



Computer-Based Readability Testing of Information Booklets for German Cancer Patients

Christian Keinki¹ · Richard Zowalla² · Monika Pobiruchin³ · Jutta Huebner¹ · Martin Wiesner²

Published online: 12 April 2018
© American Association for Cancer Education 2018

Abstract

Understandable health information is essential for treatment adherence and improved health outcomes. For readability testing, several instruments analyze the complexity of sentence structures, e.g., Flesch-Reading Ease (FRE) or Vienna-Formula (WSTF). Moreover, the vocabulary is of high relevance for readers. The aim of this study is to investigate the agreement of sentence structure and vocabulary-based (SVM) instruments. A total of 52 freely available German patient information booklets on cancer were collected from the Internet. The mean understandability level L was computed for 51 booklets. The resulting values of FRE, WSTF, and SVM were assessed pairwise for agreement with Bland–Altman plots and two-sided, paired t tests. For the pairwise comparison, the mean L values are $L_{FRE} = 6.81$, $L_{WSTF} = 7.39$, $L_{SVM} = 5.09$. The sentence structure-based metrics gave significantly different scores ($P < 0.001$) for all assessed booklets, confirmed by the Bland–Altman analysis. The study findings suggest that vocabulary-based instruments cannot be interchanged with FRE/WSTF. However, both analytical aspects should be considered and checked by authors to linguistically refine texts with respect to the individual target group. Authors of health information can be supported by automated readability analysis. Health professionals can benefit by direct booklet comparisons allowing for time-effective selection of suitable booklets for patients.

Keywords Evidence-based health information · Patient information booklets · Readability · Understandability · Support vector machine

Introduction

Low health literacy of patients is one of the most important reasons for worse prognosis [1]. It leads to less use of screening (e.g., mammography), preventions (e.g., influenza vaccination), and decreased therapy adherence. This eventually results in an overall poorer general health status, as well as greater barriers to the availment of adequate medical care. Yet, a greater involvement of patients in decision-making processes leads to higher treatment satisfaction, improved treatment adherence, less decision conflicts, improved health behaviors, and a better global health status. Moreover, further

improvements are higher therapeutic safety and a decrease of unnecessary medical procedures, less hospitalizations, lower emergency room visits, and less false receipts of drugs resulting in less consumption of resources [1–3].

Thus, the improvement of health literacy in general and provision of suitable information for the individual patient is a major concern in the German National Cancer Plan and similar initiatives in other western countries [4]. These include a low-threshold access to target group-oriented, quality-assured information, advice, and offers of assistance, leading to a strengthening of shared decision-making [5].

One of the main problems of written patient information material is the gap between the language of experts and lay people. Even with a higher level of health literacy, most vocabulary and concepts of diagnosis and treatment may not be easy to understand. In addition, even for patients with higher levels of education, understanding medical concepts might be difficult. Obviously, this also applies for patients with lower (health) literacy. Furthermore, the concepts of oncologists and patients concerning the origin of cancer are different [6, 7]. As information on diagnostic procedures, treatments, and follow-up treatments is highly complex [8], it is often not easy to

✉ Christian Keinki
christian.keinki@med.uni-jena.de

¹ Department of Hematology and Medical Oncology, University Hospital Jena, Am Klinikum 1, 07747 Jena, Germany

² Department of Medical Informatics, Heilbronn University, Max-Planck-Str. 39, 74081 Heilbronn, Germany

³ GECKO Institute, Heilbronn University, Max-Planck-Str. 39, 74081 Heilbronn, Germany

describe the underlying concepts in laymen-friendly language. Many cancer patients report that with the diagnosis of cancer, concentration and understanding decline noticeably [9]. Therefore, written health information should be made even easier to comprehend for those affected by such a diagnosis.

Readability is a term to describe the properties of written text with respect to the readers’ understanding of the document. It depends on the complexity of the text structure, the sentence structure, and the vocabulary used. In this context, the readability of written information can be checked with different instruments. Through each of these instruments, a specific readability score can be calculated. Examples for the English language include the Flesch Reading Ease Scale [10], the Flesch-Kincaid Readability Formula, the Fog Index, the SMOG Readability Formula, or the Fry Readability Graph [11]. However, the manual computation of these related formulas can be difficult and time-consuming and is therefore associated with a high demand of human and/or financial resources.

In addition to the aforementioned readability metrics, estimating the difficulty of a text can be interpreted as a text classification problem. Thus, methods from the field of natural language processing and machine learning can be applied for this task.

This paper presents the comparison of a computer-based readability analysis based on German patient information booklets for oncological patients. The assessment of the difficulty of 51 booklet texts was conducted by applying a German adaptation of the Flesch-Reading Ease Scale (FRE) [12], the Vienna Formula (Wiener Sachtextformel, WSTF) [13], and a purely vocabulary-oriented method, i.e., Support Vector Machines (SVMs).

Methods

Selection of Patient Information Booklets

We collected patient information booklets on cancer in the German language. All booklets had to be available for free on the Internet. From a former analysis of websites with information for cancer patients, we were able to identify websites with booklets for cancer patients [14]. This procedure was described in detail in a previous publication [15].

We focused on the six most common types of cancer in Germany for women and men respectively [16]. Thus, nine different types of cancer were taken into consideration: breast cancer, prostate cancer, lung cancer, colorectal cancer, endometrial cancer, bladder cancer, melanoma, pancreatic cancer, and oropharyngeal cancer.

FRE Scale

A well-established readability metric for the English language is the Flesch-Reading Ease Scale [10]. The FRE measures the readability of a text via its average sentence length (ASL) and average number of syllables per word (ASW). It relies on the fact that short words or sentences are usually easier to understand than longer ones. However, for this analysis, we applied the modified FRE for the German language by Toni Amstad [12]:

$$FRE = 180 - ASL - (58.5 \cdot ASW)$$

Vienna Formula (WSTF)

In contrast to the FRE, the Vienna Formula (WSTF) is not an adapted version for the German language. Instead, it relies on the work by Bamberger and Vanacek [13], who conducted an analysis on the bases of German text material. They derived at least five versions of the so-called Vienna formula for prose and non-fictional text. Typically, the fourth WSTF is used for text analysis. This metric also relies on the average sentence length (ASL) and on the proportion of words with three or more syllables (MS):

$$4th\ WSTF = 0.2656 \cdot ASL + 0.2744 \cdot MS - 1.6939$$

Transformation of FRE and WSTF Scales

For the purpose of comparison with the SVM results, the obtained readability scores of FRE and WSTF were transformed to the value system of L . Hence, normalized FRE values L_{FRE} were transformed as follows:

$$L_{FRE} = \begin{cases} x_{FRE} \leq 90 : & 10 \frac{x_{FRE}}{10} \\ x_{FRE} > 90 : & 1 \end{cases}$$

with $x_{FRE} \in [0, 100]$, $x_{FRE} \in \mathbb{R}$.

Analogously, WSTF scores were transformed to the scale of L according to

$$L_{WSTF} = \frac{x_{WSTF} - 3}{1.2} = (x_{WSTF} - 3) \cdot \frac{5}{6}$$

with $x_{WSTF} \in [4, 15]$, $x_{WSTF} \in \mathbb{R}$.

Understandability Level L

Modern methods from the field of machine learning can be leveraged to compensate for the limitations of and enhance established readability measures [17, 18]. In previous work, a vocabulary-based computation of an “expert level” using a

specifically trained SVM was presented, which is also applicable to German cancer booklets [18].

For this analysis, the raw material of all cancer booklets was originally downloaded as PDF documents. For this reason, a tool to transform and extract the raw text material was applied. As a second step, the extracted material was cleaned from disturbance artifacts (such as OCR artifacts). Then, *stop word removal* and *stemming* techniques were applied [19]. Next, unclassified documents (here: paragraphs from the booklets) were transformed into document vectors v based on the previously (i.e., in the training phase) computed feature set F [20]. The SVM probability output computed via *Platt scaling* is then transformed into L [21]. After these steps, the resulting SVM classifier allows to discriminate between laymen- and expert-centric text materials. The procedure was described in detail in [15].

Computation of L

First, the mean understandability level L for every patient information booklet was computed by the use of the specifically trained SVM. Second, the mean FRE and WSTF levels for every booklet were calculated, according to the original formulas. The computation was conducted via a self-implemented software tool which was tested for correctness against reference text material from the literature.

Next, the transformation was applied according to the aforementioned section. Thereby, a direct comparison of two classic metrics for readability analysis against a vocabulary-based metric (SVM) is made possible on the same scale (L).

Interpretation of Scores

The corresponding measure $L \in [1, \dots, 10]$ describes the understandability of text material, here booklets. A value of $L \geq 8-10$ represents a high degree of difficulty (academics); $L \geq 6 < 8$, a difficult text; $L \geq 4 < 6$, a moderate text for laymen with medical educational background; $L \geq 3 < 4$, an easy text (intermediate level/junior high school); and L between 1 and < 3 ,

a very easy text (elementary level/elementary school). Higher values of L require an academic (medical) background knowledge or working experience in the medical domain. The corresponding scores of FRE, WSTF, and L_{SVM} , as well as the corresponding class labels of all three scores, are presented in Table 1.

Statistical Methods

For a pairwise comparison of the aforementioned readability metrics, the differences (delta) of their respective values were plotted against their mean according to the methodology introduced by Bland and Altman [22]. The upper and lower limits of agreement (LOA) are defined as $\bar{d} \pm 1.96s$ where \bar{d} denotes the mean difference and s denotes the standard deviation of pairwise differences. If 95% of computed readability score differences lay inside the LOA, both methods can be considered interchangeable, i.e., equally appropriate. According to Bland and Altman, the resulting absolute delta of upper and lower limits should be carefully evaluated for practical implications. In the presented setting, an observed LOA delta of 1 would equal 10% of the normalized readability scale L . In case less than 5% of the observed values are observed outside of the LOA range, this might suggest that the two metrics could be used interchangeably. However, if the actual LOA delta spans a large amount of the scale, the results should be interpreted very carefully, as done by Bland and Altman [23]: “How far apart measurements can be without leading to problems will depend on the use to which the result is put, and is a question of clinical judgment. Statistical methods cannot answer such a question.”

In addition to the Bland–Altman analysis, all L_{SVM} scores were tested pairwise ($n = 51$) against L_{FRE} and L_{WSTF} values with a two-sided, two-sample t test ($\alpha = 0.05$). The alternative hypothesis is thus defined as the difference of means for L_{SVM} and L_{FRE} scores, L_{WSTF} , respectively.

The software suite R was used in version 3.3.3 (2017-03-06) for the statistical analysis on a Windows 10 Enterprise LTS 2016/64bit computer.

Table 1 Interpretation of scores with corresponding class label

| Description | FRE $\in [0, 100]$ | WSTF $\in [4, 15]$ | $L \in [1, 10]$ | Class label |
|--------------------------|--------------------|--------------------|-----------------|-------------|
| Very difficult to read | [0–29] | [14–15] | [8–10] | VD |
| Difficult to read | [30–49] | [13–14] | [6–8] | D |
| Fairly difficult to read | [50–59] | [10–13] | [5–6] | M |
| Average readability | [60–69] | [8–10] | [4–5] | M |
| Fairly easy to read | [70–79] | [7–8] | [3–4] | E |
| Easy to read | [80–89] | [5–7] | [2–3] | VE |
| Very easy to read | [90–100] | [4–5] | [1–2] | VE |

FRE readability by Flesch-Reading Ease Scale; *WSTF* Readability by Vienna Formula; L understandability level by Support Vector Machine; class label: *VE* very easy; *E* easy; *M* moderate; *D* difficult; *VD* very difficult

Results

Characteristics of the Patient Booklets

The acquisition of booklets was carried out between February 14 and 21, 2015. A total of 52 different patient booklets were downloaded. One booklet from the pharmaceutical industry about lung carcinoma could not be investigated, since the length of the related text sections was not sufficient for the analysis. Finally, 51 booklets could be assessed within this study (see Table 2). In sum, 26 booklets (51.0%) were available from the different regional Cancer Societies in Germany, 10 (19.6%) from the German Cancer Aid, 1 (2.0%) from the Cancer Information Service, 6 (11.8%) from the websites of the pharmaceutical industry, 5 (9.8%) from the websites of different statutory health insurance companies, and 3 (5.9%) from the websites of the self-help organizations. Most booklets referred to breast cancer (14; 27.5%) followed by prostate cancer (8; 15.7%) and colorectal cancer (7; 13.7%). For more detailed information, see [15].

Understandability of the Booklets (SVM)

The booklets had a mean understandability value of $\bar{L}_{SVM} = 5.09$. As listed in Table 2, two of the 51 booklets (3.9%) had a mean score ≥ 8 and therefore are only suitable for an academic readership. Seven of the 51 booklets (13.7%) had a mean score < 4 and were therefore suitable for a lay audience. In only five of the 51 booklets (9.7%) was a mean score < 3 reached, so that an elementary education was sufficient for understandability of these booklets. For the remaining 42 booklets (82.4%), a mean score between ≥ 4 and < 8 shows a level suitable for laymen or persons with medical knowledge or strong medical background. The distribution of the computed values of L_{SVM} is depicted in Fig. 1a.

The group of endometrial cancer booklets showed the lowest mean value $\bar{L}_{SVM, \text{endometrial}} = 4.08$. The highest understandability level was observed for the booklet group on prostate cancer with $\bar{L}_{SVM, \text{prostate}} = 6.08$.

Readability of the Booklets (FRE)

The mean FRE score for all booklets resulted in a value of $\bar{L}_{FRE} = 6.81$ (see Table 2). An academic background—i.e., mean readability score ≥ 8 —is necessary for three of the 51 booklets (5.9%). All of the other booklets (94.1%) scored an L between ≥ 4 and < 8 . This indicates that a reader should have (basic) medical knowledge or an educational background in the field of medicine to understand these booklets. None of the 51 booklets showed a readability score < 4 which corresponds to the non-existence of class labels VE and E in Table 2. Figure 1b depicts the L_{FRE} value distribution.

Table 2 Listing of scores (L) and associated class labels (C) for all metrics under investigation

| Booklet topic | L_{SVM} | L_{FRE} | L_{WSTF} | C_{SVM} | C_{FRE} | C_{WSTF} |
|----------------|-------------------|-------------------|-------------------|-----------|-----------|------------|
| Breast_01 | 4.91 | 6.49 | 6.95 | M | D | D |
| Breast_02 | 6.57 | 6.50 | 7.10 | D | D | D |
| Breast_03 | 4.25 | 7.08 | 7.55 | M | D | D |
| Breast_04 | 8.48 | 9.00 | 9.65 | VD | VD | VD |
| Breast_05 | 6.75 | 7.29 | 8.04 | D | D | VD |
| Breast_06 | 5.00 | 7.45 | 8.24 | M | D | VD |
| Breast_07 | 5.50 | 5.84 | 5.93 | M | M | M |
| Breast_08 | 1.80 | 7.10 | 7.52 | VE | D | D |
| Breast_09 | 5.00 | 6.13 | 6.19 | M | D | D |
| Breast_10 | 4.62 | 7.03 | 7.73 | M | D | D |
| Breast_11 | 5.23 | 6.78 | 7.38 | M | D | D |
| Breast_12 | 5.81 | 7.28 | 8.20 | M | D | VD |
| Breast_13 | 4.92 | 6.68 | 7.15 | M | D | D |
| Breast_14 | 4.30 | 5.58 | 5.36 | M | M | M |
| Mean | 5.22 | 6.87 | 7.36 | | | |
| Prostate_01 | 6.75 | 7.67 | 8.30 | D | D | VD |
| Prostate_02 | 6.62 | 6.51 | 7.12 | D | D | D |
| Prostate_03 | 5.24 | 6.98 | 7.55 | M | D | D |
| Prostate_04 | 1.67 | 5.88 | 6.80 | VE | M | D |
| Prostate_05 | 6.68 | 7.47 | 8.53 | D | D | VD |
| Prostate_06 | 5.50 | 6.81 | 6.95 | M | D | D |
| Prostate_07 | 6.29 | 7.37 | 8.14 | D | D | VD |
| Prostate_08 | 9.79 | 9.15 | 9.83 | VD | VD | VD |
| Mean | 6.08 [‡] | 7.23 | 7.90 [‡] | | | |
| Lung_01 | 4.77 | 6.50 | 7.04 | M | D | D |
| Lung_02 | 4.84 | 7.34 | 7.92 | M | D | D |
| Lung_04 | 2.53 | 5.64 | 6.11 | VE | M | D |
| Lung_05 | 5.44 | 6.82 | 7.48 | M | D | D |
| Mean | 4.40 | 6.58 | 7.14 | | | |
| Colorectal_01 | 4.41 | 6.40 | 6.94 | M | D | D |
| Colorectal_02 | 5.74 | 6.54 | 6.70 | M | D | D |
| Colorectal_03 | 5.49 | 6.99 | 7.57 | M | D | D |
| Colorectal_04 | 4.84 | 6.61 | 7.03 | M | D | D |
| Colorectal_05 | 5.44 | 7.04 | 7.84 | M | D | D |
| Colorectal_06 | 4.94 | 6.71 | 7.45 | M | D | D |
| Colorectal_07 | 4.71 | 6.64 | 7.25 | M | D | D |
| Mean | 5.08 | 6.70 | 7.25 | | | |
| Bladder_01 | 5.83 | 6.64 | 7.38 | M | D | D |
| Endometrial_01 | 4.44 | 6.99 | 7.62 | M | D | D |
| Endometrial_02 | 2.67 | 8.10 | 8.37 | VE | VD | VD |
| Endometrial_03 | 4.00 | 6.82 | 7.39 | M | D | D |
| Endometrial_04 | 5.22 | 7.04 | 7.92 | M | D | D |
| Mean | 4.08 [†] | 7.24 [‡] | 7.83 | | | |
| Melanoma_01 | 4.58 | 6.32 | 6.66 | M | D | D |
| Melanoma_02 | 7.70 | 6.99 | 7.80 | D | D | D |
| Melanoma_03 | 1.88 | 5.32 | 5.22 | VE | M | M |
| Melanoma_04 | 5.87 | 7.27 | 8.32 | M | D | VD |
| Melanoma_05 | 4.45 | 6.76 | 7.74 | M | D | D |

Table 2 (continued)

| Booklet topic | L_{SVM} | L_{FRE} | L_{WSTF} | C_{SVM} | C_{FRE} | C_{WSTF} |
|------------------|-----------|-------------------|-------------------|-----------|-----------|------------|
| Mean | 4.90 | 6.53 | 7.15 | | | |
| Oropharyngeal_01 | 4.38 | 6.53 | 7.13 | M | D | D |
| Oropharyngeal_02 | 3.33 | 6.98 | 7.94 | E | D | D |
| Oropharyngeal_03 | 6.41 | 6.57 | 7.06 | D | D | D |
| Oropharyngeal_04 | 5.96 | 6.64 | 7.17 | M | D | D |
| Mean | 5.02 | 6.68 | 7.33 | | | |
| Pancreatic_01 | 4.33 | 6.69 | 7.47 | M | D | D |
| Pancreatic_02 | 3.38 | 5.42 | 5.65 | E | M | M |
| Pancreatic_03 | 4.97 | 6.54 | 6.87 | M | D | D |
| Pancreatic_04 | 5.38 | 6.53 | 7.36 | M | D | D |
| Mean | 4.52 | 6.30 [†] | 6.84 [‡] | | | |
| Mean total | 5.09 | 6.81 | 7.39 | M | D | D |

^x Class labels: *VE* very easy, *E* easy, *M* moderate, *D* difficult, *VD* very difficult

[†] Lowest mean of L among all cancer types

[‡] Highest mean of L among all cancer types

The booklets on pancreatic cancer achieved the lowest mean value $\bar{L}_{FRE,pancreatic} = 6.30$. The highest readability level was observed for the text material on endometrial cancer: $\bar{L}_{FRE,endometrial} = 7.24$.

Readability of the Booklets (WSTF)

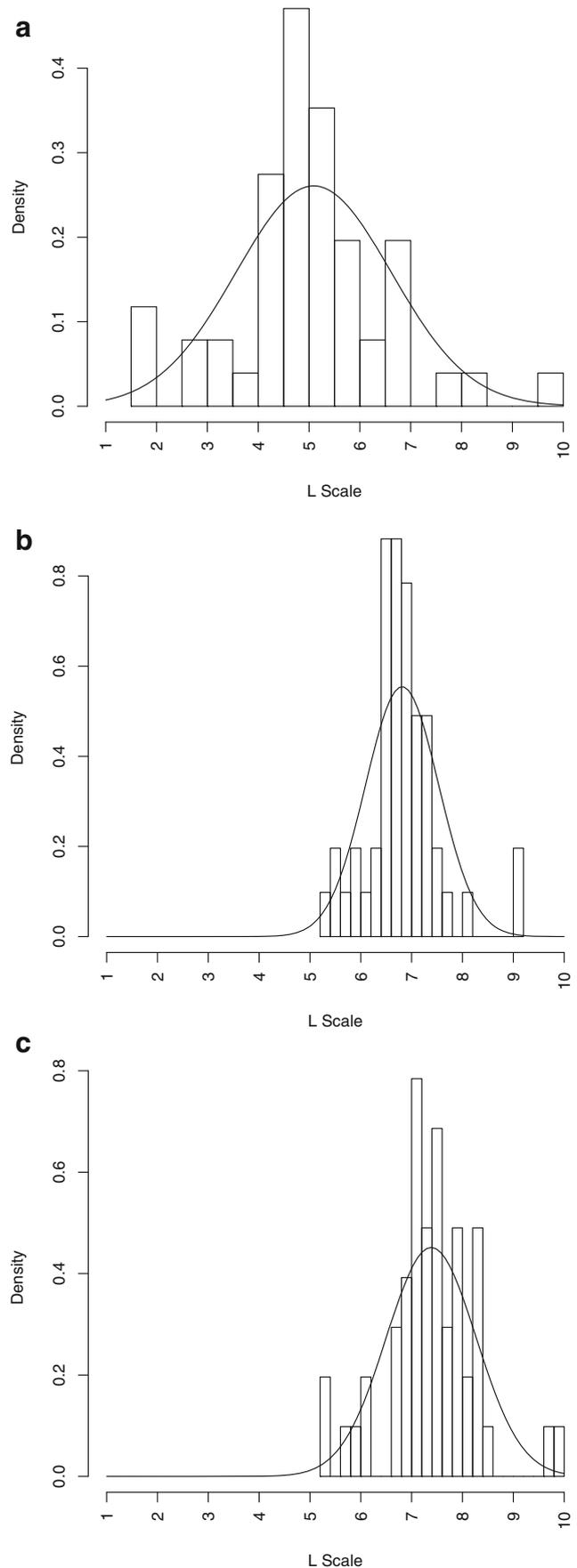
For the Vienna Formula, the booklets scored an \bar{L}_{WSTF} of 7.39. A total of 19.6% (10 of 51) of booklets were identified as very difficult to read (VD); see Table 2. The remainder of the booklets (80.4%) were classified as M and D which correspond to an L between ≥ 4 and < 8 . None of the booklets showed levels of L below 4, as depicted in Fig. 1c.

With a mean value of $\bar{L}_{WSTF,pancreatic} = 6.84$, booklets on pancreatic cancer achieved the lowest value for all cancer types. The highest mean WSTF score was observed for the prostate cancer group: $\bar{L}_{WSTF,prostate} = 7.90$.

Comparison of Instruments

The results of a quantitative comparison analysis of all booklets are presented in Fig. 2. It depicts a Bland–Altman plot for the delta of $L_{SVM} - L_{FRE}$. A total of 48 of the 51 data points are located within the upper and lower LOA. The LOA range spans from -4.181 to 0.736 which represents approximately 50% of the scale of L . The two-sided, two-sample t test

Fig. 1 **a** Distribution of the SVM vocabulary-based scores of L_{SVM} ($n = 51$, mean = 5.090, sd = 1.530). **b** Distribution of the FRE scores, transformed to L_{FRE} ($n = 51$, mean = 6.813, sd = 0.719). **c** Distribution of the WSTF scores, transformed to L_{WSTF} ($n = 51$, mean = 7.385, sd = 0.883)



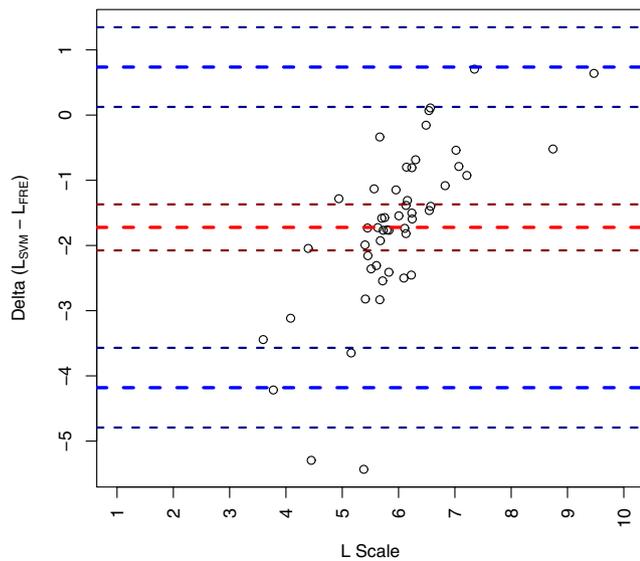


Fig. 2 Bland–Altman plot (SVM–FRE), comparison of $n = 51$ samples. Mean of differences $\bar{d} = -1.723$ (95% CI [-2.075, -1.370]), lower LOA $\bar{d} - 1.96s = -4.181$ (95% CI [-4.792, -3.570]), upper LOA $\bar{d} + 1.96s = 0.736$ (95% CI [0.125, 1.347])

showed a significant difference between the means of L_{SVM} when compared to L_{FRE} ($P < 0.001$).

A similar result is given by the Bland–Altman plot in Fig. 3 (comparison between L_{SVM} and L_{WSTF}). In this case, the LOA ranges from -4.774 to 0.185 (again, approx. 50% of the L scale). Again, three data points are not within the limits of agreement. In accordance with the prior comparison, the two-sided, two-sample t test revealed a significant difference between L_{SVM} and L_{WSTF} ($P < 0.001$).

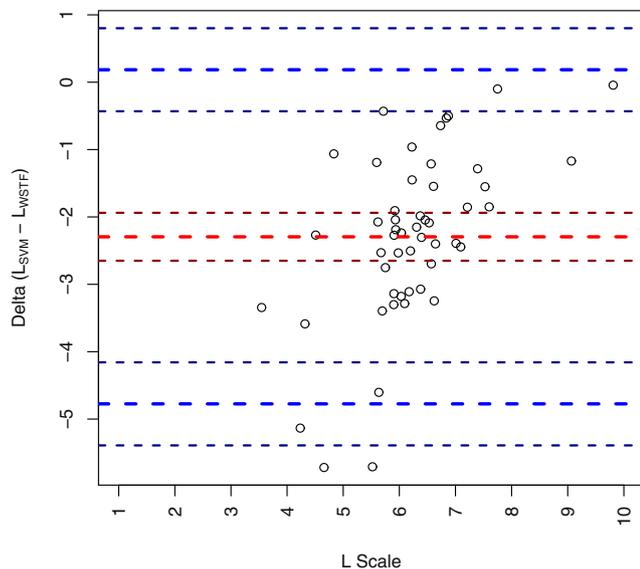


Fig. 3 Bland–Altman plot (SVM–WSTF), comparison of $n = 51$ samples. Mean of differences $\bar{d} = -2.295$ (95% CI [-2.651, -1.939]), lower LOA $\bar{d} - 1.96s = -4.774$ (95% CI [-5.391, -4.158]), upper LOA $\bar{d} + 1.96s = 0.185$ (95% CI [-0.431, 0.801])

Figure 4 depicts a direct comparison between the “classic” readability metrics ($L_{WSTF} - L_{FRE}$). A total of two data points are found outside the respective LOA which ranges from 0.025 to 1.119, reflecting merely 11% of the L scale. By contrast to the plots in Figs. 2 and 3, the direct comparison of WSTF and FRE suggests that both readability formulas can be applied interchangeably for the assessed text material on various cancer types. Pairwise delta values of all direct comparisons are given in Table 3 for all 51 booklets.

When compared as a triplet, eight out of 51 booklets are assigned the same class label, either M, D, or VD (see Table 2). Only two of the booklets were rated as “very difficult to read” by all applied readability formulas: Breast_04 and Prostate_08. In contrast, none of the booklets was rated as “easy or very easy to read” by FRE and WSTF. But seven booklets were classified as E or VE by SVM. In our assessment, we found only four booklets (7.84%) in which all three metrics showed agreement, that the related text material was adequate for laymen with basic medical background knowledge (breast_07, breast_14, melanoma_03, pancreatic_02).

In the case of four booklets, the classification result of the SVM resulted in a higher absolute score than the corresponding result of the FRE (see Table 3). In detail, it was one booklet about breast cancer from a cancer society (breast_02), two booklets about prostate cancer, one from a cancer society (prostate_02), one expert guideline which was found on a webpage of a self-help group, and finally one booklet about melanoma (melanoma_02) from a cancer society. Yet, no booklets resulted in a higher SVM value compared to the WSTF value.

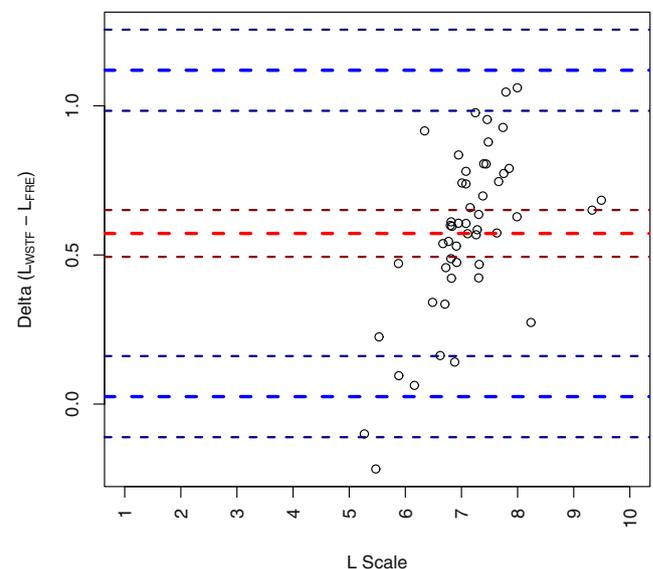


Fig. 4 Bland–Altman plot (WSTF–FRE), comparison of $n = 51$ samples. Mean of differences $\bar{d} = 0.572$ (95% CI [0.494, 0.651]), lower LOA $\bar{d} - 1.96s = 0.025$ (95% CI [-0.111, 0.161]), upper LOA $\bar{d} + 1.96s = 1.119$ (95% CI [0.983, 1.255])

Table 3 Listing of pairwise differences (denoted as δ) of all metrics under comparison

| Booklet topic | $\delta_{\text{SVM-FRE}}$ | $\delta_{\text{SVM-WSTF}}$ | $\delta_{\text{WSTF-FRE}}$ |
|------------------|---------------------------|----------------------------|----------------------------|
| Breast_01 | -1.58 | -2.04 | 0.46 |
| Breast_02 | 0.07 | -0.53 | 0.6 |
| Breast_03 | -2.83 | -3.3 | 0.47 |
| Breast_04 | -0.52 | -1.17 | 0.65 |
| Breast_05 | -0.54 | -1.29 | 0.75 |
| Breast_06 | -2.45 | -3.24 | 0.79 |
| Breast_07 | -0.34 | -0.43 | 0.09 |
| Breast_08 | -5.3 | -5.72 | 0.42 |
| Breast_09 | -1.13 | -1.19 | 0.06 |
| Breast_10 | -2.41 | -3.11 | 0.7 |
| Breast_11 | -1.55 | -2.15 | 0.6 |
| Breast_12 | -1.47 | -2.39 | 0.92 |
| Breast_13 | -1.76 | -2.23 | 0.47 |
| Breast_14 | -1.28 | -1.06 | -0.22 |
| Prostate_01 | -0.92 | -1.55 | 0.63 |
| Prostate_02 | 0.11 | -0.5 | 0.61 |
| Prostate_03 | -1.74 | -2.31 | 0.57 |
| Prostate_04 | -4.21 | -5.13 | 0.92 |
| Prostate_05 | -0.79 | -1.85 | 1.06 |
| Prostate_06 | -1.31 | -1.45 | 0.14 |
| Prostate_07 | -1.08 | -1.85 | 0.77 |
| Prostate_08 | 0.64 | -0.04 | 0.68 |
| Lung_01 | -1.73 | -2.27 | 0.54 |
| Lung_02 | -2.5 | -3.08 | 0.58 |
| Lung_04 | -3.11 | -3.58 | 0.47 |
| Lung_05 | -1.38 | -2.04 | 0.66 |
| Colorectal_01 | -1.99 | -2.53 | 0.54 |
| Colorectal_02 | -0.8 | -0.96 | 0.16 |
| Colorectal_03 | -1.5 | -2.08 | 0.58 |
| Colorectal_04 | -1.77 | -2.19 | 0.42 |
| Colorectal_05 | -1.6 | -2.4 | 0.8 |
| Colorectal_06 | -1.77 | -2.51 | 0.74 |
| Colorectal_07 | -1.93 | -2.54 | 0.61 |
| Bladder_01 | -0.81 | -1.55 | 0.74 |
| Endometrial_01 | -2.55 | -3.18 | 0.63 |
| Endometrial_02 | -5.43 | -5.7 | 0.27 |
| Endometrial_03 | -2.82 | -3.39 | 0.57 |
| Endometrial_04 | -1.82 | -2.7 | 0.88 |
| Melanoma_01 | -1.74 | -2.08 | 0.34 |
| Melanoma_02 | 0.71 | -0.1 | 0.81 |
| Melanoma_03 | -3.44 | -3.34 | -0.1 |
| Melanoma_04 | -1.4 | -2.45 | 1.05 |
| Melanoma_05 | -2.31 | -3.29 | 0.98 |
| Oropharyngeal_01 | -2.15 | -2.75 | 0.6 |
| Oropharyngeal_02 | -3.65 | -4.61 | 0.96 |
| Oropharyngeal_03 | -0.16 | -0.65 | 0.49 |
| Oropharyngeal_04 | -0.68 | -1.21 | 0.53 |
| Pancreatic_01 | -2.36 | -3.14 | 0.78 |

Table 3 (continued)

| Booklet topic | $\delta_{\text{SVM-FRE}}$ | $\delta_{\text{SVM-WSTF}}$ | $\delta_{\text{WSTF-FRE}}$ |
|---------------|---------------------------|----------------------------|----------------------------|
| Pancreatic_02 | -2.04 | -2.27 | 0.23 |
| Pancreatic_03 | -1.57 | -1.9 | 0.33 |
| Pancreatic_04 | -1.15 | -1.98 | 0.83 |
| Mean | -1.72 | -2.3 | 0.58 |

Discussion and Conclusion

Discussion

A high-quality health information must not only include the best available external evidence; it must also be readable and correspond to patient preferences [24]. In order to comply with these requirements, the application of easy language is essential [25]. In addition, patients show a high level of satisfaction when information was written in a simple language and was available prior to therapy [26].

However, our findings suggest that the majority of the 51 booklets (92.16%) is hard to read (see Table 2). Thus, it can be difficult to understand for the intended group of readers. The primary reason for this finding is attributable to the scores of the sentence-structure-based FRE and WSTF metrics and the corresponding classification into difficult or very difficult texts. This corresponds to the results of other authors, which also report on high readability levels [27–31]. Moreover, to understand a medical text, the vocabulary used by authors is of high relevance for the understandability and might be even more decisive than the sentence structure [32].

With our evaluation of patient booklets on cancer, we demonstrated that a specifically trained support vector machine is able to analyze text material with respect to its vocabulary. By contrast to established instruments, i.e., FRE and WSTF, for most analyzed documents, the SVM instrument resulted in lower readability scores, i.e., text documents are easier to comprehend for readers.

In terms of text production, a selection of one of the three instruments may also depend on the aims of the authors. An author, who is already experienced in writing text material for cancer patients, could benefit by a quick check for expert vocabulary for a certain paragraph or the whole document. In contrast, a less experienced author might be more interested in checking sentence structures. Our study demonstrates that both needs can be satisfied automatically during the text production phase, even though we conducted a post-publish analysis.

Several limitations apply for the study setting. The PDF documents were automatically converted to Microsoft Word documents via text recognition. Furthermore, disturbance artifacts—e.g., different kinds of hyphens, markup language (e.g., XML or HTML tags), mis-encoded characters from

different encoding schemes, etc.—may still be included in the extracted raw text material. The FRE and WSTF metrics are computed mainly on the mean sentence length, the mean number of syllables per word, and language-specific weighting factors. However, detecting syllables is not a trivial task for the German language and does not work reliably in some rare circumstances [33]. For this reason, the computed FRE or WSTF scores are influenced by the aforementioned inaccuracies, which constitutes a systematic bias; yet, this affects all NLP tools for analysis on German text material.

Furthermore, it has to be considered for every applied instrument that an exclusive computation of readability disregards individual knowledge and motivation of readers [13]. Aspects related to illustration and design were not included in our analysis. Consequently, the suitability of health information cannot only be judged on the basis of readability of information [11].

Conclusion

Methods from the field of Machine Learning may add valuable information for authors of cancer brochures, thus, complementing established readability assessment methods, such as FRE and WSTF. In addition to these metrics, the readability of patient information should be analyzed in terms of its vocabulary, e.g., via SVM-based metrics. The results of the presented study show that the automatic computation of related scores can support authors either (a) during the text production process or (b) at the final quality assurance stage. Yet, our Bland-Altman analysis revealed that the computed SVM scores cannot be interchanged with FRE or WSTF as they express different aspects of written text. This is confirmed by the results of the conducted *t* tests ($P < 0.001$). However, both aspects of text analysis, i.e., sentence structure vs. vocabulary, can be examined by booklet authors to refine their texts linguistically, especially with respect to the intended target group of readers.

The authors of the study recommend the use of both vocabulary and established readability assessment instruments as a supportive measure to ensure and provide understandable health information.

Practice Implications

Authors of health information can benefit by automatic analysis tooling which provides feedback on the readability of written text. This includes both aspects of readability, i.e., sentence structure complexity and the expert level of the used vocabulary. For 51 German cancer brochures, our study reveals that a significant difference between both aspects exists. Authors should therefore re-evaluate existing brochures and reduce sentence complexity, as our study shows higher scores for FRE and WSTF when compared to the vocabulary-based

metric (see Figs. 2 and 3). Moreover, the results of the study suggest that authors should, however, check their text material with all of the aforementioned instruments during the writing process.

Professionals or institutions looking for information material can benefit by a direct comparison of several brochures (see Table 2), as they do not have to compare cancer booklets by themselves, thus saving time. But more importantly, it helps to decide which brochures are easy—or at least easier—to understand and could be recommended to patients in case of a certain tumor type.

Compliance with Ethical Standards

Conflict of Interest The authors declare that they have no conflict of interest.

Ethical Approval This article does not contain any studies with human participants performed by any of the authors.

References

- Berkman ND, Sheridan SL, Donahue KE, Halpern DJ, Crotty K (2011) Low health literacy and health outcomes: an updated systematic review. *Ann Intern Med* 155:97–107. <https://doi.org/10.7326/0003-4819-155-2-201107190-00005>
- Coulter A, Ellins J (2007) Effectiveness of strategies for informing, educating, and involving patients. *BMJ* 335:24–27. <https://doi.org/10.1136/bmj.39246.581169.80>
- Dirmaier J, Härter M (2012) Partizipative Entscheidungsfindung: Patientenbeteiligung bei Behandlungsentscheidungen in der medizinischen Versorgung [Shared decision making: patient involvement in treatment decisions in medical care]. *BARMER GEK Gesundheitswesen Aktuell* 212–235
- Gorgojo L, Harris M, Garcia-Lopez E, Group CW (2012) National cancer control programmes: analysis of primary data from questionnaires Final Prelim Rep Eur Partnersh Action Cancer EPAACURL <http://www.epaac.eu/national-cancer-plans> Accessed Novemb 2015
- Bundesministerium für Gesundheit (2012), Nationaler Krebsplan - Handlungsfelder, Ziele und Umsetzungsempfehlungen [National cancer plan—fields of action, objectives and implementation of recommendations]. Druckerei im Bundesministerium für Arbeit und Soziales, Berlin
- Huebner J, Muenstedt K, Prott FJ, Stoll C, Micke O, Buentzel J, Muecke R, Senf B (2014) Online survey of patients with breast cancer on complementary and alternative medicine. *Breast Care Basel Switz* 9:60–63. <https://doi.org/10.1159/000360381>
- Paul M, Davey B, Senf B, Stoll C, Muenstedt K, Mücke R, Micke O, Prott FJ, Buentzel J, Hübner J (2013) Patients with advanced cancer and their usage of complementary and alternative medicine. *J Cancer Res Clin Oncol* 139:1515–1522. <https://doi.org/10.1007/s00432-013-1460-y>
- Keinki C, Seilacher E, Ebel M, Ruetters D, Kessler I, Stellamanns J, Rudolph I, Huebner J (2015) Information needs of cancer patients and perception of impact of the disease, of self-efficacy, and locus of control. *J Cancer Educ Off J Am Assoc Cancer Educ* 31:610–616. <https://doi.org/10.1007/s13187-015-0860-x>
- Huebner J, Micke O, Muecke R, Buentzel J, Prott FJ, Kleeberg U, Senf B, Muenstedt K, PRIO (Working Group Prevention and Integrative Oncology of the German Cancer Society) (2014) User

- rate of complementary and alternative medicine (CAM) of patients visiting a counseling facility for CAM of a German comprehensive cancer center. *Anticancer Res* 34:943–948
10. Flesch R (1948) A new readability yardstick. *J Appl Psychol* 32: 221–233
 11. Friedman DB, Hoffman-Goetz L (2006) A systematic review of readability and comprehension instruments used for print and web-based cancer information. *Health Educ Behav off Publ Soc public Health Educ* 33:352–373. <https://doi.org/10.1177/1090198105277329>
 12. Amstad T (1978) *Wie verständlich sind unsere Zeitungen? [How readable are our newspapers?]*. Doctoral thesis, Universität Zürich, Switzerland
 13. Bamberger R, Vanacek (1984) *Lesen-Verstehen-Lernen-Schreiben [Reading-Comprehension-Learning-Writing]*. Diesterweg
 14. Liebl P, Seilacher E, Koester M-J, Stellamanns J, Zell J, Hübner J (2015) What cancer patients find in the internet: the visibility of evidence-based patient information—analysis of information on German websites. *Oncol Res Treat* 38:212–218. <https://doi.org/10.1159/000381739>
 15. Keinki C, Zowalla R, Wiesner M, Koester MJ, Huebner J (2016) Understandability of patient information booklets for patients with cancer. *J Cancer Educ Off J Am Assoc Cancer Educ*. <https://doi.org/10.1007/s13187-016-1121-3>
 16. Robert Koch-Institut, die Gesellschaft der epidemiologischen Krebsregister in Deutschland e.V. (2013) *Krebs in Deutschland 2009/2010 [Cancer in Germany 2009/2010]*. Robert Koch-Institut, Berlin
 17. Leroy G, Miller T, Roseblat G, Browne A (2008) A balanced approach to health information evaluation: a vocabulary-based naïve Bayes classifier and readability formulas. *J Am Soc Inf Sci Technol* 59:1409–1419. <https://doi.org/10.1002/asi.20837>
 18. Zowalla R, Wiesner M, Pfeifer D (2014) Automatically assessing the expert degree of online health content using SVMs. *Stud Health Technol Inform* 202:48–51
 19. Joachims T (1998) Text categorization with support vector machines: learning with many relevant features. In: Nédellec C, Rouveirol C (eds) *Machine learning: ECML-98*. Springer Berlin Heidelberg, pp 137–142
 20. Salton G, Buckley C (1988) Term-weighting approaches in automatic text retrieval. *Inf Process Manag* 24:513–523. [https://doi.org/10.1016/0306-4573\(88\)90021-0](https://doi.org/10.1016/0306-4573(88)90021-0)
 21. Platt JC (1999) Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. In: *Advances in large margin classifiers*, MIT Press pp 61–74
 22. Bland JM, Altman DG (1986) Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet Lond Engl* 1:307–310
 23. Bland JM, Altman DG (1999) Measuring agreement in method comparison studies. *Stat Methods Med Res* 8:135–160. <https://doi.org/10.1177/096228029900800204>
 24. Hoefert H-W (2011) *Wandel der Patientenrolle: neue Interaktionsformen im Gesundheitswesen [Change of the patient role: new forms of interaction in healthcare]*. Hogrefe, Göttingen
 25. Fagerlin A, Zikmund-Fisher BJ, Ubel PA (2011) Helping patients decide: ten steps to better risk communication. *J Natl Cancer Inst* 103:1436–1443. <https://doi.org/10.1093/jnci/djr318>
 26. Butow P, Brindle E, McConnell D, Boakes R, Tattersall M (1998) Information booklets about cancer: factors influencing patient satisfaction and utilization. *Patient Educ Couns* 33:129–141
 27. Cooley ME, Moriarty H, Berger MS, Selm-Orr D, Coyle B, Short T (1995) Patient literacy and the readability of written cancer educational materials. *Oncol Nurs Forum* 22:1345–1351
 28. Garcia SF, Hahn EA, Jacobs EA (2010) Addressing low literacy and health literacy in clinical oncology practice. *J Support Oncol* 8: 64–69
 29. Nicholls S, Hankins M, Hooley C, Smith H (2009) A survey of the quality and accuracy of information leaflets about skin cancer and sun-protective behaviour available from UK general practices and community pharmacies. *J Eur Acad Dermatol Venereol JEADV* 23: 566–569. <https://doi.org/10.1111/j.1468-3083.2008.03017.x>
 30. Singh J (2003) Reading grade level and readability of printed cancer education materials. *Oncol Nurs Forum* 30:867–870. <https://doi.org/10.1188/03.ONF.867-870>
 31. Weintraub D, Maliski SL, Fink A, Choe S, Litwin MS (2004) Suitability of prostate cancer education materials: applying a standardized assessment tool to currently available materials. *Patient Educ Couns* 55:275–280. <https://doi.org/10.1016/j.pec.2003.10.003>
 32. Hasan M, Kotov A, Carcone A et al (2016) A study of the effectiveness of machine learning methods for classification of clinical interview fragments into a large number of categories. *J Biomed Inform* 62:21–31. <https://doi.org/10.1016/j.jbi.2016.05.004>
 33. Müller K (2001) Automatic detection of syllable boundaries combining the advantages of treebank and bracketed corpora training. In: *Proceedings of the 39th annual meeting on Association for Computational Linguistics Association for Computational Linguistics*, pp 410–417