1. Introduction

How people respond to healthcare information crucially depends on how this information is designed [1,2]. This study assesses the quality of three patient information leaflets. Substantial regulatory efforts have been made to improve the design of patient information leaflets within the EU [3]. In 1998 the European Community issued a directive that requires pharmaceutical companies to base their leaflets on a template [4]. This so-called QRD template regulates four aspects of package leaflets: (1) its content elements; (2) the order in which these topics should be discussed; (3) the headings for paragraphs and subparagraphs; (4) the wording of a number of specific passages. Another important EU criterion is that the information in the leaflet for the patient conforms entirely to the report on the clinical studies, the so-called Summary of Product Characteristics (SPC). Furthermore, EC Directive 2001/83 stipulated that “patient leaflets should reflect the results of consultations with target patient groups to ensure that it is legible, clear and easy to use”[5]. As of November 2005, there is an obligation to test new leaflets. In many states of the EU, a test procedure similar to the one proposed by Sless and Wiseman [6] has become the standard. In this procedure, 20 participants answer questions orally about the leaflet. Although Sless and Wiseman stress that the test is primarily a diagnostic tool, the European Guideline suggests a success criterion: each of the questions should be answered correctly by 80% of the consumers. Other publications are less strict and require that the mean correct percentage over all questions should be at least 80%[7].

Perhaps paradoxically, these extensive efforts have produced a problematic set of constraints [8] for the design of package leaflets:

- the template’s headings and text structure have never been tested and thus do not guarantee readability for patients;
- the template’s extensive information requirements lead to leaflets that may be more than 2000 words in length (some authors doubt whether a user-friendly leaflet is even possible given these requirements [9]);
- the testing success criterion puts a pressure on testing agencies to deliver successful tests, which makes it rewarding to ask easy questions to higher educated participants;
- it is unclear what should be done if the test results are incompatible with the template and with the obligation to follow the SPC[10].
As a result, it may still be doubted whether the European leaflet is a usable document. A German survey from 2005 found that many readers consider the leaflet too long, are unnerved by the ‘negative’ information in it, and would prefer more concrete action-directed information [9].

In the present study, we assess the usability of three recent leaflets approved by the Dutch Medicine Evaluation Board (MEB). We redesigned the leaflets and tested them again, using what Day [11] has called the “alternative representations” approach: original displays are analyzed and redesigned according to basic cognitive principles, retaining the information in both documents. Our dependent variables follow Wright’s [12] account of how people read functional documents. Key processes are accessing, comprehending, and applying the information. Our questions were:

- Do these leaflets enable patients to find, comprehend and apply relevant information?
- Can the leaflets be improved while remaining within the European regulatory framework?

The last question makes our study different from the few medicine leaflet redesign studies we know of. Morrow et al. [13] studied US pharmacy leaflets, reducing their length from 557 to 251 words, while Fuchs and Hippipus [2] shortened German patient information leaflets to roughly 600 words. Both studies have demonstrated significant improvements in the usability of leaflets, but their revisions were unconstrained by government regulations and drastically changed both the content and the global structure of the leaflet.

Our text revisions were based on a set of evidence-based Document Design principles concerning the quality of the text’s structure and its visual signalling:

- Answering comprehension questions becomes difficult when the text provides several passages that might be relevant (so-called distractors) [14]. Hence we integrated information on the same topic, whenever this was allowed by the template. If not, we added cross-references between passages on related topics.
- Headings facilitate text search [15]. We added two headings in the section on how to use the medicine.
- Headings in text need to be visually discriminable [16]. Hence we introduced a clearer hierarchy of section and subsection headings (presented in bold, with different font sizes).
- Heavy text signalling (e.g. by underlined, capital or bold text), deteriorates text recall [17]. Hence we removed bold fonts from the body text.
- Listed information is better processed when given in an explicit list than when given in prose format [18]. All sentences containing lists were transformed into an introductory segment followed by a bulleted list.
- Readers expect the main idea of paragraphs to be given in the beginning of their paragraphs and made them more explicit if needed.
- The leaflet structure needs to follow readers’ pre-existing schemata [20]. This requirement is violated when the side effects are grouped according to organ systems. We grouped them instead in terms of frequency and severity, and accordingly changed the subheadings.

Three leaflets were redesigned according to these principles. In the main study, all three originals and revisions were tested among users. In a follow-up study, one original-revision pair was tested again, but in a different paper format.

2. Methods

2.1. Original texts

Patient information leaflets were tested for three commonly used medicines: (1) Oxazepam, a member of the benzodiazepine family, (2) Bisoprolol, prescribed for patients with hypertension, and (3) Rosuvastatin, given to patients with elevated levels of LDL-cholesterol. The first leaflet (1128 words) dated from before the EU template and was shorter than the other two (2157 and 2666 words, respectively). The leaflets contained no illustrations.

For the main study, we used the A4versions of the original leaflets that had been approved by the Medicine Evaluation Board. The revisions were presented in the same A4 format (210 mm × 297 mm; 8.3 in. × 11.7 in.). The main differences between the A4 leaflets and actual inserts are that the A4 leaflets use larger fonts (10.5 points), are in black and white and use ‘normal’ paper quality, while the inserts use a smaller font, are in colour, have a two-column lay out and are printed on thinner paper.

The unusual A4 format may affect localization performance. Hence we performed a follow-up study for one of the leaflets (Rosuvastatin); in this study, the original and revision were presented in the normal package insert format.

2.2. Interview procedure

Participants were visited at home by the interviewers. We trained 30 students to carry out the interviews and the tests. The procedure started with a quick literacy test, inspired by Keselman et al. [21]. This test contained thirty multiple choice items containing potentially difficult words taken from the three original leaflets. Eighteen words were medical terms (e.g. ‘side effect’), twelve words were non-medical (e.g. ‘recommendation’). The test enabled us to control for differences in linguistic ability.

The second part of the procedure was the user test. The participants received a package leaflet and were asked to answer 15 questions. These were read aloud and handed to them on cards. According to Wright [12], users of healthcare information look for information relevant to their personal situation. The design implication is that such information should be found swiftly. In order to test this implication, our participants were not allowed to read the document before the test started.

For every question, participants had to locate the relevant information and answer the question orally. The interviewer stopped her stopwatch when the answer started. Interviewers were instructed to allow a maximum search time of 5 min. Most unsuccessful searches were stopped before this deadline.

Interviewers asked the participants to paraphrase answers when they were unsure whether a participant really understood a fragment. However, they were not allowed to prompt participants who gave incomplete answers (“is there anything else that could be relevant?”).

After the user test, participants evaluated the leaflet by completing the Consumer Information Rating Form (CIRF) developed by Koo et al. [22] and answered questions about their age, education, mother tongue and their use of medicines. The whole procedure took between 30 and 50 min.

2.3. Questions in the user test

Fifteen questions were asked about different sections of the leaflets. Instead of asking simple reproduction of text information (e.g. “please mention two side effects from this leaflet”), our questions required the participants to apply information in the leaflet to real life situations. For example, the participants were
told to imagine that they were allergic for titian dioxide and then had to find out whether they would be allowed to take Rosuvastatin. The questions focused on five functions of patient information leaflets [23]. The number of questions for a topic varied between leaflets, since the leaflets differed in the amount of information on these topics:

- Patients know whether the medicine fits their complaints (1 question).
- Patients know whether or not they can safely take the medicine (contra-indications and interactions; 2–4 questions).
- Patients know how to use the drug (3–5 questions).
- Patients know what side effects may occur and what to do in case they occur (3–4 questions).
- Patients know whether using the drug may affect certain activities in everyday life (3 questions concerning alcohol use, driving and breast feeding).

The tests for original and redesigned leaflets used the same questionnaire. Almost all question scenarios were explicitly discussed in the leaflet. In our 45 questions, only two questions required participants to make a (modest) inference to relate the text to the scenario. An example is a question about a patient who cannot remember what to buy in the supermarket, which must be related to a text passage on “memory loss”.

2.4. Redesign

Our main redesign principles have been discussed above. We need only add here that we used active sentences where possible and we simplified complex sentences. This led to a reduction in mean sentence length for the three leaflets (original: 13.5 words; revision: 11.0 words). Although some lexical simplifications were made, the revisions did not differ from the originals in the proportion of frequent words.

In conclusion, our revisions mainly affected the text structure and its visual signalling. However, the paper size and font size of the original and the revised documents were kept the same. An example of our revision strategy is provided in Fig. 1, which contains the directions for use in the Oxazepam leaflet.

The text length of the original and revised versions was different for two of our leaflets. The Oxazepam leaflet was entirely restructured to conform to the EU template format, which resulted in a longer text (1128 words original version, 1555 words revised version). For Rosuvastatin, which was originally 2666 words long, the revision was shorter (1892), mainly due to deletions of overlapping information.

Our revisions were constrained by the EU regulations, so that some reader problems could not be addressed satisfactorily. For example, participants had problems finding information about ingredients that could produce allergic reactions. According to the template, this information must be presented in the final section, under the obligatory heading Further Information. We inserted a reference to this section in the earlier paragraph headed Before you use X, where participants expected to find this information. But this is not optimal, because the heading Further Information does not help to locate information. Likewise, the requirement of conformity between the leaflet and the SPC affects the position of the information that children should not use the medicine. Such information would be best placed under the heading do not take X if..., but cannot be placed there if it is not listed as a contra-indication in the SPC.

The follow-up study used the Rosuvastatin text of the earlier revision except for a few minor changes that were made in consultation with Dutch MEB members in order to maximally comply with the EU regulations.

<table>
<thead>
<tr>
<th>Original fragment</th>
<th>Revised fragment</th>
</tr>
</thead>
</table>
| The doctor decides the appropriate dose, taking into account the nature of the complaints. In cases of anxiety and tension, the usual dose is one 10 mg tablet, taken 3 to 4 times a day. In serious cases it may be necessary to increase the dose to 150 mg a day with a maximum of 300 mg. | How and when should you take Oxazepam?  
Take the tablet with water. Swallow it whole with a glass of water. Do not dissolve the tablet in water and do not chew on it.
For sleeping problems you should take Oxazepam at least one hour before going to bed.  
How much Oxazepam should you take?  
- In case of sleeping problems 20 to 50 mg, to be taken at least one hour before going to bed.  
- One should start with the lowest dose, as the risk of side effects increases with higher doses.  
A lower dose is prescribed for elderly patients, children and patients suffering from liver or kidney problems or from a chronic respiratory disease called hypercapnia. Take the tablets with water.  
You will be given the lowest dose to start with, as the risk of side effects increases with higher doses.  
Furthermore, you will be given a lower dose when you belong to one of the following groups:  
- elderly  
- children  
- patients suffering from liver problems  
- patients suffering from kidney problems  
- patients suffering from a chronic respiratory disease called hypercapnia. |
2.5. Dependent measures

There were four dependent measures. Localization Success means whether or not participants found the relevant information in 5 min. Localization Time refers to the time participants needed to find the relevant information in the patient information leaflet. Average Localization Time (over questions and subjects) was computed for each of the three package leaflets. Localization time was only taken for successful localizations. Comprehension was measured with the help of key words for every question. Several questions had more than one key word. For example, participants were asked whether they were allowed to drink alcohol while using the drug; in case they thought this was not allowed they were asked to motivate their answer. Synonyms for key words were scored as correct answers. Perception of Usability was measured using the Consumer Information Rating Form (CIRF) [22]. Participants had to rate the patient information leaflet after the user test on the following criteria: (1) easy to read, (2) easy to understand, (3) easy to remember, (4) easy to locate information, and (5) easy to keep. The answers were given on a scale ranging from 1 (very hard) to 5 (very easy). The five questions had a Cronbach’s Alpha of .85 for our 316 participants. They were averaged in the analysis.

A composite measure named Task Success was computed by multiplying Localization Success and Comprehension. For example, if the information was located successfully for 90% of the questions, and if 90% of these questions were answered correctly, the Task Success score was 81%.

Finally, the control variable Literacy was measured with a multiple choice vocabulary test of 30 items. The literacy score was the percentage of test questions answered correctly. After removing 3 items, Cronbach’s Alpha was .72. The medical and non-medical items in the test correlated highly, suggesting that we measured a general linguistic ability and not health literacy.

2.6. Participants

The original texts were tested with 154 users, while 164 participants read the revised texts. A number of participants were recruited from the databases of TNS Consumer Research and the Dutch Consumers’ Association. Other participants were recruited through the networks of the student interviewers. All participants received a 10 Euro gift voucher. Participants who (had) used the specific drug deleted from the data set.

There were more female than male participants (65% versus 35%). The mean age was 51 years. There were no differences in gender or age between text versions. Gender did not affect performance. Age correlated negatively with localization performance (\(p < .01\) in both cases). Participants had a mean number of 11.8 years of formal education.

The follow-up study used 25 additional participants for both revision and original text.

3. Results

3.1. Literacy

Participants who read the revised texts had a higher literacy score (mean proportion correctly answered items: .81, SD .13) than the participants who read the original texts (.74, SD .13). Table 1 shows that the performance of our participants is substantially correlated with literacy. Literacy was a better predictor of performance than years of formal education. Hence all subsequent ANOVA’s used participants’ literacy score as a covariate.

3.2. Locating the information

Tables 2 and 3 show the means for localization success and localization times, corrected for literacy.

First of all, the results indicate that the QRD template does not guarantee usability. The original leaflets that follow the QRD template are no better than the original Oxazepam leaflet, which lacks the template structure.

As Table 2 makes clear, the revisions led to significantly better localization performance. And Table 3 shows substantially reduced localization times for the revisions. It takes about a minute to find information in the old versions, and 35–40 s in the new versions. In our study, readers did their very best to find the information. In real life, they may give up more easily. Hence the reduced localization times are important in that shorter times will produce less ‘drop-outs’. We conclude that our revisions in

### Table 1

<table>
<thead>
<tr>
<th>Correlation of literacy score with:</th>
<th>Location success</th>
<th>Location time</th>
<th>Comprehension score</th>
<th>Evaluation score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leaflet</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bisoprolol (n = 107)</td>
<td>.57**</td>
<td>−.45**</td>
<td>.21*</td>
<td>.10</td>
</tr>
<tr>
<td>Oxazepam (n = 101)</td>
<td>.43**</td>
<td>−.35**</td>
<td>.21*</td>
<td>.33**</td>
</tr>
<tr>
<td>Rosuvastatin (n = 105)</td>
<td>.58**</td>
<td>−.39**</td>
<td>.08</td>
<td>.27**</td>
</tr>
</tbody>
</table>

### Table 2

Localization success scores (mean proportion of questions correctly located, corrected for literacy).

<table>
<thead>
<tr>
<th></th>
<th>Original (SE)</th>
<th>Revision (SE)</th>
<th>Significance of differences for means</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bisoprolol</td>
<td>.761 (.016)</td>
<td>.910 (.016)</td>
<td>F (1,102) = 42.17, (p = .001), (\eta^2 = .289)</td>
</tr>
<tr>
<td>Oxazepam</td>
<td>.764 (.019)</td>
<td>.825 (.018)</td>
<td>F (1,98) = 5.54, (p = .021), (\eta^2 = .054)</td>
</tr>
<tr>
<td>Rosuvastatin</td>
<td>.789 (.017)</td>
<td>.873 (.016)</td>
<td>F (1,102) = 12.82, (p = .001), (\eta^2 = .112)</td>
</tr>
</tbody>
</table>

### Table 3

Localization times (mean time in seconds for successful locations, corrected for literacy).

<table>
<thead>
<tr>
<th></th>
<th>Original (SE)</th>
<th>Revision (SE)</th>
<th>Significance of differences for means</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bisoprolol</td>
<td>64.4 (3.1)</td>
<td>35.7 (3.0)</td>
<td>F (1,104) = 41.48, (p &lt; .001), (\eta^2 = .285)</td>
</tr>
<tr>
<td>Oxazepam</td>
<td>55.9 (3.7)</td>
<td>36.8 (3.5)</td>
<td>F (1,98) = 13.97, (p &lt; .001), (\eta^2 = .125)</td>
</tr>
<tr>
<td>Rosuvastatin</td>
<td>63.1 (3.5)</td>
<td>39.0 (3.2)</td>
<td>F (1,102) = 24.53, (p &lt; .001), (\eta^2 = .194)</td>
</tr>
</tbody>
</table>
structure and signalling produce substantial gains in localization performance.

3.3. Comprehension and appreciation

The comprehension scores in Table 4 show significant improvements as well, although they were already high for the original texts. A question-by-question analysis suggests that the following local revision strategies were particularly effective:

- restructured descriptions of how the drug works;
- explicit warnings (e.g. do not drive when taking this drug ...);
- explicit instruction how the drug should be used (e.g. the pill should be swallowed in its entirety, not dissolved or bitten into pieces).

The European Guideline on Readability [3] requires that all items in a readability test should be answered correctly by at least 80% of the participants, while EFPIA [7] is less demanding and requires a mean percentage of 80% over all questions and users. Do these three patient information leaflets meet these criteria? Table 5 provides success scores for the original and revised patient information leaflets.

Table 5 shows that all three original documents failed to meet the 80%-criterion over all questions and users. Our revisions were successful for Bisoprolol (.65 → .83) and Rosuvastatin (.74 → .85), but the Oxazepam patient information leaflet did not improve enough to meet the criterion (.62 → .75). The stricter criterion requiring an 80% score for every single question was not met by any of the leaflets.

Our results are remarkable given that two of the three leaflets have previously passed user tests approved of by the Dutch MEB. Our tests seem to have been more demanding than customary leaflet tests. This may have to do with our avoidance of simple reproduction questions, with our ban on prompting participants giving incomplete answers, with a lower number of participants from higher educational levels, or with the fact that our participants were not allowed to study the leaflet before being tested. We feel that our procedure is a valid one on all these counts, although a short 'leaflet preview' is an option.

Our final dependent measure was the participants' appreciation of the document, see Table 6.

Appreciation improved for all three package leaflets. Higher appreciations may enhance the motivation to actually read or continue reading package leaflets. However, the evaluations did not correlate with localization and comprehension performance.

3.4. Results of the paper format study

In the main study, our revisions resulted in better localization, comprehension and appreciation of the information. In order to rule out the possibility that this result is due to the unusual A4 format, the follow-up study repeated the tests for one of the leaflets (Rosuvastatin), using the folded presentation format of actual package inserts.

We analyzed the follow-up data together with the data of the two A4-versions of Rosuvastatin gathered in the main study. This yielded a 2 × 2 design with paper format (A4 versus insert) and text version (original versus revision) as factors; the results are presented in Table 7. The minor differences between the scores for the A4 format in Tables 7 and 2 are due to the presence of new participants in the covariance analysis.

<table>
<thead>
<tr>
<th>Table 4</th>
<th>Comprehension scores (mean proportion of correctly located questions that were correctly paraphrased, corrected for literacy).</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Original (SE)</td>
</tr>
<tr>
<td>Bisoprolol</td>
<td>.855 (.014)</td>
</tr>
<tr>
<td>Oxazepam</td>
<td>.806 (.018)</td>
</tr>
<tr>
<td>Rosuvastatin</td>
<td>.936 (.010)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 5</th>
<th>Success scores for three patient information leaflets (based on localization and comprehension scores corrected for literacy).</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Success scores</td>
</tr>
<tr>
<td></td>
<td>Original</td>
</tr>
<tr>
<td>Bisoprolol</td>
<td>.65</td>
</tr>
<tr>
<td>Oxazepam</td>
<td>.62</td>
</tr>
<tr>
<td>Rosuvastatin</td>
<td>.74</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 6</th>
<th>Evaluation scores (means on a scale from 1 (negative) to 5 (positive); corrected for literacy).</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Original (SE)</td>
</tr>
<tr>
<td>Bisoprolol</td>
<td>2.77 (.084)</td>
</tr>
<tr>
<td>Oxazepam</td>
<td>3.02 (.086)</td>
</tr>
<tr>
<td>Rosuvastatin</td>
<td>2.72 (.087)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 7</th>
<th>Results for the two Rosuvastatin formats, original and revised; corrected for literacy.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A4 format</td>
</tr>
<tr>
<td></td>
<td>Original version</td>
</tr>
<tr>
<td>Localization success</td>
<td>.783 (.016)</td>
</tr>
<tr>
<td>Localization times</td>
<td>64.46 (3.45)</td>
</tr>
<tr>
<td>Comprehension scores</td>
<td>.937 (.009)</td>
</tr>
<tr>
<td>Evaluations</td>
<td>2.72 (0.080)</td>
</tr>
<tr>
<td>Localization times</td>
<td>.783 (.016)</td>
</tr>
<tr>
<td>Localization times</td>
<td>64.46 (3.45)</td>
</tr>
<tr>
<td>Localization times</td>
<td>.783 (.016)</td>
</tr>
</tbody>
</table>
The format variable had a modest effect on localization success, the insert format having a 4% higher score. Furthermore, the original insert was evaluated more favorably than the original A4 format, while this difference disappeared in the revised versions—hence the format by version interaction in the evaluation scores. There was no effect of format on search time and comprehension score. Most importantly however, there is a robust main effect of text revision on localization performance and comprehension. Thus, our revision effects can be generalized over paper formats.

4. Discussion and conclusion

4.1. Discussion

We tested three patient information leaflets in order to find out whether readers could locate and comprehend relevant information. None of these documents achieved a success score of 80%, as is required by the European Guideline. We then revised the leaflets, drawing on the first test and on principles of Document Design. All three revisions resulted in increased localization, comprehension and appreciation, but one patient information leaflet still did not meet the 80% criterion. A follow-up study with one of the leaflets established that the revision gains are independent from the paper format used.

Our text materials were not tightly controlled as is possible in standard laboratory experiments. We did not isolate specific text features. And of course our three leaflets differed in multiple ways (e.g. medicine group, text length, template conformity). Although further research is needed, we may expect our study to be generalizable across the genre. First, the processing problems that we identified largely conform to those noted by other leaflet users. Thus, our revision effects can be generalized over this genre. Second, our revision strategies are not unique for this study since they are evidence-based (see Section 1). An important restriction to be noted here is that our structural revision strategies were confined to lower level structural features like optional subheadings, paragraphs and bulleted lists. This is because we revised the leaflets within the current QRD template. Quite likely, this template imposes a sub-optimal structure on the European leaflet. It is mainly based on the structure of the SPC, a document that differs greatly from patient information leaflets both in its purposes and in its reader group.

The close link between leaflet and SPC reflects the content-focused perspective that until recently has dominated the design of medicine information. In this perspective, the main issues are:

• What do we want to tell people about the medicine?
• How do we make sure that the information we give is correct?

In the performance-focused perspective advocated by Sless and Shrensky [24], the main question is: what are the tasks the reader must be able to perform with the help of the medicine information?

Such a performance-focused perspective will require a reconsideration of the template. Such a reconsideration should be based upon data about the pre-existing medication instruction schema that patients bring to the leaflet reading task. For instance, it is possible that patients expect to find the directions for using the medicine at the beginning of the medication instruction and not somewhere in the middle. Sorting tasks are a promising method to study these schemata [20] and we hope to report such a sorting study soon [25]. Eventually, revisions based on alternative leaflet templates may yield larger usability improvements than the current study has shown.

4.2. Conclusion

This study shows that the EU regulations, while producing uniformity in leaflet structure and content, do not guarantee leaflet usability and may even create problems for readers. On the other hand, leaflets can be improved within the regulatory framework, simply by applying Document Design principles. This study shows that interventions like these are not only effective in short 'informal' medication instructions [2,13] but also in the much longer and heavily regulated EU patient leaflets.

While this is good news, we welcome serious reflection on the ways in which the EU regulatory framework, taken as a whole, affects the usability of patient information leaflets.

4.3. Practice implications

On the short run, this work has two implications. Existing leaflets need to be tested in real performance tests (given that evaluations and performance levels do not correlate), And they need to be (re-)written according to basic evidence-based Document Design principles.

Acknowledgements

This work was partly supported by the Dutch Consumers’ Association. We also thank Hanneke Molenaar for her work on the follow-up study, and the students who participated in the project as interviewers.

References


