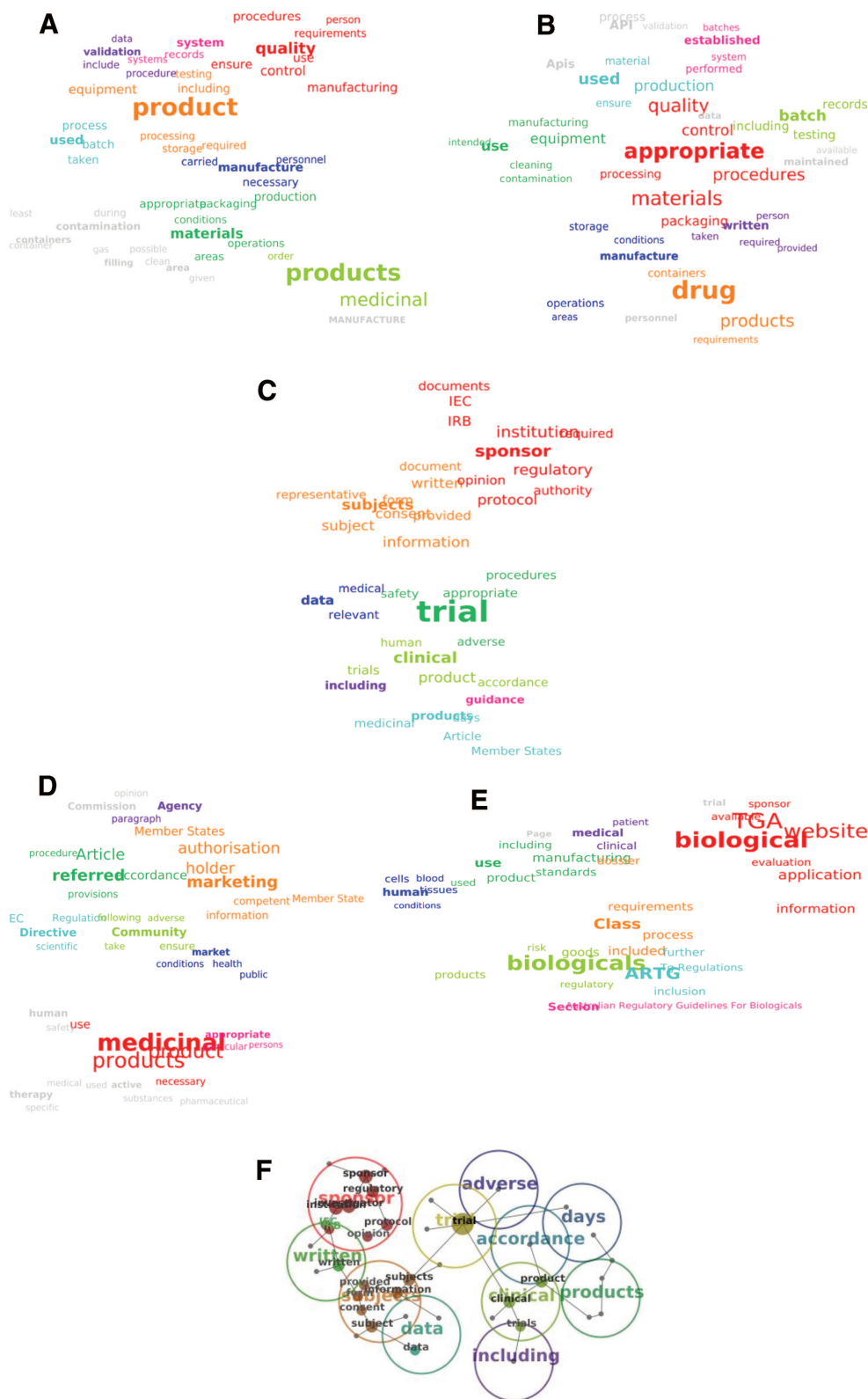


Figure 2. Concept clouds and themes. **(A–C):** Concept clouds identified by the software in documents across regulatory frameworks: Good Manufacturing Practice (GMP) (1) documents **(A)**, GMP (2) documents **(B)**, and Good Clinical Practice (GCP) documents **(C)**. **(D–F):** Concept clouds identified in individual regulatory frameworks' documents **(D, E)** and overlay of themes **(F)** identified by the software across regulatory frameworks in GCP documents: European Medicines Agency documents **(D)**, TGA documents **(E)**, and GCP documents related concepts and themes (overlay of words is a reflection of the concepts' closeness) **(F)**. For a list of the specific documents in each group, see Table 1. Abbreviations: API, Active Pharmaceutical Ingredients; ARTG, Australian Register of Therapeutic Goods; EC, Ethics Committee; IEC, Institutional (Independent) Ethics Committee; IRB, Institutional Review Board; TGA, Therapeutic Goods Administration.

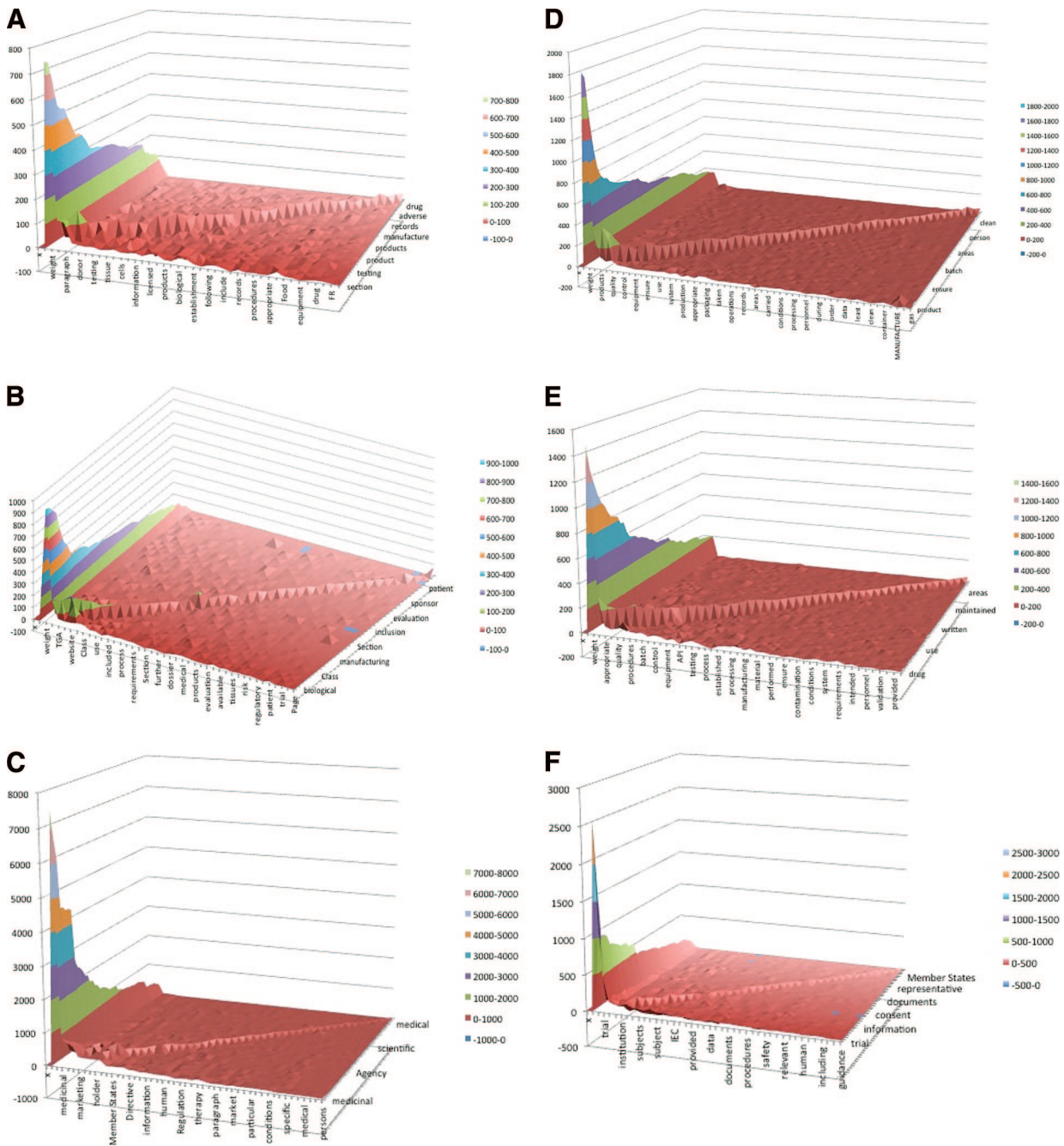


Figure 3. Correlation of concepts across data corpus. Pairs of concepts correlation (co-occurrence) identified by the software across the entire data corpus for the Food and Drug Administration documents (A), Therapeutic Goods Administration documents (B), European Medicines Agency (C), Good Manufacturing Practice (GMP) (1) (D) and GMP (2) documents (E), and Good Clinical Practice document analysis (F). Colored slices on the graph present relative frequencies of the key co-occurring concepts. For a list of the specific documents in each group, see Table 1. Abbreviations: API, Active Pharmaceutical Ingredients; FR, Federal Regulations; IEC, Institutional Ethics Committee; TGA, Therapeutic Goods Administration.

For example, the map in Figure 2F presents analysis of GCP documents across different regulatory frameworks with a similar “separateness” of concepts and their specific connectivity—represented not only by color but also by the distance between them. The GCP documents displayed considerably higher concordance between different regulatory frameworks than other analyzed documents. The central theme “trial” was closely surrounded by the themes “sponsor,” “subjects,” “data,” “clinical,”

“accordance,” and “adverse.” Themes such as “written,” “including,” and “days” were reasonably remote but still linked to the central theme “trial.” The theme “sponsor” consisted of the concepts “sponsor,” “regulatory,” “protocol,” and “opinion,” and the theme “subjects” contained concepts such as “provided,” “information,” and “consent” closely linked to the theme “data” (and the concept “data”). The themes “clinical” and “products” were also in close proximity—demonstrating their relatedness

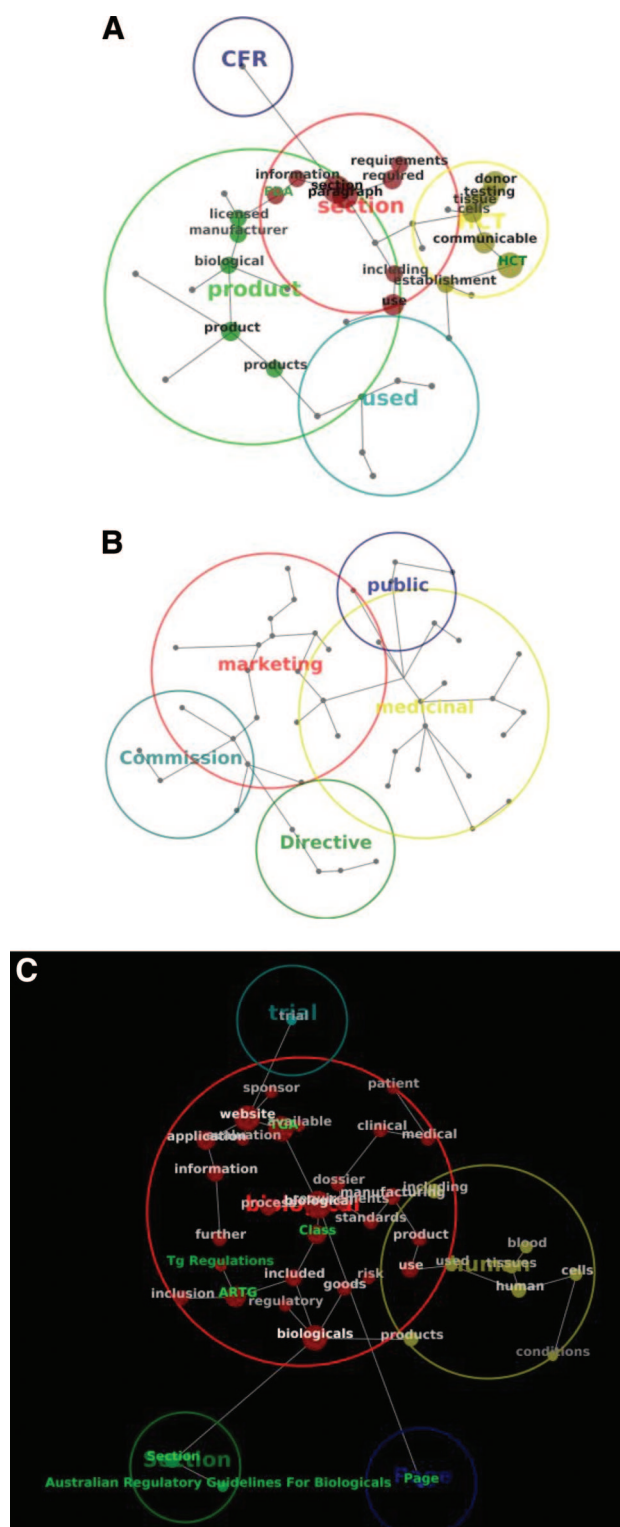


Figure 4. Most frequent themes and concepts. Most frequent themes and their concepts identified by the software in FDA documents (overlay of words is a reflection of concepts' closeness) (A), European Medicines Agency documents (B), and Therapeutic Goods Administration documents (presented here as an example of a strong contextual similarity) (C). For a list of the specific documents in each group, see Table 1. Abbreviations: ARTG, Australian Register of Therapeutic Goods; CFR, Code of Federal Regulations; FDA, Food and Drug Administration; HCT, human cell and tissues.

in the documents but slightly more remote from the themes “sponsor,” “written,” or “data,” as these did not occur near each other in the documents. The example of the close co-occurrence and remoteness of the themes is obvious in the following excerpts from the GCP documents: “(a) the informed consent of the legal representative has been obtained; consent must represent the subject’s presumed will and may be revoked at any time, without detriment to the subject; ...” (Directive 2003/20/EC [EMA]) and, “In obtaining and documenting informed consent, the investigator should comply with the applicable regulatory requirement(s), and should adhere to GCP and to the ethical principles that have their origin in the Declaration of Helsinki” (Guidance for Industry E6 Good Clinical Practice [FDA]).

Furthermore, the software automatically extracted the most important concepts discussed in the documents being processed. In the maps of the analyzed documents, the five most frequent concepts were very different, as shown in Table 2.

The major themes in the TGA, FDA, and EMA regulatory frameworks and the same group documents (i.e., GMP and GCP) across different regulatory frameworks are presented as thematic summaries in Table 3. The thematic summaries of analyzed documents mostly consist of a broad range of themes and their frequency significantly varies. However, some common ground was found between them. For example, in two groups of documents, GMP (1) and GMP (2), 8 of top 12 concepts (66.66%) were exactly the same (“product(s),” “quality,” “materials,” “control,” “used,” “equipment,” “procedures,” and “use”), and 2 of 12 concepts (16.66%) were interchangeable (“medicinal”/“drug” and “manufacturing”/“production”). Hence, 10 of the 12 most frequently co-occurring concepts (83.33%) were shared between all GMP documents in the analysis, irrespective of the frameworks from which they originate, whereas only 2 of the most frequent concepts were entirely different (“appropriate” and “batch”). Further analysis presented in Table 4 shows connectivity of themes as a measure of connectedness between them. If all the themes across regulatory frameworks and regulatory areas were grouped according to their connectivity range, a high range (81%–100%) of connectivity was found in only 9 themes (of 65 themes in total), whereas a majority of themes convened within a medium connectivity range (11%–80%). Sometimes a theme with a relatively high frequency was not connected to other themes, and hence its connectivity range was low (i.e., “drug” in the FDA regulatory documents).

The co-occurrence between concepts in the text was an important measure of the degree of association between them. Hence, the three-dimensional (3D) graphs in Figure 3 are presented to illustrate the co-occurrence of the concepts and proximity analysis in individual frameworks’ documents (FDA, TGA, EMA), as well as across the set of documents applicable to GMP and GCP requirements. Proximity analysis measured the co-occurrence of concepts found within the text. In this approach a window, based on a length of words or sentences, was specified and moved sequentially through the text, taking note of the concepts that occurred together. In the EMA documents the co-occurrence of pairs of concepts was in the maximum frequency range of 7,000–8,000 (Fig. 3C), whereas the FDA and the TGA documents had maximum concept co-occurrence of up to 1,000 (Fig. 3A and 3B, respectively). This measure depends not only on individual frequency of the concepts but also on the co-occurrence of the concepts (their connectedness) [19, 22]. For instance, the individual frequency of the concept “medicinal” was

greater than 1,000 in three EMA documents analyzed manually. Nevertheless, because of their connectedness, the total of their frequency presented on the co-occurrence graph generated by the software analysis was in the range of 7,000–8,000 (Fig. 3C).

A group of concepts, which appear in the text close to each other, was identified as part of the same theme. The most central themes (and concepts within those) are presented in Figure 4. The central themes encompassed “product,” “medicinal,” and “biological” in FDA, EMA, and TGA frameworks, respectively, and the central concepts contained within the central themes ranged from “biological,” “licensed,” and “manufacturer” (FDA) to “dossier,” “goods,” and “clinical” (TGA). In some instances, central themes included the abbreviated name of a regulatory registry (i.e., “ARTG”), reference to a legal entity (i.e., “Commission”), or reference to a legal document (i.e., “Directive” or “CFR”). Representation of central themes in Figure 4 to some extent reflects the main focus of regulatory documents scrutinized in this study. It appeared that the FDA documents revolve around themes “HCT” (human cell and tissues) or “product” with frequent references to “CFR” and “section,” whereas the EMA documents were mostly dominated by “medicinal” (product), “marketing,” and “public,” with numerous citations of “Commission” and “Directive.” Interestingly, the TGA document analysis demonstrated only two main central themes, “biological” (with numerous concepts) and “human”; its reference is mostly to “section” and “Australian Regulatory Guidelines for Biologicals” (mentioned as one of the concepts within the central themes).

The last block of information visible from the map was the contextual similarity between concepts. The map was constructed by initially placing the concepts randomly on the grid. Each concept exerted a pull on each other concept with a strength related to their co-occurrence value. Concepts can be thought of as being connected to each other with springs of various lengths [19, 22]. The more frequently two concepts co-occur, the stronger are the forces of attraction that bring frequently co-occurring concepts together [19, 22]. However, because there are many forces of attraction acting on each concept, it is impossible to create a map in which every concept is at the expected distance away from every other concept. Rather, concepts with similar attractions to all other concepts will cluster together [19, 22]. If regulatory agencies talk about similar issues, these issues (or concepts) will appear closer together than those of different agencies with different agendas or different intentions.

An example of contextual similarity was evident in the TGA documents outlining the Biologicals Framework, where the most frequent co-occurring themes (and concepts within those themes) are presented in Figure 4C. Although the concepts “application” and “information” do not have a direct link in the text (they never appear together), they appeared adjacent on the map due to the similarities of the contexts (i.e., themes) in which they appear. Furthermore, these words appeared clustered close together with other significant words, such as “dossier,” “processing,” and “goods,” that appear in similar contexts in the documents. A thematic grouping of “blood,” “human,” “cells,” and “product” (or “products”) was also apparent. An entirely separate domain on the map presented the “Australian Regulatory Guidelines for Biologicals” and “section” drawn together by the similarities of the contexts in which they appear (but distant from the other concepts on the map).

In addition to a document summary in supplemental online data 1, the software revealed differences in the themes and their frequencies in the form of a thematic summary. An example of the thematic summary is presented in supplemental online data 2A. It demonstrated the statistical breakdown of themes with 100% frequency assigned to the most frequently occurring theme, i.e., “medicinal” in the EMA documents. The frequency of other themes’ appearance (i.e., “marketing,” “human,” and “therapy”) or references (i.e., “Community,” “Directive,” and “Agency”) was expressed as a percentage of the number using this leading theme frequency as 100% (frequency assigned to “medicinal”). The wording behind each theme is presented in the thematic summary shown in supplemental online data 2B. It includes explanation of the context of a theme (e.g., for “human,” the context example is: “Substance: Any matter irrespective of origin which may be: human, e.g., human blood and human blood products”), as well as the number of occurrences (“hits”) for each theme.

DISCUSSION

Differences and Similarities Among the Three Regulatory Frameworks Analyzed

The conceptual map of documents provided three sources of information about the content of documents: (a) main concepts contained within the text and their relative importance, (2) strengths of links between concepts (how often they co-occur), and (c) similarities in the context in which they occur. This information provided a powerful metric for statistically comparing different documents. Arguably, this may be interpreted as a representation of the document’s intent or purpose. Furthermore, this map can be used to simply provide a bird’s eye view of a large amount of material such as compliance with multiple regulatory frameworks’ requirements.

The document summary provides blocks of sentences containing the most important concepts that appear on the map. They are crucial segments of text demonstrating the relationship between the defined concepts. Parts of the document summaries are presented in supplemental online data 1, as an illustration of this type of analysis.

We found evidence of some shared aspects and terminology within analyzed documents in this study. The 83.33% of most frequently co-occurring concepts were shared between all GMP documents in the analysis, irrespective of the frameworks from which they originated. The GCP documents displayed considerably higher concordance between different regulatory frameworks than other analyzed documents. The GCP concept cloud was the most consistent, with a fairly narrow range of concepts surrounding “sponsor” and “trial” occupying the central positions. This clearly reflects the focus of all GCP documents revolving around one central concept regardless of the regulatory framework from which they originated. The findings also suggested that the most frequent concepts in the TGA, FDA and EMA frameworks were “biological,” “product,” and “medicinal,” respectively—all denoting a product as their focus point.

However, we also found a range of differences. For example, it appeared that the FDA documents revolve around the themes “HCT” (human cell and tissues) or “product,” with frequent references to “CFR” and “section,” whereas the EMA documents are mostly dominated by “medicinal,” “marketing,” and “public,” with

numerous citations of “Commission” and “Directive.” The TGA documents analysis has demonstrated only two main central themes: “biological” (with numerous concepts) and “human.”

The lack of similarity found in both manual and software analysis (i.e., different terminology, use of different phrases and the difference in their co-occurrence frequency) supports part of the hypothesis relating to differences between the frameworks. It also demonstrates that formal terminology differences exist widely throughout the regulatory frameworks, manifested in their high-tier documents.

The complexity of the global regulatory environment revealed in this study advocates application of a regulatory engineering approach that would allow for a whole array of regulatory requirements to be met in more than one regulatory framework. Similar to process engineering, the purpose of regulatory engineering would be to provide a solid setting for the development of biologic drugs (which carries a significant amount of uncertainty in its own intrinsic element). Development of new biologic drugs and biopharmaceuticals (i.e., cell and tissue products) has led the existing regulatory agencies of the developed world to propound new regulatory approaches—it “has become one of the most research-intensive sectors with a great potential for delivering innovative human medicines in the future” [3]. Regulation of cell and tissue products is closely linked not only to sensitive areas of public health policy and funding but also to the development of novel disciplines such as regulatory science [4, 5]. In view of the harmonization efforts between the analyzed regulatory frameworks, this paper represents a contribution to the recognition of similarities and differences between them. It may also fill the literature gap and provide some foundation for future comparative research of biologic drug regulations on a global level.

Methodology and Software Used in the Study

Qualitative research often deals with in-depth analysis, fine distinction, and complexity; it is also based upon values of reason, truth, and validity, though it can be context-contingent to some degree [46, 47]. However, it is the researcher who is the research instrument in a qualitative study. This means that qualitative research cannot entirely rely on standardized procedures to deal with concepts such as bias and reproducibility [48]. Hence, researchers are required to reflect constantly and critically on the decisions they make during the course of a study, to maintain transparency about the actual research design, and to keep an audit trail regarding access to and collection and analysis of the data [49]. For Leximancer software analysis, concepts are defined as “collections of words that generally travel together throughout the text” [19]. The presence of each word in a sentence provides an appropriate contribution to the accumulated evidence for the presence of a concept [22]. It means that a sentence (or group of sentences) is tagged as a concept only if it contains enough accumulated evidence—if the sum of the keyword loads is above the threshold set in the software [22]. The software conducted two forms of analysis. It measured the presence of defined concepts in the text (conceptual analysis) and how they were interrelated (relational or semantic analysis) (Figs. 1, 2). Apart from allowing these differences to be explored visually, the software explicitly stored the data relating to the frequency of concept use (main concept frequencies represented in Table 2) and frequency of concept co-occurrence in a spreadsheet

file (represented in the form of 3D graphs in Fig. 3). Concepts can either be explicit (i.e., particular words or phrases) or implicit (i.e., implied, but not explicitly stated in a set of pre-defined terms) [23]. Leximancer can automatically extract its own dictionary of terms for each document set while using this information. It is capable of assuming that the concept classes that are present within the text are correct, unambiguously extracting a vocabulary of terms for each concept [19, 22]. There are commonly two applicable forms of reliability in content analysis: stability and reproducibility. Stability refers to the tendency of a coder (or a system) to consistently recode the same information in the same way over time [23]. In the automatic coding system (the software approach) many such inconsistencies are avoided, leading to a high level of coding stability [22]. Reproducibility refers to the consistency of classification provided by several coders within the same marking scheme [23]. In the software, this issue is most pertinent to the generation of the conceptual map. The process of map generation is stochastic, leading to the possibility of different final positions for the extracted concepts each time the map is generated [19]. In order to maintain the maximal reproducibility of the software, which also affects its validity, we chose not to change any of the software’s parameters and to use only minimal manual intervention while performing the software analysis. In some other qualitative content analysis tools (e.g., NVivo; QSR International Pty. Ltd., Doncaster, VIC, Australia, <http://www.qsrinternational.com>) it is required that the analyst derive the list of codes and rules for attaching them to the data [30]. This is not a prerequisite for the Leximancer application, hence eliminating the need for reliability and validity checks. Leximancer, as a data-mining tool, uses Bayesian logic to identify and visually present main concepts within the text [45], and it is considered to be highly reliable, stable, and valid in different applications [29, 33]. Relational (semantic) analysis measures how identified concepts are related to each other within the documents. Two types of relational analysis were used in this study in order to measure the relationships between the identified concepts. Proximity analysis looks into co-occurrence of concepts found within the text and cognitive mapping presents the information visually for comparison.

CONCLUSION

This paper has described the processes that were obtained while developing a method of software documentary analysis for higher-tier regulatory documents. The results obtained in software analysis are in concordance with a manual analysis obtained [17], and both sets of results are comparable. The research did not provide any evidence about superiority either a manual or a software-based approach, since they produced qualitatively different sets of data (i.e., terminology and phrase frequency vs. contextual analysis). The weaknesses of this type of study need to be acknowledged. Data are studied out of context, and the selectiveness of the data or representativeness of the sample can be questioned. One of the main areas of criticism is the fact that the whole study process is open to the subjective interpretation of the researcher. Despite this criticism, this approach to documentary analysis is particularly useful when the researcher is faced with a vast number of complex documents with no common format, scope, or terminology.

Our findings indicate that some common ground was identified between the documents originating from different regulatory frameworks. This was consistent with the manual terminology search of the same documents [17]. There was no evidence in the co-occurrence analysis to establish any shared conceptual vocabulary between the regulatory frameworks.

As pointed out in the first part of the study [17], this research draws attention to the importance of understanding the complexity of a global regulatory environment attained by three regional regulatory agencies. Although a vast array of differences was identified in terms of terminology, concepts, and themes, some shared aspects and common principles were also identified between the analyzed regulatory frameworks.

REFERENCES

- Halme DG, Kessler DA. FDA regulation of stem-cell-based therapies. *N Engl J Med* 2006; 355:1730–1735.
- FDA Guidance for industry: PAT—A framework for innovative pharmaceutical development, manufacturing, and quality assurance. 2004. Available at <http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm070305.pdf>. Accessed January 13, 2011.
- The Financing of Biopharmaceutical Product Development in Europe. Final Report. European Commission, Enterprise and Industry. October 2009. Available at http://ec.europa.eu/enterprise/sectors/biotechnology/files/docs/financing_biopharma_product_dev_en.pdf. Accessed December 5, 2011.
- Advancing Regulatory Science at FDA, Strategic Plan: August 2011. Food and Drug Administration, U.S. Department of Health and Human Services. Available at <http://www.fda.gov/regulatoryscience>. Accessed December 5, 2011.
- Implementing the European Medicines Agency's Road Map to 2015: The Agency's Contribution to Science, Medicine, Health—From Vision to Reality. EMA/MB/550544/2011. October 6, 2011. Available at <http://www.ema.europa.eu>. Accessed November 22, 2011.
- Burger SR. Current regulatory issues in cell and tissue therapy. *Cytotherapy* 2003;5:289–298.
- TGA basics. Therapeutic Goods Administration. Available at <http://www.tga.gov.au/about/tga.htm>. Accessed September 1, 2011.
- Clinical trials. Therapeutic Goods Administration. Available at <http://www.tga.gov.au/industry/clinical-trials.htm>. Accessed September 1, 2011.
- Clinical trials at a glance. Therapeutic Goods Administration. Available at <http://www.tga.gov.au/industry/clinical-trials-glance.htm>. Accessed September 1, 2011.
- Note for guidance on good clinical practice. Therapeutic Goods Administration. Available at <http://www.tga.gov.au/industry/clinical-trials-note-ich13595.htm>. Accessed August 6, 2011.
- European Medicines Agency. Available at <http://www.ema.europa.eu/ema>. Accessed October 15, 2011.
- About us. European Medicines Agency. Available at http://www.ema.europa.eu/ema/index.jsp?curl=pages/about_us/general/general_content_000235.jsp&murl=menus/about_us/about_us.jsp. Accessed October 15, 2011.
- Food and Drug Administration. Available at <http://www.fda.gov>. Accessed October 27, 2011.
- Development & Approval Process (CBER). Center for Biologics Evaluation and Research, Food and Drug Administration. Available at <http://www.fda.gov/BiologicsBloodVaccines/DevelopmentApprovalProcess/default.htm>. Accessed July 20, 2011.
- The Center for Drug Evaluation and Research (CDER), Food and Drug Administration. Guidance, compliance, and regulatory information. Available at <http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/default.htm>. Accessed July 20, 2011.
- New Drug Application (NDA) Process (CBER). Center for Biologics Evaluation and Research, Food and Drug Administration. Available at <http://www.fda.gov/BiologicsBloodVaccines/DevelopmentApprovalProcess/NewDrugApplicationNDAProcess/default.htm>. Accessed July 21, 2011.
- Ilic N, Savic S, Siegel E et al. Examination of the regulatory frameworks applicable to biologic drugs (including stem cells and their progeny) in Europe, the U.S., and Australia: Part I—A method of manual documentary analysis. *STEM CELLS TRANSLATIONAL MEDICINE* 2012; 1:898–908.
- Mason J. *Qualitative Researching*. 2nd ed. London, UK: Sage Publications, 2002.
- Smith AE. Leximancer tutorials. Available at <https://www.leximancer.com/support/>. Accessed August 2, 2011.
- Bailey KD. *Methods of Social Research*. 2nd ed. New York, NY: Free Press, 1982.
- Robson C. *Real World Research*. 2nd ed. Oxford, U.K.: Blackwell, 2002.
- Smith AE. *Leximancer Manual* (Version 2.2) 2005. Available at https://www.leximancer.com/wiki/images/7/77/Leximancer_V2_Manual.pdf. Accessed August 2, 2011.
- Weber RP. *Basic Content Analysis*. 2nd ed. Thousand Oaks, CA: Sage Publications, 1990.
- Cabr  Castellv  MT. Theories of terminology: Their description, prescription and explanation. *Terminology* 2003;9:163–199.
- Meyer I, Eck K, Skuce D. Systematic concept analysis within a knowledge-based approach to terminology. In: Wright SE, Buding G, eds. *Handbook of Terminology Management*. Amsterdam, The Netherlands: John Benjamins, 1997:110.
- Felber H. *Terminology Manual*. Paris: General Information Programme & UNISIST, UNESCO. Paris, France: Infoterm, 1984.
- Kageura K. *The Dynamics of Terminology: A Descriptive Theory of Term Formation and Terminological Growth*. Amsterdam, The Netherlands: John Benjamins, 2002.
- Ilic N, Khalil D, Hancock S et al. Regulatory considerations applicable to manufacturing of human placenta-derived mesenchymal stromal cells (MSC) used in clinical trials in Australia and comparison to USA and European regulatory frameworks. *Mesenchymal Stem Cell Therapy*. Springer/Humana Press (in press).
- Smith AE, Humphreys MS. Evaluation of unsupervised semantic mapping of natural language with Leximancer concept mapping. *Behav Res Methods* 2006;38:262–279.
- Cretchley J, Gallois C, Chenery H et al. Conversations between carers and people with schizophrenia: A qualitative analysis using Leximancer. *Qual Health Res* 2010;20:1611–1628.
- McKenna B, Waddell N. Media-ted political oratory following terrorist events. *J Lang Pol* 2007;6:377–399.
- Martin NJ, Rice JL. Profiling enterprise risks in large computer companies using the Leximancer software tool. *Risk Manag* 2007;9:188–206.
- Scott N, Smith AE. Use of automated content analysis techniques for event image assessment. *Tourism Recreation Res* 2005;30:87–91.
- Grech MR, Horberry T, Smith AE. Human error in maritime operations: Analyzes of accidents reports using the Leximancer tool. Presented at the Human Factor and Ergonomics Society 46th Annual Meeting, Baltimore, MD, 2002.
- Loke K, Bartlett B. What distinguishes a distinction? Perceptions of quality academic performance and strategies for achieving it. In: *Reimagining Practice: Researching Change*, Vol. 2. Conference Proceedings of the International Conference on Cognition, Language, and Special Education. Nathan, QLD, Australia: Griffith University, 2003; 175–182.
- Rooney D, Gallois C, Cretchley J. Mapping a 40-year history with Leximancer:

AUTHOR CONTRIBUTIONS

N.I.: conception and design, collection and assembly of data, data analysis and interpretation, manuscript writing; S.S., E.S., and K.A.: final approval of manuscript; L.T.: conception and design, final approval of manuscript.

DISCLOSURE OF POTENTIAL CONFLICTS OF INTEREST

E.S. holds a compensated employment/leadership in Ground Zero Pharmaceuticals, Inc.

Themes and concepts in the *Journal of Cross-Cultural Psychology*. *J Cross-Cultural Psychol* 2010;41:318–328.

37 Baker SC, Gallois C, Driedger SM et al. Communication accommodation and managing musculoskeletal disorders: doctors' and patients' perspectives. *Health Commun* 2011;26:379–388.

38 Kyle GJ, Nissen L, Tett S. Perception of prescription medicine sample packs among Australian professional, government, industry, and consumer organizations, based on automated textual analysis of one-on-one interviews. *Clin Ther* 2008;30:2461–2473.

39 Travaglia JF, Westbrook MT, Braithwaite J. Implementation of a patient safety incident management system as viewed by doctors, nurses and allied health professionals. *Health (London)* 2009;13:277–296.

40 Hepworth N, Paxton, SJ. Pathways to help-seeking in bulimia nervosa, binge eating

problems: A concept mapping approach. *Int J Eat Disord* 2007;40:493–504.

41 McGivern G, Fischer MD. Reactivity and reactions to regulatory transparency in medicine, psychotherapy and counselling. *Soc Sci Med* 2012;74:289–296.

42 Noble C, O'Brien M, Coombes I et al. Concept mapping to evaluate an undergraduate pharmacy curriculum. *Am J Pharm Educ* 2011;75:55.

43 Watson M, Smith A, Watter S. Leximancer concept mapping of patient case studies. In: Khosla R, ed. *Knowledge-Based Intelligent Information and Engineering Systems*. Lecture Notes in Computer Science, Vol. 3683. Springer, 2005;1232–1238.

44 Grimbeek P, Bartlett B, Loke K. Using Leximancer to identify themes and patterns in the talk of three high-distinction students. In: Bartlett B, Bryer F, Roebuck D, eds. *Second Annual Conference on Cognition, Language, and Special Education Research*: Vol.

2: *Educating: Weaving Research into Practice*. Brisbane, QLD, Australia: Griffith University, 2004;122–128.

45 Zuell C, Harkness J, Juergen HP et al., eds. *Computer-Assisted Text Analysis for the Social Sciences: The General Inquirer III*. Mannheim, Germany: Center for Surveys, Methods and Analysis (ZUMA), 1989.

46 Miles MB, Huberman AM. *Qualitative Data Analysis*. 2nd ed. Thousand Oaks, CA: Sage Publications, 1994.

47 Patton MQ. *Qualitative Research & Evaluation Methods*. 3rd ed. Thousand Oaks, CA: Sage Publications, 2002.

48 Creswell JW. *Qualitative Inquiry & Research*. 2nd ed. Thousand Oaks, CA: Sage Publications, 2007.

49 Maxwell JA. *Qualitative Research Design*. 2nd ed. Thousand Oaks, CA: Sage Publications, 2005.



See www.StemCellsTM.com for supporting information available online.

See the accompanying article on pages 898–908 of this issue.