

# A framework for the evaluation of patient information leaflets

Mark Garner PhD,\* Zhenye Ning MA† and Jill Francis PhD‡

\*Head of Linguistics, School of Language and Literature, University of Aberdeen, Aberdeen, UK, †Senior Lecturer in Translation, College of Foreign languages and Literature, Northwest Normal University, Lanzhou, Gansu, China and ‡Reader in Health Psychology, Health Services Research Unit and Health Psychology Group, University of Aberdeen, Aberdeen, UK

## Abstract

### Correspondence

Mark Garner  
School of Language and Literature  
University of Aberdeen  
Aberdeen  
UK  
E-mail: m.garner@abdn.ac.uk

### Accepted for publication

30 December 2010

**Keywords:** evaluation, framework, linguistics, patient information leaflets, theoretical model, written communication

**Background** The provision of patient information leaflets (PILs) is an important part of health care. PILs require evaluation, but the frameworks that are used for evaluation are largely under-informed by theory. Most evaluation to date has been based on indices of readability, yet several writers argue that readability is not enough. We propose a framework for evaluating PILs that reflect the central role of the patient perspective in communication and use methods for evaluation based on simple linguistic principles.

**The proposed framework** The framework has three elements that give rise to three approaches to evaluation. Each element is a necessary but not sufficient condition for effective communication. Readability (focussing on text) may be assessed using existing well-established procedures. Comprehensibility (focussing on reader and text) may be assessed using multiple-choice questions based on the lexical and semantic features of the text. Communicative effectiveness (focussing on reader) explores the relationship between the emotional, cognitive and behavioural responses of the reader and the objectives of the PIL. Suggested methods for assessment are described, based on our preliminary empirical investigations.

**Conclusions** The tripartite model of communicative effectiveness is a patient-centred framework for evaluating PILs. It may assist the field in moving beyond readability to broader indicators of the quality and appropriateness of printed information provided to patients.

## Background

Current UK government policy is to provide patients with health information that is accessible and of high quality<sup>1</sup> as 'quality information empowers people to make choices that are right for them'.<sup>2</sup> Patient information leaflets (PILs) play a crucial role in this, and it is vitally important to assess their effectiveness in com-

munication with the target readership. It is essential, therefore, to have a means of evaluating PILs to identify potential for miscommunication as a basis for enhancement. In this paper, we present a theoretical framework for evaluation based on a linguistic model of interpersonal communication.

There is an extensive research literature on PILs evaluation,<sup>3-5</sup> largely focussed on patients'

capacity to read a particular leaflet. A preliminary search of Ovid MEDLINE and EMBASE (to Week 40, 2009) yielded 166 abstracts in which the key terms 'patient information' and 'readability' or 'reading ease' were present. Of these, 91 also considered patients' comprehension or understanding of the material (although several merely assumed that comprehension was the consequence of high readability).<sup>6,7</sup> Only 16 abstracts included the terms 'communication' or 'linguistic(s)'. Five of these took a patient perspective when evaluating printed information for patients, and of these, three<sup>8-10</sup> tested readers' comprehension of specific aspects of the text, whilst the other two<sup>11,12</sup> measured general approval ratings by patients. A sixth study<sup>13</sup> considered three aspects – readability, usability and likeability – as the criteria for quality of patient information. No studies, however, presented an explicit model for methods of evaluation.

The overwhelming majority of studies in our brief search used one or more well-established standard measures of readability, some of which are outlined as follows. Readability measures for evaluating leaflets have, however, been challenged. Mumford,<sup>14</sup> for example, notes that they are not sufficiently sensitive to distinguish complex but familiar terminology or common words that are used with different force in health care. She sees some value in them, nonetheless, provided they are used in conjunction with directed training for nurses or other leaflet writers. Others point out that readability is only an approximate guide to whether a text will be understood, and in particular, if it will be acted upon:

Readability scores provide information about the surface of the text, but [do] not directly provide information about the comprehensibility of the text. [They] do not acknowledge the specific needs of the target readers, ... [or] provide information about the coherence of a piece of text.<sup>15</sup>

In an important paper, Dixon-Woods<sup>16</sup> argues that the focus on readability arises from conceptions of the purposes of leaflets as well as from assumptions about the process of communication itself. The dominant conception derives from a biomedical perspective: PILs are a means of

patient education. Their purpose is to save time and energy and to provide medicolegal security for the providers of health care. In this view, the PIL is aimed at effecting cognitive, attitudinal or behavioural changes in patients, who are irrational, passive, forgetful and incompetent.<sup>16</sup> An alternative, less common conception of the purpose of PILs is comparatively new and is driven by a consumer advocacy agenda, which aims at empowerment rather than correction.

Prevailing assumptions about communication see it as unproblematic provided that potential sources of interference are removed. This has led to a preoccupation with readability as the major source of interference, a view that:

excludes the voice of patients from the evaluation of printed information, since the value of leaflets can be predicted by a formula about the relationship between syllables and sentence length.<sup>16</sup>

By contrast, conceiving of the purpose of PILs as patient empowerment values patients' rationality, competence, resourcefulness and reflexivity.<sup>16</sup> If communication is to be effective, the PIL 'must be noticed, read, understood, believed and remembered'.<sup>17</sup> Mayberry and Mayberry agree: a proper evaluation of patient information must consist of three necessary parts, which are a test of readability, a test of comprehension and a test of the long-term effects of the written materials.<sup>18</sup> What is needed is a theoretical basis for evaluation, in which readability tests have a role, but only insofar as they serve as an initial, mechanistic check that patients are likely to be able actually to read through a PIL. Some ideas that have been used include acceptability,<sup>19</sup> suitability,<sup>20</sup> cultural appropriateness,<sup>21</sup> usability<sup>22</sup> and comprehensibility.<sup>23</sup>

These patient-centred approaches are an important step forward in the evaluation of PILs. There is as yet, however, no theoretical framework that connects them with readability and hence provides a sound rationale for evaluation and improvement of leaflets. Very little research in this field has been informed by linguistics or by communication theory, and the literature remains largely descriptive and atheoretical. A notable exception is Kealley *et al.*,<sup>24</sup> who used

Systemic Functional Linguistics to analyse an information pamphlet written by nurses for patients' relatives. They concluded that the language of this pamphlet restricts and constrains the relatives while constructing the staff as ethical experts. By problematising the assumed neutrality of information in the communication process, Kealley *et al.*'s approach points the way to a more sophisticated and complex appreciation of the processes of communication between health-care providers and their clients. It has the disadvantage, however, of requiring an advanced knowledge of a highly technical linguistic theory, which makes it difficult for those in the health-care field who lack such knowledge to undertake such theory-based evaluations themselves.

In this paper, we present the broad outlines of a framework for evaluation. It is, we believe, relatively easy for professionals outside of the discipline of applied linguistics to understand and apply. Our starting point is alluded to in the aforementioned quote from Mayberry and Mayberry,<sup>18</sup> which mentions three elements essential to evaluating the likely success of PILs:

1. readability of the language of the text
2. readers' comprehension of the text
3. patients' long-term responses to the text

Our model is based on a theory of written communication that systematises these three elements and provides a basis for both empirical measurement of each element and a holistic evaluation of PILs. By investigating each element separately, the model enables the writers of a PIL to identify those specific aspects of the text and the target audience that act as obstacles to the responses that are the intended outcomes of the leaflet in the context for which it is being written. As will become evident, different factors influence each element, which are therefore evaluated using different methods.

In particular, the third element – the outcomes in terms of patients' reactions, expectations and decisions in the light of the writer's intention – is rarely represented in the literature. This is probably due partly to the logistical and ethical difficulties in obtaining relevant data from patients, but mainly it is because, we believe, of

the lack of a communication theory that connects such responses to specific characteristics of the text itself. The model we propose for text evaluation should improve researchers' ability to assess the draft of a PIL and also to enhance the final versions.

It should be noted, however, that a genuinely comprehensive evaluation of PILs also requires a clearer picture of the totality of the communication activities of which it is only a part, albeit sometimes a crucial part. An example of such an 'ecological' study in a very different area is Garner and Johnson's investigation of the total communicative context of emergency calls to the police.<sup>25</sup> This kind of fully contextualised approach, however, is beyond the scope of the present paper.

### A communication perspective

A great deal of research into the effectiveness of various forms of communication, in health care as in other fields, has been hampered by a fundamental misconception of communication itself.<sup>26</sup> The assumption that communication consists in the 'transmission' of 'information' has long been, and largely remains, the commonsense view. It was lent ostensible validity by the 'mathematical' formulation of Shannon and Weaver,<sup>27</sup> which consists in essence of a three-part linear model: sender → message → receiver. Until the latter part of the twentieth century, various versions of this basic model were widely accepted in the research literature. One of the first systematically to question the assumption was the philosopher Reddy,<sup>28</sup> who demonstrated how it fails to account for the most fundamental and significant characteristics of human communication. Reddy proposed a constructivist alternative, in which in Garner's terms, communication is inherently 'situated' (i.e. deeply embedded in its context), dynamic and interactive.<sup>29</sup> Mechanistic views of sending and receiving messages (often conceived of as mere 'information'), however, still implicitly, and at times explicitly, underlie much research into spoken and written communication in health care.

In contemporary communication theory literature, no single model is predominant. There

is, nonetheless, general agreement that any formulaic or graphical representation is misleading, simplistic or, conversely, so complicated as to be impossible to apply.<sup>30</sup> The conception of 'message' as a pre-ordained body of 'information' conveyed from one person to another should be abandoned. Messages are processes, in a constant state of flux, as potential meanings are continuously 'co-constructed' or 'negotiated' by the participants in a communicative event.<sup>29</sup> This is most evident in face-to-face conversation, from which all forms of human communication derive,<sup>31</sup> but equally, in a written text, which normally involves no personal interaction between writer and reader, the words on the page do not express a single, fixed meaning. On the contrary, a written text makes potential meanings available to a more or less limitless number of readers, each of whom will construct a meaning that is different, however subtly, from that of all other readers.

One reason why constructivist approaches to the study of communication are not yet the norm in health-care research is that they are highly inconvenient. The evaluation of PILs, for example, is incomparably more difficult when it is accepted that communication is a dynamic process. The one certainty is that the meaning constructed by a reader can never be precisely the same as that intended by the writer. There can be no failsafe linguistic template for writing PILs, or a neat formula for evaluating their effectiveness with patients: investigators have to be content with probabilities and approximations.

There is, however, we would argue, no alternative. To treat the leaflet as, in effect, a container of information will almost inevitably result in less than optimal communication between provider and patient. It may even, arguably, result in unethical practice. For example, it may be assumed that to provide patients with the necessary information, all that is required is to give them the leaflet 'containing' it and that any consent given (or withheld) is thereby 'informed'.

In summary, the framework we are proposing for PILs research takes a constructivist approach based on a series of postulates about the nature of communication. These are that:

1. communication is much more than sending and receiving information;
2. communication is an interactive process – directly between reader and text and indirectly between reader and author;
3. meaning is not inherent in the text, but is constructed by the reader.

### **A tripartite model of reading for evaluating and enhancing PILs**

The framework for evaluation that we propose takes a constructivist view of communication, which will enable researchers systematically to:

1. Evaluate any particular PIL
2. Make comparisons, between, for example, PILs written for different contexts or different versions of the same PIL
3. Explain why some PILs are more effective than others
4. Develop guidelines for writing and using PILs

The framework attempts to account for the complex and relatively unpredictable nature of the communication process without oversimplification, yet is simple and coherent enough to be applied by researchers without an extensive background in communication theory or linguistics. It has been used in some preliminary and small-scale studies, and the findings are positive, but it requires much more extensive and systematic empirical validation before it can be recommended for general adoption in PILs research and development. We hope that by publishing it at this still formative stage, we will encourage other researchers to contribute to testing and, where necessary, modifying it.

The reader of a PIL is involved in three overlapping but not coterminous processes:

1. Reading to the end
2. Constructing a coherent meaning
3. Responding

*Reading* is proceeding through the text and attempting to assign meanings to the words and phrases. *Constructing a coherent meaning* involves the reader in attempting to assign to the text *as a whole* a sufficiently clear, unified and

contextually relevant meaning. The *response* may be cognitive, affective or behavioural and varies according to the wording of the leaflet and the reader's circumstances.

All three of these processes are indispensable and occur to a fair extent simultaneously. A reader does not read the text, then interpret it, then respond, but is continuously constructing and reconstructing the meaning while moving through the text, and forming and re-forming the envisaged response. There is, nonetheless, a *logical* sequence, by which (i) is interpretatively prior to (ii), which is prior to (iii). To put it another way, successful completion of (i) is a necessary but not sufficient condition for (ii), which is a necessary but not sufficient condition for (iii).

To avoid the imprecision that characterises much discussion of what makes a text 'good' or 'successful', it is important to keep these processes conceptually, and hence analytically, distinct. Evaluation of a PIL requires a systematic assessment of each. That is why, as we argued earlier, a readability score is only the starting point of an evaluation. The ultimate test of the PIL as a means of communication is (3.): the reader's response. It is possible to conduct a valid evaluation of the communicative effect of a PIL purely by investigating responses, but such an approach is of limited applicability. If, for example, the response is deemed to be inappropriate to the communicative aims of the leaflet, in the absence of a clear picture of the interpretative challenges encountered by the reader during processes (1.) and (2.), it is very hard for the instigators (i.e. writers of the leaflet and/or the health-care professional who passes it to the patient) to know

what has led to the inappropriate response and how best to modify the text accordingly.

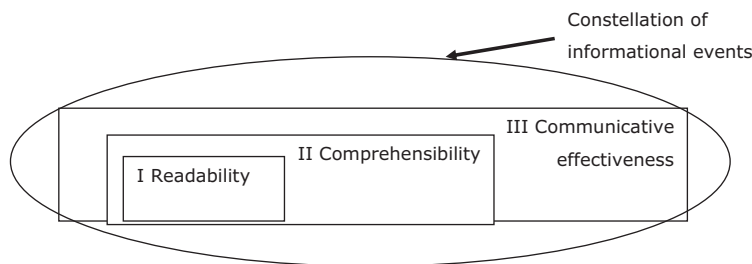
The relations between these three processes are represented in Fig. 1.

In Fig. 1, the 'constellation of communicative events' is included for completeness; it is not considered further in this paper. The term refers to the totality of interactions that take place in relation to a particular clinical intervention or clinical condition. These may, and almost certainly do, influence the construction of the meaning by the patient, but do not appear to have been the object of research to date. Such research is needed before a genuinely complete picture can be presented of the role of PILs in communication with patients.

Because these three processes are discrete, they are evaluated by means of different measures and methods. Because they are interrelated, however, the approaches to evaluation must be complementary. Together they can provide a holistic evaluation of a PIL and its parts. On this basis, (i) any hindrances to understanding can be identified and (ii) appropriate modifications made. We propose three indicative kinds of measures that meet these conditions. They are, respectively, measures of *readability*, *comprehensibility* and *communicative effectiveness*.

### Readability

Readability predicts the relative ease with which a reader can assign meanings to words and phrases. It has both a visual and a linguistic aspect. Visually, the print must be large enough and adequately spaced and sufficiently distinct from its background. Readability can be enhanced by



**Figure 1** Representation of the framework for evaluation of patient information leaflets, depicting the relations between the processes involved in making sense of a text.

the effective use of highlights, colours, graphics and the like. These visual aspects, although important, are not our concern here; rather, we focus on the linguistic aspects of readability, which include the length and syllabic make-up of words, as well as their likely familiarity to readers with a specified level of education.

Research aimed at developing valid readability ratings has been undertaken for at least seven decades, and more than 40 measures have been published. They differ in their intended field of application, the means used to arrive at a rating and the form in which the rating is given. Most use a formula based on some combination of the number of words per sentence; word length (in terms of number of syllables or letters); and word familiarity (derived from validated lists of words according to commonness of use). For example, the Fry Readability Formula<sup>32</sup> applies a simple formula based on the ratio of words of three or more syllables in 100-word excerpts from the beginning, middle and end sections of a text. A similar approach is taken in SMOG (Simple Measure Of Gobbledegook)<sup>33</sup> and the FOG index.<sup>34</sup> More complex formulae are employed in two of the most commonly used measures, the Flesch Reading Ease<sup>35</sup> score and the Flesch–Kincaid Reading Grade Level (FKRGL).<sup>36</sup> These were designed principally to assess reading texts for schools, but are also frequently used for other kinds of text.

A different kind of measure is given by the ‘cloze’ test,<sup>37</sup> in which every fifth word is blanked out and readers are asked to fill in the gap. Cloze has been shown in several studies<sup>38</sup> to be as reliable an indicator of readability as those that calculate it on the basis of word length or familiarity, numbers of syllables and the like. Although simpler in application, cloze is somewhat more sophisticated from a communicative perspective. An important element of interpretation is the ability to predict what will come next on the basis of what has gone before, so a cloze score is to an extent an indicator of the ongoing construction of meaning as the reader moves through the text. In other words, it is partly a measure of *comprehensibility* of the text as well as of its *readability*. Cloze is, however, at

best an approximate guide to the meaning-making process. Some missing words can be guessed simply because they are part of a longer phrase (such as ‘by and large’ and ‘over and above’), or because of the grammatical structure of the sentence, even if the words themselves are poorly understood. A reader can thus make correct guesses even in the absence of a coherent interpretation of the whole text. Moreover, cloze does not of itself indicate where the obstacles to an appropriate interpretation might lie.

Ley and Florio<sup>4</sup> provide an informative summary of a number of readability tests and report a high correlation between the results obtained in a range of texts using different methods. Their general conclusion, as far as readability is concerned, is that ‘much of the literature produced for patients, clients and the general public is too difficult’.

These tests have considerable limitations, nonetheless. At a purely practical level, the validity of a readability score requires a minimum word count (e.g. 100 words of continuous text), which are in excess of those in many PILs, such as those that accompany over-the-counter medicines. More fundamentally, these tests ignore factors such as the nature of the topic, the ordering of ideas, choices of sentence structure which do not affect length, and the reader’s background knowledge and stylistic and personal expectations.<sup>39</sup> Davison and Kantor object to readability formulas on the grounds that ‘[r]eading difficulty may be affected by the purposes and background of the reader and the inherent difficulties of the subject matter’. Meade *et al.*<sup>40</sup> stress the crucial role played in interpretation by communicative characteristics that have little to do with word length or familiarity, for example, ‘technical accuracy, format, learnability, motivational messages, ... legibility ... and experiential or cultural factors of the target group’.

The popularity of readability tests is attributable in part to the fact that they are generally quick and easy to apply (most contemporary word-processing software packages include a readability measure facility) and are expressed as a clear numerical value. Readability is at best,

however, unspecific as a guide to comprehensibility and even less as a predictor of response. Our tripartite model builds on the extensive findings of readability research, by using any of the well-validated scoring systems as the basis for evaluation. Nevertheless, an appropriate level of readability does not by itself guarantee that the other two essential criteria will be met. This leads us to the second stage of evaluation in the model.

### Comprehensibility

Sometimes a reader may arrive at the end of a page of apparently readable language only to realise that the whole does not make sense. That is why the notion of comprehensibility is necessary. The two most basic determinants of comprehensibility are vocabulary (*lexical items*) and grammatical structure (*syntax*), to each of which the reader must be able to attribute relevant and integrated meanings.<sup>41</sup>

Lexical items include not only individual words, but also groups of words, (e.g. compounds such as *General Practitioner* or phrases such as *as often as possible*) that are the semantic equivalent of a single word. Lexical items are included in some readability tests, in which they are given a familiarity score, based on frequency of use in everyday language. Some research has been conducted into the effects of rephrasing an instruction on the comprehension and compliance (referred to below as communicative effectiveness).<sup>42</sup> This is an aspect of the relationship between lexical elements that merits further exploration in relation to PILs, and it is hoped to develop it as our work on evaluation progresses. Objective measures of frequency have become available in recent years through analyses of vast databases of actual language in use in a wide range of contexts.<sup>43,44</sup>

Frequency, however, must not be confused with comprehensibility, which is defined as the reader's capacity to assign *contextually relevant* meanings to the items. There is thus only an indirect relationship between frequency and comprehensibility, as the following three points show. First, some infrequent items will almost

inevitably be necessary in a PIL: for example, those relating to certain surgical procedures. Provided that suitably worded definitions or explanations are given where appropriate in the text, the use of such terms need not detract from comprehensibility.

Secondly, and more much more problematically, items that are in frequent use, and therefore easily recognized, may not be assigned a contextually relevant meaning by the reader. Many everyday terms are used in health care with a different or more specific meaning. *Vein*, for example, is often used in non-technical discourse to refer to any blood vessel and therefore encompasses the meaning of the less widely used *artery*. Other common lexical items may have implications that are significant to the interpretation but are overlooked by the reader. In the PIL for a clinical trial<sup>45</sup> that was the subject of the evaluation outlined later, the compound *outpatients' department* was used in the section describing the locations in which the trial procedures would be conducted. This item would be readily recognized by any reader who had visited a large hospital, yet 10% of the participants in the research failed to realise the essential point that patients undergoing the procedure in this department would not be required to stay overnight. Another instructive example from the same PIL is *laser*. Despite its relatively high frequency in modern English (among the 2600 most common nouns in the British National Corpus,<sup>46</sup> with the same frequency as, for example, *surplus* and *tablet*), anecdotal evidence obtained during the trial suggested that a number of patients did not attribute the appropriate meaning to it when reading the PIL. They interpreted a laser solely as a source of light (not heat) and were thus unprepared for the painful burning sensation occasioned by this form of treatment.

A third complexity in the relationship between frequency and comprehensibility arises when two (or more) high-frequency words occur in a compound lexical item with a much lower frequency: the generic term *blood vessel* is an example. All of the preceding examples show that even apparently non-technical, high-

frequency items with a high readability score need to be evaluated for comprehensibility as well.

*Syntax*, or the structuring of sentences, is the other major linguistic factor in the reader's attempt to construct a coherent meaning. Most readers process a text sentence-by-sentence. Each new sentence is expected both to make sense in itself and to add some new meaning, however slight, or to re-emphasise what has preceded it. They 'attempt to interpret whatever structure they encounter ... immediately rather than engaging in a wait-and-see strategy'.<sup>47</sup> Each sentence (and indeed each clause within a sentence) must therefore be sufficiently short and readily interpretable to make the new meaning accessible. Many 'plain English' guides give practical and potentially helpful advice in structuring a sentence<sup>48</sup> and, when used as a checklist, can serve as a helpful basis for rating the likely comprehensibility of sentences when drafting a document. Equally important is the semantic relationship (or *cohesion*) between the preceding and following clauses or sentences. Even apparently straightforward sentences can impede comprehensibility, if, for example, the referent of a particular pronoun ('this', 'those', 'it', 'which', etc.) is unclear.

The aforementioned are only a few examples of the many linguistic factors that influence a reader's interpretation. Each PIL presents its own potential challenges to comprehensibility, and a full evaluation is required at this level for each. An on-going study evaluating a PIL for a varicose veins trial utilizes one approach to measuring comprehensibility that appears to have promise. Full details are reported elsewhere<sup>49</sup> but, briefly, comprehensibility was evaluated as follows. After a series of readability tests, the PIL was divided into five-sentence chunks, and a group of readers was asked to read it chunk by chunk and then answer multiple-choice questions, some of which related to lexical items, and some to sentence structures, within that chunk (see examples in Box 1). Two experienced linguists (MG, ZN), with the assistance of a health-care researcher (JF), developed the questions.

**Box 1** Text from PIL: The scars on your legs are easily noticeable to start with, but will continue to fade for many months after the operation. Very occasionally, some people develop a little brown staining where the veins were removed. Another uncommon but disappointing problem is the appearance of tiny thread veins or 'blushes' on the skin in the areas where varicose veins were removed.

*Nerve damage.* Nerves under the skin can be damaged when removing varicose veins close to them and small areas of numbness are quite common.

1. "Brown staining" is \_\_\_\_\_.  
(Please tick the box before the answer which you think is most appropriate here)
  - A. a side-effect of the treatment which has no serious result
  - B. a dirty spot in the skin which looks brown
  - C. a change in skin colour after using a certain kind of medicine
  - D. a change to light brown in the skin after the skin recovers from a cut
2. When the scars occur, we \_\_\_\_\_.  
(Please tick the box before the answer which you think is most appropriate here)
  - A. should keep the adhesive strips on them
  - B. should wear some dressings to make them fade
  - C. should not worry about them because they will fade after some time
  - D. should have a close observation at them
3. What information would you like to come next?  
(Write as long or as briefly as you like)

Considerable care was taken to try to ensure that the respondents' choice between options on each question was not influenced by factors other than their comprehension of the PIL. For example, we tried to ensure that the non-'correct' options in every question were *prima facie* reasonable (i.e. not so absurd as the be unselectable). We also worded the 'correct' options in such a way as to remove the possibility of guessing on the basis of pattern matching with the text.

The most appropriate answer to each was decided, on the basis of the text alone, by the linguists, working at first independently, and then in consultation. These answers were taken to be as reasonable as possible an indicator of the meanings expressed by the PIL text itself, independently of what the writers' (clinical researchers') intended meanings were. (These answers and the readers' responses were examined after the study, in conjunction with one of



the writers, who endorsed all of the linguists' classifications.)

The research participants were 30 honours-level university students of Linguistics. Despite the relatively high readability scores of the text and the advanced educational level of the participants, no individual was able to select the most appropriate answer to every question and some questions yielded no such answers. It is highly likely that a more genuinely representative sample of the target readership (i.e. patients) would have scored even lower. The fine-grained, stage-by-stage analysis of the comprehensibility of this PIL enabled us to identify very specifically both the lexical and the structural elements of the text that would require adjusting to improve the comprehensibility level.

Unlike readability, which can be measured solely on the basis of the text itself, comprehensibility is a function of the interaction of the reader with the text, and systematic evaluation must, therefore, include readers' constructions of the text. This is time-consuming, but if PILs are to be made as effective as possible by means of rigorous pre-publication assessment, a careful evaluation of comprehensibility is indispensable. The researcher's task may in time become easier by the establishment of a set of templates for use with many kinds of PILs: this is a fruitful field for further research.

Although investigating comprehensibility provides a much more sophisticated kind of evaluation than readability alone, the communicative success of a PIL is not guaranteed even when the readability and comprehensibility are high. It is quite possible for a reader to construct a meaning from the text that is coherent, but so divergent from that intended by the writer that it gives rise to an inappropriate response. (This is different from the situation in which the reader comprehends the intended meaning but makes a considered judgement not to comply with it: see below.) This brings us to the third analytical aspect of the tripartite model. Moreover, it cannot be assumed that readers will read systematically through the text from beginning to end – see, for example, Frantz.<sup>42</sup> Readers may begin by scanning the leaflet, seeking those parts that appear to be most

relevant or interesting, and may consciously or unconsciously skip over portions of the text. In the absence of research into this aspect of readers' behaviour with PILs, a linear reading process has to be assumed, but the actual situation is undoubtedly more complex.

### Communicative effectiveness

As defined earlier, communicative effectiveness is a function of the reader's cognitions (e.g. expectations, understandings), affect (e.g. relief, concern, worry) and often intentions and behaviour (e.g. taking a pill before eating). These are formed by the reader as a result of reading the text and cannot be identified by textual analysis alone. They must be ascertained from the target readership. This idea, incorporating the notion of 'usability', has been explored in the context of human-computer interaction by systematically examining the actions (and not only reported comprehension) of the target user to evaluate and optimise system design.<sup>50</sup>

Research into communicative effectiveness explores the nature of the readers' actual or intended responses. Any form of communication gives rise to variant interpretations, as a result of the expectations, motivations, prior knowledge and personal circumstances of the addressee, together with other factors. As with other types of effectiveness in relation to health care (e.g. clinical effectiveness, cost-effectiveness), we propose that the communicative effectiveness of a PIL be assessed on the basis of specified outcomes. In the case of a PIL, these outcomes can be ascertained by asking some basic research questions about readers' intention to respond to the PIL, e.g.:

1. What are these intentions?
2. Do these intentions correspond with the objectives of the PIL as defined by the instigator?
3. If the answer to (2.) is no, what are the reasons for the lack of intention to comply?

We argued earlier that communication does not consist in sending messages. The notion that communicative success can be judged according to whether the message 'received' will be the same

as the message 'sent' is untenable. Some variation in the meanings constructed by PIL readers is inevitable, and in many cases, there is a range of ways to respond appropriately. If, however, the response arises from a failure to grasp the PIL's objectives, the PIL has failed in its purpose. This is not to say that the ultimate criterion of assessment is whether the PIL persuades the reader to behave in a specified way. There are circumstances in which non-adherent responses may be appropriate. For example, PILs accompanying certain medicines frequently advise the reader to continue taking them even if experiencing 'relatively minor' side-effects. An individual may, however, have previously found these side-effects to be debilitating and decide not to comply. This decision is not in itself an indication of low communicative ineffectiveness – indeed, it may be a good decision based on effective communication.

There are therefore, variously, one or two phases in the evaluation. In the first phase, patients' responses are compared with the communicative aims of the instigator. One parameter of comparison is provided where leaflets specify the intended objectives. Instigators of a PIL have, however, a complex set of intended meanings – not all of them elaborated or even entirely conscious – so what can be made explicit within the text can serve only as the starting point. A second phase of evaluation is required if there is a significant divergence between response and intention (including a lack of any response). If there is, the reasons for the divergence are explored with the reader. The aim of this phase is to discover whether the unintended response is the result of a misunderstanding (i.e. the PIL has low communicative effectiveness) or of a considered decision not to comply with a fully understood objective. In the latter case, the PIL has high communicative effectiveness; the writer may then decide whether it would be helpful to incorporate such variant but appropriate responses in a revised text. In the example given in the preceding paragraph, the writer might add a rider such as, 'If, however, these side-effects are serious for you, you may decide to discontinue the medication. If you do, you should consult your health-care professional as soon as possible.'

A systematic set of criteria for evaluating reader responses is yet to be developed; they will necessarily include the degree of convergence between instigator and readers about both the importance and appropriateness of responses. One possible technique is the use of a measure of 'simulated behaviour'<sup>51</sup> in which respondents report what they would do in response to a written scenario (i.e. a simulated context). Because readers construct an array of meanings (including unexpected interpretations about what are the key messages), the communicative effectiveness of a PIL is also likely to be informed by less constrained methods of data collection, such as the semi-structured interview. For example, an interview prompt might be

Having read this PIL, what are you going to do?

Content analysis of interview transcripts can then be applied; codes can be assessed by the instigators and classified in an appropriate manner, such as:

1. understands and intends to comply with PIL objectives
2. intends not to comply owing to failure to understand PIL objectives
3. understands PIL objectives and makes a considered decision not to comply

## Conclusion

The question 'are information leaflets for patients effective?' is highly important, as the production of PILs is a significant cost to the health service. PILs play a vital role in attempts at assisting patients to make informed choices, to take treatments appropriately and to participate in research that may improve health. Usable answers to the question require an evaluation model based on a theoretically valid understanding of how written communication works.

Yet despite extensive published guidance about leaflet development, there are few models, tools or measures that can be used to evaluate whether such guidance results in improved outcomes – in other words, whether PILs are communicatively effective. Readability scores, while

important, are not sufficient for evaluating PILs. The proposed model, using insights from linguistics and communication theory, distinguishes three aspects of the process of communicating in writing, with a focus on communicative effectiveness as the ultimate criterion of evaluation. The model and methods that we propose are fundamentally patient-centred. By focusing separately on readability and comprehensibility, the model incorporates analysis of linguistic features of the text into the meaning construction by readers and guides the researcher to critical points of difficulty for the reader. The reader's response is, however, the determining endpoint of communication. By including communicative effectiveness, the model for evaluation is designed to draw attention to a crucial point. The meaning of the text of a PIL is not constituted by what is encoded by the writer. The reader constructs the meaning, and the outcome of the communication is his or her behavioural, cognitive and/or affective response. Effectiveness can be ascertained on the basis of a comparison of the writer's intended outcome with the actual response. Any significant difference between them needs to be further investigated as a possible indicator that the text would benefit from being reworked.

The next steps in this programme of work include applying the model diagnostically. Once potential points of difficulty are identified, linguistic principles will be formulated and applied to leaflets that require enhancement. Experimental studies will then be undertaken to test whether linguistic enhancement leads to improved communicative effectiveness of leaflets. We invite the research community to take the field forward by devising further tests of both the proposed model and the methods used in applying it.

## References

- 1 Department of Health. The information standard, 2010. Available at: <http://www.dh.gov.uk/en/Healthcare/PatientChoice/BetterInformationChoicesHealth/Informationstandard/index.htm>, accessed 3 February 2010.
- 2 Department of Health. The information accreditation scheme, 2010. Available at: <http://www.dh.gov.uk/en/Healthcare/PatientChoice/BetterInformationChoices>

- Health/Informationaccreditation/index.htm, accessed 3 February 2010.
- 3 Moulton B, Franck LS, Brady H. Ensuring quality information for patients: development and preliminary validation of a new instrument to improve the quality of written health care information. *Health Expectations*, 2004; **7**: 165–175.
- 4 Ley P, Florio T. The use of readability formulas in health care. *Psychology, Health and Medicine*, 1996; **1**: 7–28.
- 5 Helitzer D, Hollis C, Cotner J, Oestreicher N. Health literacy demands of written health information materials: an assessment of cervical cancer prevention materials. *Cancer Control*, 2009; **16**: 70–78.
- 6 Sarma M, Alpers JH, Prideaux DJ, Kroemer DJ. The comprehensibility of Australian educational literature for patients with asthma. *Medical Journal of Australia*, 1995; **162**: 360–363.
- 7 Shrank W, Avorn J, Rolon C, Shekelle P. Effect of content and format of prescription drug labels on readability, understanding and medication use: a systematic review. *Annals of Pharmacotherapy*, 2007; **41**: 783–801.
- 8 Ordoñas Baines JP, Lopez Briz E, Urbieto Sanz E, Torregrosa Sanchez R, Jimenez Torres NV. An analysis of patient information sheets for obtaining informed consent in clinical trials. *Medicina Clinica*, 1999; **112**: 90–94.
- 9 Mansoor LE, Dowse R. Effect of pictograms on readability of patient information materials. *Annals of Pharmacotherapy*, 2003; **37**: 1003–1009.
- 10 Mwingira B, Dowse R. Development of written information for antiretroviral therapy: comprehension in a Tanzanian population. *Pharmacy World and Science*, 2007; **29**: 173–182.
- 11 Sustersic M, Meneau A, Dremont R, Paris A, Laborde L, Bosson JL. Developing patient information sheets in general practice. Method proposal. *Revue du Praticien*, 2008; **58** (19 Suppl): 17–24.
- 12 Joshi HB, Newns N, Stainthorpe A, MacDonagh RP, Keeley FXJ, Timoney AG. The development and validation of a patient-information booklet on ureteric stents. *BJU International*, 2001; **88**: 329–334.
- 13 Wright P. Criteria and ingredients for successful patient information. *Journal of Audiovisual Media in Medicine*, 2003; **26**: 6–10.
- 14 Mumford ME. A descriptive study of the readability of patient information leaflets designed by nurses. *Journal of Advanced Nursing*, 1997; **26**: 985–991.
- 15 Pothier L, Day R, Harris C, Pothier DD. Readability statistics of patient information leaflets in a speech and language therapy department. *International Journal of Language and Communication Disorders*, 2008; **43**: 712–722.
- 16 Dixon-Woods M. Writing wrong? An analysis of published discourse about the use of patient infor-

- mation leaflets. *Social Science and Medicine*, 2001; **52**: 1417–1432.
- 17 Ley P. *Communication with Patients: Improving Communication, Satisfaction and Compliance*. London: Chapman and Hall, 1992.
  - 18 Mayberry JF, Mayberry MK. Effective instruction for patients. *The Journal of the Royal College of Physicians of London*, 1996; **30**: 205–208.
  - 19 O'Donnell M, Entwistle V. Consumer involvement in research projects: the activities of research funders. *Health Policy*, 2004; **69**: 229–238.
  - 20 Rees CE, Ford JE, Sheard CE. Patient information leaflets for prostate cancer: which leaflets should healthcare professionals recommend? *Patient Education and Counseling*, 2003; **49**: 263–272.
  - 21 Doak LG, Doak CC. Patient comprehension profiles: recent findings and strategies. *Patient Counselling and Health Education*, 1980; **1**: 101–106.
  - 22 Iddo G, Prigat A. Why organizations continue to create patient information leaflets with readability and usability problems: an exploratory study. *Health Education Research*, 2005; **20**: 485–493.
  - 23 Ley P. The measurement of comprehensibility. *International Journal of Health Education*, 1973; **11**: 17–20.
  - 24 Kealley J, Smith C, Winsor B. Information empowers but who is empowered? *Communication and Medicine*, 2004; **1**: 119–129.
  - 25 Garner M, Johnson E. The transformation of discourse in emergency calls to the police, Oxford University Press (in press).
  - 26 Nelson CK. The haunting of communication research by dead metaphors: for reflexive analyses of the communication research literature. *Language & Communication*, 2001; **21**: 245–272.
  - 27 Shannon CE, Weaver W. *The Mathematical Theory of Communication*. Urbana: University of Illinois Press, 1964.
  - 28 Reddy MJ. *The Conduit Metaphor – A Case of Frame Conflict in our Language About Language*. Ortony, *Metaphor and Thought*, 2nd edn. Cambridge: Cambridge University Press, 1993: 164–201.
  - 29 Garner M. *Language: An Ecological View*. Oxford: Peter Lang, 2004.
  - 30 Sless D. Repairing messages: the hidden cost of the inappropriate theory. *Australian Journal of Communication*, 1986; **9**: 82–93.
  - 31 Leeds-Hurwitz W. *Social Approaches to Communication*. New York: The Guildford Press, 1995.
  - 32 Fry E. A readability formula that saves time. *Journal of Reading*, 1967; **11**: 513–516.
  - 33 McLaughlin HG. SMOG grading: a new readability formula. *Journal of Reading*, 1969; **12**: 639–646.
  - 34 Gunning R. The FOG index after twenty years. *Journal of Business Communication*, 1968; **6**: 3–13.
  - 35 Flesch R. A new readability yardstick. *Journal of Applied Psychology*, 1948; **32**: 221–233.
  - 36 Kincaid JP, Fishburne RP, Rogers RL, Chissom BS. *Derivation of New Readability Formula for Navy Enlisted Personnel*. Millington, TN: Navy Research Branch, 1975.
  - 37 Taylor WL. Cloze procedure: a new tool for measuring readability. *Journalism Quarterly*, 1953; **30**: 415–433.
  - 38 Friedman DB, Hoffman-Goetz L. A systematic review of readability and comprehension instruments used for print ad web-based cancer information. *Health Education & Behavior*, 2006; **33**: 352–373.
  - 39 Davison A, Kantor RN. On the failure of readability formulas to define readable text: a case study from adaptations. *Reading Research Quarterly*, 1982; **17**: 187–209.
  - 40 Meade CD, Smith CF. Readability formulas: cautions and criteria. *Patient Education and Counseling*, 1991; **17**: 153–158.
  - 41 Beck IL, McKeown MG, Sinatra GM, Loxterman JA. Revising social studies texts from a text-processing perspective: evidence of improved comprehensibility. *Reading Research Quarterly*, 1991; **26**: 251–276.
  - 42 Frantz JP. Effect of location and procedural explicitness on user processing of and compliance with product warnings. *Human Factors*, 1994; **36**: 532–546.
  - 43 Wynne M. *Developing Linguistics Corpora: A Guide to Good Practice*. Oxford: Oxbow Books, 2005.
  - 44 McEnery T, Wilson A. *Corpus Linguistics*. Edinburgh: Edinburgh University Press, 1998.
  - 45 CLASS Writing Group: Brittenden J, Burr J *et al*. The CLASS Study: Comparison of LASer, Surgery and foam Sclerotherapy as a treatment for varicose veins. *Lancet*, 2008. Available at: <http://www.lancet.com/protocol-reviews/09PRT-2275>, accessed 31 January 2011.
  - 46 Leech G, Rayson P, Wilson A. Word Frequencies in Written and Spoken English: based on the British National Corpus. Available at: <http://uc-rel.lancs.ac.uk/bncfreq/>, accessed 19 March 2010.
  - 47 Kinstch W. *Comprehension: A Paradigm for Cognition*. Cambridge: Cambridge University Press, 1998: 124.
  - 48 Cutts M. *Oxford Guide to Plain English*. Cambridge: Oxford University Press, 2007.
  - 49 Forrest A. *The Patient Information Leaflet as Communication: A Mixed Methods Study to Inform a Trial of Surgical Interventions*. Unpublished dissertation, University of Aberdeen, 2009.
  - 50 Preece J, Rogers Y, Sharp H. *Interaction Design: Beyond Human-Computer Interaction*. New York: John Wiley & Sons., 2002.
  - 51 Hrisos S, Dickinson HO, Eccles MP, Francis J, Johnston M. Are there valid proxy measures of clinical behaviour? *Implementation Science*, 2009; **4**: 37.