The Way Forward in Package Insert User Tests From a CRO's Perspective

Introduction: A significant increase in the amount of text in package inserts has been observed over the last years. This study investigated the consequences of this increase. Method: Five package inserts available on the German medicine market in 2002 and five developed model versions were investigated in a crossover procedure using the written readability test.

Results: The more extensive the package inserts, the worse patients feel informed. Increasing the amount of text significantly decreases the ability to locate information, thus putting people off from reading the contents ($P \le 0.021$).

Discussion: The results suggest that decreasing the amount of text is a key factor, whereby a maximum of 1,500 words per package insert should be the aim. Conclusion: The way forward in package insert user testing is to concentrate on patient requirements and package insert improvements. Appropriate solutions are required for the further development of guidelines, templates, and directives.

Joerg Fuchs

PAINT-Consult, Jena,

University of Bonn, Bonn, Germany

Germany; Department of

Drug Regulatory Affairs at

the Institute of Pharmacy,

Key Words User test; Readability test; Package insert; Amount of text; Guidelines

Correspondence Address

Dr. Joerg Fuchs, PAINT-Consult, Friedrich-Engels Str. 19, 07749 Jena, Germany (email: joerg.fuchs@ paint-consult.com).

Presented at the Joint DIA-EMEA-CMD(h) Workshop on User Testing, December 5, 2008, Devonport House, London.

INTRODUCTION

Since Directive 2004/27/EC came into force, package inserts of medicines that are sold in the European Union require user tests to ensure that they are legible, clear, and easy to use (1). One possible way to carry out these tests is the verbal face-to-face interview method, developed by the Australian communications researchers Sless et al. (2,3). Using this method, 12 to 15 questions concerning the package insert key messages are posed orally by an interviewer to a minimum of 20 medical laymen divided into two rounds.

Other methods can be used if they comply with the guidelines and if they are able to detect problems in locating and understanding information (2). One example is the written readability test, which was developed and validated in the PAINT1 survey. In this method the test participants are given the instructions and a minimum of 15 questions regarding the package insert key messages using a questionnaire (4).

Different stakeholders are involved in package inserts and their associated user tests, although patients form the key group in Europe. Patients should be able to locate, understand, and use the information provided in the package inserts. Other interested parties are the pharmaceutical companies, who must create package inserts that comply with the laws and guidelines. Beyond these groups, there are the national and supranational agencies that monitor to ensure package inserts and their user tests comply with legal requirements. Therefore, the task of a contract research organization (CRO) is to support the pharmaceutical companies to develop and test package inserts that meet the requirements of each stakeholder group. This means that compromises have to be found, which cannot always result in the most appropriate package insert for every patient.

In comparison to European package inserts, which are delivered to patients in the packaging of each over-the-counter and prescribed medicine, package inserts in the United States—also called prescription drug labels—are primarily written to give health care professionals the information they need to appropriately prescribe medicines (5). However, both the European and US package inserts have grown over the years in length, detail, and complexity. Therefore, appropriate improvements and measures, such as user tests, are essential (4–9).

Currently, there are no published results available regarding what can be achieved with the package insert user tests required in the European Union. Table 1 provides optimizations of four selected aspects that one CRO company has achieved using the written readability test Package Insert Optimizations Through the Most Recent User Tests Carried Out by PAINT-Consult Using the Written Readability Test (N = 40, English and German Package Inserts)

T 1					
	N I	K I		-	н
	4 1	ו ט	L	L	н
	-	_	_		-

Aspect		Original Package Inserts (n)	Final Package Inserts (n)	Difference (%)
Difficult words	Average	86	14*	84
	Minimum	4	3*	25
	Maximum	426	15*	96
Abbreviations	Average	17	4	76
	Minimum	4	2	50
	Maximum	61	5	92
Long sentences (over 20 words per sentence)	Average	29	7	76
	Minimum	4	2	50
	Maximum	140	13	91
Amount of text (number of words)	Average	2,505	2,002	20
	Minimum	841	834	1
	Maximum	6,777	3,758	45
*All difficult words were explained.				

method. On average, the number of difficult words in the first package inserts provided by the pharmaceutical companies are reduced, from 86 to 14 in the final tested versions. This is a difference of 84% (10). In addition, the 14 remaining difficult words are always explained.

There are maximum and minimum results. The minimums mostly occur with long-term clients, who have learned from the user test company's recommendations and test results to provide better package inserts. This is a clear step forward through user testing.

Another aspect is the amount of text. On average, the number of words contained in package inserts was reduced by 20% (Table 1) (10). One survey with patients and another with medical and pharmaceutical experts showed that this text reduction meets patients' and experts' requirements, as both want the text of package inserts to be shorter and limited to only essential points (11). However, the current situation indicates there has been a large text increase over the last years, due to new directives, templates, and guideline demands and new knowledge about the medicines. Therefore, a subanalysis of the PAINT1 survey results was done to investigate the advantages and disadvantages of this text increase.

MATERIALS AND METHODS

The PAINT1 study was a written readability test using a questionnaire consisting of 15 questions relating to the package inserts' key messages and 17 questions concerning the participants' opinions.

Five original package inserts, available on the German medicine market, and five previously developed model versions were investigated from September 2002 to April 2003. The models contained the same information as the originals; however, they were optimized using a set of quality criteria such that the amount of text was reduced to one page of A4 paper, printed on both sides. Furthermore, the models contained an optimized design and a larger font size of 11 points (4).

The study was a crossover comprehensibility test in two rounds with a break between each round of a minimum of 4 weeks. In the first round, half of the participants received an original and the other half a model version. The versions were swapped in the second round so that, at the end of this study, every person had tested an original and the corresponding model package insert (4).

Percentages for information not found and incorrectly comprehended information were determined for the total of tested key messages of each package insert. Furthermore, the median of the time required to locate all 15 answers was calculated per package insert.

In the section that addressed the participants' personal opinions of the package inserts, each participant used a five-point scale to assess the comprehensibility, legibility, complexity of information, clarity and structure, and their confidence in the described medicine. The participants' answers were coded as follows: "yes" = 1, "mostly yes" = 2, "other" = 3, "mostly no" = 4, and "no" = 5. Medians were calculated for each of the 17 questions relating to the personal opinions.

Afterward, Pearson's correlation coefficients were calculated using the SPSS 14.0 statistics software program between:

• The package insert specifics, such as the amount of text on one side

• The percentages of information not located, incorrectly comprehended information, the time needed to locate the information, and the personal opinions on the other side

Relationships between participants' demographic background and their opinions relating to the package inserts were calculated using the Cramer V calculation procedure from the SPSS 14.0 statistics program for both package insert groups—originals and models. In addition, for testing differences, Pearson's chi-square test was used. Afterward, Pearson's correlation coefficients were also calculated for each package insert.

RESULTS

In the first round, 1,105 people participated, and of these 1,057 took part in the second test round (age 10–92 years; average 38 years) (4).

Figures 1 and 2 illustrate that the participants located the information in each model package insert significantly better than in the corresponding original version. There was a very high correlation between locatability in the original package inserts and the amount of text they contained. An increase in the number of words in the originals led to a significant decrease in



FIGURE 1

Relationship between the amount of text in package inserts and the percentage of not located information from all 15 tested key messages. O, original; M, model; 1, Enalapril; 2, Ibuprofen; 3, Paracetamol (acetaminophen); 4, Repaglinide; 5, Telmisartan.

FIGURE 2

Relationship between the amount of text in package inserts and the time needed to locate all 15 tested key messages (17). Definitions as in Figure 1.



ability to locate the contents (P = 0.019), with patients requiring significantly more time to find the information (P = 0.006).

A high correlation was also found in the model versions between locatability and the amount of text. However, this was not significant.

Furthermore, there was no general relationship between the comprehensibility and the amount of text; long texts, such as from the ibuprofen original, were also well comprehended (Figure 3).

In addition, relationships in the group of the originals were found as follows. The locatability of the tested key messages decreased in relation to increases in the percentage of:

- Nonquantifiable phrases per total number of words. Words such as "longtime use" are nonquantifiable phrases. They do not enable the reader to clearly rate the importance of the information being communicated. For example, "longtime use" could be interpreted as either a period lasting at least 1 month or a period lasting up to 1 year or more (12).
- Difficult words per total number of words.
- Sentences longer than 20 words per total number of sentences.

These findings were valid for the percentage of located information and for the time needed to

find the contents. However, these six relationships were not significant, although their correlation coefficients were between 0.509 and 0.819.

Furthermore, significantly more participants stated that the information was easy to locate if:

- They needed less time to locate the requested information (*P* = 0.001).
- They found more of the 15 tested key messages (P = 0.039).

With the exception of the amount of text, the correlation coefficients concerning the comprehensibility of the 15 tested key messages were always less than 0.5. Significant correlations between comprehensibility and participant opinions were not found in either package insert group.

Further significant relationships were found concerning the amount of text to the following 8 of the 17 participant opinions about the package inserts. An increase in the number of words meant that the:

- First impression of the originals deterred the participants from reading further (originals: *P* = 0.021; Figure 4).
- Confidence about using the medicine decreased (models: *P* = 0.022; Figure 5).



FIGURE 3

Relationship between the amount of text in package inserts and the percentage of incorrectly comprehended information from all 15 tested key messages. Definitions as in Figure 1.

FIGURE 4

Relationship between the amount of text in package inserts and the motivation to read the package insert (17). Definitions as in Figure 1.

- Participants felt worse informed by the information contained in the package inserts (originals: *P* = 0.014).
- Participants more frequently did not want similar package inserts in future (originals: *P* = 0.015).
- Participants more frequently expressed the opinion that the package inserts contained too

much information (originals: P = 0.033, models: P = 0.014).

- Information provided in the package insert was more frequently difficult to locate (originals: *P* = 0.001).
- Information provided was difficult to understand (models: *P* = 0.014).

FIGURE 5

Relationship between the amount of text in package inserts and the confidence to use the medicine after reading the package insert. Definitions as in Figure 1.



• Participants more frequently stated the text was difficult to read (originals: *P* = 0.032).

A similar high number in significant relationships of participants' opinions about the originals was found in the percentage of difficult words per total number of words. Terms were assessed as potentially difficult for patients based on both the experience of the study leader and their occurrence in medical dictionaries. The higher the percentage of medical terms, the more participants stated:

- They lost confidence in using the medicine (P = 0.033).
- The package insert did not explain all important information (*P* = 0.049).
- The package insert contained too much information (*P* = 0.027).
- The package insert was difficult to understand (P = 0.014).
- Complicated sentences were contained (P = 0.017).
- Difficult words were in the package insert (P = 0.035).
- The information provided was not precise (*P* = 0.008).
- At the beginning of the package insert there was less important information (P = 0.042).

The following further relationships concerning the participants' opinions were detected in the group of the originals:

- An increase of the average number of words per sentence reduced confidence in using the medicine (*P* = 0.043).
- The higher the percentage of sentences with subjunctive tenses, the more people stated, after reading the package insert, that the first impression put them off reading the information (P = 0.005), the package insert contained too much information (P= 0.049), and the important information was not provided at the beginning (P = 0.01).
- The increase in the percentage of words in brackets per total number of words led to more patients feeling worse informed about the medicine (P = 0.038); more people stated that the package insert contained too much information (P = 0.019); fewer participants wanted the package insert in the future (P = 0.02); and the information provided was more difficult to locate (P = 0.043) and to read (P = 0.002).
- The higher the percentage of words longer than 20 letters, the more participants stated that the text was difficult to understand (P = 0.042).

Significant influences of demographic data on the participants' opinions were often found $(P \le 0.005)$. However, the Cramer V was never over 0.195, which showed that only very weak relationships exist. The most significant dependencies within the 17 investigated opinions were found in the aspect of age (originals 15×, models 14×) followed by education level (originals 14×, models 11×), sex (originals 5×, models 12×), participant's mood (originals 8×, models 10×), and finally, the medicine use (originals 2×, models 10×).

For example, the correlation calculation showed that an increase in age led to higher motivation to read the information. This was significant in 9 of the 10 investigated package inserts ($P \le 0.029$). However, the older the participants, the more frequently the opinion was stated for every original that too much information was contained ($P \le 0.029$).

DISCUSSION

Apart from the number of difficult words contained in package inserts, the amount of text is also a very important key factor in the use of package inserts. Even if extensive package inserts are not generally less comprehensible, they significantly reduce the possibilities of locating the information and patients more frequently feel worse informed than with shorter versions. In addition, long texts decrease the motivation to read the provided contents and in the end only a few people will read them.

Furthermore, patients who have less trust in their medicines caused by extensive texts will more frequently not comply with the instructions. In the worst case they will not use the prescribed medicines. More significant influences of the amount of text on the use of package inserts as described in the results are anticipated as only two different leaflet types, each with five versions, were investigated in this survey.

In a readability test study with 40 people, Dickinson et al. (13) also found problems as a result of long package inserts. Therefore some of the participants recommended shortening of the texts.

An investigation of 68 German package inserts from frequently used medicines selected in the year 2000 found an average text amount of 1,496 words (12). The unpublished PAINT2 survey of 271 package inserts, randomly selected from all versions available on the German medicine market in 2005, showed a significant text increase over 5 years to an average of 2,004 words per leaflet. Further, rises in the amount of text are expected, for example, due to the new demands of Directive 2004/27/EC (1), the text increase of the QRD-template (a text frame for package inserts in the European Union) (14) and continuing practical experiences with the medicines.

Recommendations, such as those from the FDA, to provide a summary of the most important contents at the beginning of package inserts would cause a further increase in the volume of text (15). However, the results of the five model package inserts showed that shorter leaflets without this summary are appropriate to inform patients about the medicines, so the FDA approach cannot be recommended for European package inserts. It follows that the current proposal for amending European Directive 2001/83/EC to require a summary of the essential package insert information (16) similar to the FDA requirements must therefore be assessed as inappropriate. In addition, no evidence-based research is available to suggest that such a summary, as used in the United States to inform health professionals, is helpful for patients when included in European package inserts. This is particularly the case as other layout and design aspects, such as bold print, were successfully used in each of the five models to emphasize the most important information. Apart from the negative effects of increasing the amount of text, a summary could more frequently lead to patients not reading the entire package insert.

The serious negative effects of extensive package inserts should be more considered in future approaches. Therefore, shortening package inserts is an important aspect to consider. A first step would be to undertake measures that we can immediately put into practice before or during the package insert user tests. Examples are:

- Avoiding repetitions and extensive explanations
- Using short points instead of long sentences
- Reducing the text that is intended only for doctors

GUIDELINES, TEMPLATES, AND DIRECTIVE IMPROVEMENTS

The second step is related to ongoing guideline, template, and directive optimizations, as user tests alone cannot always reduce the text to the optimal amount of fewer than 1,000 words or a maximum of 1,500 words (17).

One possibility for optimization concerns the existing Quality Review of Documents (QRD) template, which contains over 500 words (14). The results of this survey with the five model package inserts, containing a template text of around 200 words, indicate that shorter templates are sufficient and lead to significantly better results in the locatability of information. Furthermore, consistency between different guidelines, templates, and directives would be helpful. However, some documents lack this consistency, with the result that less appropriate texts are used in package inserts. For example, the QRD template contains long sentences, repetitions, and abbreviations while both the old and new readability guidelines recommend avoiding them (14,18,19).

A further aspect that should be reconsidered is the requirement to include less important information. Since the implementation of Directive 2004/27/EC, the names under which a medicine is sold in the different European Union member states have to be provided in package inserts (1). However, this information is less important for patients (11). Should it be required by individuals, every pharmacist or doctor, and especially the manufacturer, should be able to provide this information upon request.

Another example concerns the number of pharmaceutical company addresses, particularly in centralized approved package inserts (14). No patient sees the need for almost 30 different addresses of the same company. Therefore, reducing the number of addresses would be another good opportunity to shorten and optimize package inserts.

In addition, the guidelines, templates, and directives should focus more on the essential aspects. Here there are parallels to package inserts as precise, comprehensible, concise, and realistic rules can be better put into practice. If more people understand our guidelines and templates, this will help us to move forward. Guidelines, templates, and directives should more closely reflect scientific and practical experience. Data relating to the amount of text were already provided above. However, applying the more frequent research results in these documents will avoid unrealistic recommendations such as the recent proposal for amending Directive 2001/83/EC so that any new or amended text in package inserts shall, for a period of 1 year, be presented in bold, with an extra symbol and the words "New information" (16). Reasons for the impracticability of this suggestion are:

- Many medicines have a shelf life of up to 5 years. Patients receiving medicines shortly before the shelf life ends would have information presented as new that would in fact be old. In addition, patients could be confused when text is emphasized in one package insert as new, not emphasized in the next package insert, then in a later package insert again mentioned as new.
- The use of bold print is very appropriate to emphasize the most important information. But too frequent use of bold print decreases the effect of emphasizing key messages.
- The suggested presentation of new information will need an annual update of all package inserts delivered in the more than 100,000 medicines available in the European Union. This could lead to a collapse of the agencies that have to approve each amendment and increase the costs of the medicines.

Apart from the points mentioned above, guidelines, templates, and directives should be provided in fewer documents and preferably on one central website so that every person has easy access to the most up-to-date documents. The Regulatory and Procedural Guidance website of the EMEA (20) is, in this case, a good opportunity for everything that should be considered in package inserts and user tests to be provided in one place. If guidelines were presented in fewer documents, it would avoid confusion. For example, rules relating to user testing are provided in the European Commission document from 2006 (2), in the readability guideline (19), in national guidelines, and in other sources. As a result, discussions were sometimes observed between the companies and agencies about the actual requirements. A similar situation exists in the recommendations to write easily readable and comprehensible package inserts.

USER TEST SUCCESS CRITERIA

Realistic and appropriate success criteria are also required for the user tests. The current success criterion in the verbal face-to-face interview method is the 90/90% rule. In total, a minimum of 80% of the participants should be able to use each tested key message (2,3,19).

The written readability test success criteria are similar. One such is that in total 80% of the participants should be able to locate and understand each tested key message. There was a discussion relating to higher success criteria. Professor Sless, one of the verbal interview method developers, did not recommend higher user test success criteria than we have at the moment because, in his opinion, they are unattainable without falsifying the results (21). Furthermore, many people have significant reading and writing difficulties. For example, 5% of the German adult population have these problems (22). Therefore, aiming higher than the current success criteria would not be an appropriate way forward.

Nevertheless, it is necessary to focus on the improvement of the complete package insert and not only on the tested information, for example, by counting the number of difficult words and the amount of text. For this reason, improvements of the entire package insert are done as the first step of the written readability tests before the first test round with patients starts. Especially in this step, the main optimizations are achieved and the tests with people help to fine-tune the package inserts.

RIGHT TIME TO CARRY OUT USER TESTS

Another aspect is the right time to carry out user tests. The current situation is that these tests are done before the approval procedure starts. However, this can result in text changes after the user tests. For example, difficult words and extensive paragraphs may reoccur in the successfully tested version. One reason is the approval procedure. A suggestion is, therefore, that the package insert texts are first optimized and in a second step the pharmaceutical companies submit these texts to the agencies for approval. Afterward, the first test round with patients can start. This recommendation is based on practical experience, as PAINT-Consult has used it together with some companies in the past and it avoided extensive text changes and text increase after the test. However, this is only a suggestion to provoke discussion, since general changes such as this require the existing approval procedure to be modified, which is not so easy to do.

A further point to take into consideration is the harmonization of package inserts in Europe. Sometimes companies submit an optimized, successfully tested package insert, but the agencies demand to use another tested version for harmonization. As a result, the best and shortest package insert is not always used. Therefore, clear rules are required about when a harmonized text should be used or not and specifying that the best and shortest version is to be used. Otherwise we waste time and resources, and trust will be lost in the current system.

SUMMARY

User test research shows that ongoing package insert improvements are absolutely necessary. These can be achieved through user testing. However, apart from the user test success criteria, testing should focus on the improvement of the entire package inserts. Reducing the amount of text is one very important point. Establishing the right time to carry out user tests and clear rules for harmonization of the package inserts should also be considered.

In addition, ongoing guideline, template, and

directive optimizations are essential. These documents should focus more on the relevant aspects and reflect more closely scientific and practical experience. Precise, comprehensible, concise, and realistic rules can be better put into practice to achieve shorter and more understandable package inserts.

CONCLUSION

Our way forward in package insert user testing should be that we focus more on patient requirements and improvements of the package inserts. Therefore, creating appropriate European solutions especially to reduce the amount of text in package inserts is essential.

REFERENCES

- European Parliament and the Council of the European Union. Directive 2004/27/EC of the European Parliament and of the Council of 31 March 2004 amending Directive 2001/83/EC on the Community code relating to medicinal products for human use. *Off J Eur Communities*. 2004;(L136):34–57.
- European Commission. Guidance concerning consultations with target patient groups for the package leaflet. May 2006. http://ec.europa .eu/enterprise/pharmaceuticals/eudralex/vol-2/ c/user_consultation_200605.pdf. Accessed December 10, 2008.
- Sless D, Wiseman R. Writing About Medicines for People: Usability Guidelines for Consumer Medicine Information. Canberra: Department of Health and Family Services; 1997.
- Fuchs J, Hippius M. Inappropriate dosage instructions in package inserts. *Patient Educ Coun*sel. 2007;67:157–168.
- US Food and Drug Administration. An introduction to the improved FDA prescription drug labeling. https://www.cderlearn.com/fdaprescriptionlabeling/mainpage_wo_credit.cfm. Accessed April 27, 2009.
- Chin-Quee D, Wong E, Cuthbertson C. Evaluating information on oral contraceptive use: a randomized controlled trial to access missed pill instructions. *Hum Reprod.* 2006;21:3137–3145.
- Hubal R, Day R. Understanding the Frequency and Severity of Side Effects: Linguistic, Numeric and Visual Representations. American Association for Artificial Intelligence; 2006.

- Svarstad B, Mount J. Evaluation of written prescription information provided in community pharmacies, 2001. Final report to the US Department of Health and Human Services and the Food and Drug Administration. http://www.fda .gov/cder/reports/prescriptioninfo/default.htm. Accessed April 27, 2009.
- Bernardini C, Ambrogi V, Fardella G, Perioli L, Grandolini G. How to improve the readability of the patient package leaflet: a survey on the use of colour, print size and layout. *Pharmacol Res.* 2001;5:437–443.
- 10. PAINT-Consult, Jena, Germany, 2008.
- Fuchs J, Banow S, Görbert N, Hippius M. The importance of package insert information in the European Union. *PharmInd.* 2007;69(2):165–172.
- Fuchs J, Hippius M, Schaefer M. Analysis of German package inserts. Int J Clin Pharmacol Ther. 2006;44(1):8–13.
- Dickinson D, Raynor DK, Duman M. Patient information leaflets for medicines: using consumer testing to determine the most effective design. *Patient Educ Counsel*. 2001;43:147–159.
- EMEA. QRD template centralised procedure version 7.2 and MR/DC/referral procedures 1.2. http://www.emea.europa.eu/htms/human/qrd/q rdtemplate.htm. Accessed November 25, 2008.
- 15. US Food and Drug Administration. FDA News. FDA announces new prescription drug information format to improve patient safety. January 18, 2006. http://www.fda.gov/cder/regula tory/physLabel/summary.pdf. Accessed December 10, 2008.
- 16. Commission of the European Communities. Proposal for a directive of the European Parliament and of the Council amending, as regards pharmacovigilance, Directive 2001/83/EC on the Community code relating to medicinal products for human use. http://ec.europa.eu/enterprise/pharmaceuticals/pharmacos/pharmpack_12_2008/pharmacovigilance/pharmacovigilance dir_en.pdf. Accessed December 22, 2008.
- Fuchs J, Schweim H. Packungsbeilagen—Text muss kürzer werden. *Pharmazeutische Zeitung*. 2006;151:3864–3869.
- European Commission. A guideline on the readability of the label and package leaflet of medicinal products for human use, Brussels, 29 September 1998. http://ec.europa.eu/enterprise/ pharmaceuticals/eudralex/vol-2/c/gl981002
 .pdf. Accessed December 10, 2008.

- European Commission. Guideline on the readability of the labelling and package leaflet of medicinal products for human use. Revision 1, January 12, 2009. http://ec.europa.eu/enterprise/ pharmaceuticals/eudralex/vol-2/c/2009_01_12 _readability_guideline_final.pdf. Accessed January 16, 2009.
- 20. EMEA. Regulatory and procedural guidance: product information. file:///C:/Dokumente%

20und%20Einstellungen/fuchs/Desktop/pro ductinformation%20documents%20EMEA.htm. Accessed December 10, 2008.

- 21. Sless D. PIL testing: misapplied and out of context. *Regul Rapporteur J.* 2007;(11):14–15.
- 22. Döbert M, Hubertus P. *Ihr Kreuz ist die Schrift*. Münster: Bundesverband Alphabetisierung; 2000.

The author reports no relevant relationships to disclose.