

A readability assessment of online Parkinson's disease information

¹PR Fitzsimmons, ²BD Michael, ³JL Hulley, ⁴GO Scott

¹Specialist Registrar in Geriatric Medicine, Royal Liverpool Hospital; ²Academic Clinical Fellow, Walton Centre for Neuroscience and Neurosurgery; ³Foundation Year 2 Doctor, Royal Liverpool Hospital; ⁴Consultant Physician, Royal Liverpool Hospital, Liverpool, UK

ABSTRACT

Background: Patients increasingly use the internet to access health information. Inadequate health literacy is common and frequently limits patient comprehension of healthcare literature. We aimed to assess the readability of online consumer-orientated Parkinson's disease (PD) information using two validated measures.

Method: We identified the 100 highest ranked consumer-orientated PD webpages and determined webpage readability using the Flesch-Kincaid and Simple Measure Of Gobbledygook (SMOG) formulae.

Results: None of the webpages analysed complied with current readability guidelines. Commercial websites were significantly easier to read ($p=0.035$). The Flesch-Kincaid formula significantly underestimated reading difficulty ($p<0.0001$). Ease of reading correlated weakly with search engine ranking ($r=0.35$, $p=0.0004$).

Conclusions: Only 1% of the top 100 PD information webpages are fully comprehensible to the average adult. Simple Measure Of Gobbledygook should be the preferred measure of webpage readability. Parkinson's disease information websites require major text revision to comply with readability guidelines and to be comprehensible to the average patient.

Correspondence to P Fitzsimmons, Specialist Services for Older People, Royal Liverpool and Broadgreen University Hospitals NHS Trust, Thomas Drive, Liverpool L14 3LB, UK

tel. +44 (0)151 706 2000
e-mail pfitz@doctors.org.uk

KEYWORDS Consumer, literacy, online, Parkinson's disease, readability, webpages

DECLARATION OF INTERESTS No conflict of interests declared.

INTRODUCTION

Parkinson's disease (PD) is a common, progressive and incurable chronic neurodegenerative condition with a wide variety of motor and non-motor features, often requiring complex treatment regimes. As with all chronic diseases, patient education is essential to improve patient understanding of the condition and its treatment. In addition to the information patients receive from healthcare professionals, many patients seek information independently, with the internet being one commonly used source of information.

An abundance of PD-related information is available on the internet, with a Google search for 'Parkinson's disease' returning more than 4,630,000 webpages. Increasing numbers of patients access healthcare information via the internet, with an estimated 98 million Americans searching for online health information in 2006.¹ Many estimates of the prevalence of internet use for seeking healthcare information appear, without peer review, online or as private commercial releases. However, several studies published in peer-reviewed journals have suggested that 40–50% of the total population use the internet for health information purposes, falling to around 30% in the over-60s age group.^{2–4}

No published estimates of internet usage by PD patients are available. However, PD patients with impulse control disorders demonstrate pathological internet use (including

pathological internet gambling), suggesting that at least part of the PD population regularly uses the internet.^{5,6}

Although the internet is a popular and convenient source of consumer-orientated healthcare information, these resources are only useful if the consumer is able to comprehend the information presented.

Health literacy is defined as 'the ability to perform basic reading and numerical tasks required to function in the healthcare environment'. Poor health literacy is associated with poor compliance with treatment and poor disease knowledge and may increase the risk of hospitalisation.⁷ Reading ability is an integral component of health literacy; patients with inadequate literacy have difficulty reading and understanding material written at the sixth-grade level (ages 11–12), while those with marginal literacy have difficulty comprehending material written at the tenth-grade level (ages 15–16).⁸ A 2007 report estimated that 34% of US adults had only basic or below basic general document reading literacy; this figure increased to 61% in the over-65 age group.⁹ National surveys in the UK show that 16% of the population has low general literacy skills.¹⁰

Little is known about the degree of health literacy in the UK population, with small studies suggesting that inadequate health literacy is common.¹¹ A large study in a US publicly funded healthcare setting has confirmed that inadequate health literacy is a particular problem

for older patients, with one-third of patients over the age of 65 having inadequate or marginal health literacy. The prevalence of inadequate literacy increased with age, 15.6% for ages 60–65 and 58% for ages >85 years and when adjusted for education and cognitive impairment. The adjusted odds ratio for having inadequate or marginal health literacy for patients aged over 85 years was 8.62 (95% confidence interval [CI], 5.55–13.38);⁸ the majority of patients with PD fall into this older age group. To date we are not aware of any large-scale studies investigating literacy levels in patients with PD.

In view of these high levels of inadequate literacy the National Work Group on Literacy and Health, along with the US Department of Health and Human Services (USDHHS), recommends that to ensure adequate comprehension, patient-orientated literature should be written at or below the sixth-grade level (again, ages 11–12).^{12,13} The USDHHS classification system categorises material as 'easy to read' only if written at or below a sixth-grade level. Material between the seventh-grade and ninth-grade levels is rated as 'average difficulty', and material above the ninth-grade level is regarded as 'difficult'.¹³ The UK currently has no quantitative guidelines regarding levels of readability for patient-orientated literature; hence the US guidelines outlined above have been taken as the recommended standard in terms of minimum readability levels for the purposes of this study.

A previous small study of 27 websites using a single measure of readability has suggested that online consumer-orientated PD information exceeds both readability recommendations and the reading abilities of the average adult with an average grade level of 12 (i.e. ages 17–18).¹⁴ This study aimed to analyse the reading level of a large sample of consumer-oriented online PD information websites using two validated readability measures.

This study did not attempt to make an assessment of the quality, accuracy or relevance of consumer-orientated Parkinson's information on the internet. These factors are also major determinants of the usefulness of online information.

This is the largest study of readability in consumer-orientated PD literature of which we are aware to date and the first to use the accurate Simple Measure Of Gobbledygook (SMOG) readability formula.

METHODS

Sampling method

We analysed the text content of the 100 highest ranked consumer-orientated webpages (rather than whole websites). This sample was selected on the basis of Google search ranking and consumer orientation. A full description of the sampling method is given below.

We performed a Google search on 19 October 2009 for the term 'Parkinson's disease', which returned 4,630,000 English-language webpage results. Google was selected as it is the most popular search engine, accounting for more than 90% of consumer health-related search activity in 2009, with a number of other search engines making up the remaining 10%.¹⁵

Starting with the highest ranking non-sponsored webpage link, each of the webpage results returned by this Google search was reviewed by a single researcher (JL Hulley). Initially, a decision was made as to whether the webpage contained consumer-orientated healthcare information (aimed primarily at lay persons) or was aimed at healthcare professionals (journal articles, textbooks and so on). Only consumer-orientated webpages containing healthcare information regarding PD were included for further assessment.

Each consumer-orientated webpage was then classified by the following criteria:

1. Commercial or not-for-profit, with not-for-profit being defined as the website of a registered charity, educational institution or governmental organisation.
2. Website country of origin – classified as UK, USA or other country.

This classification for each webpage was determined from information contained in the 'about us' section of the hosting website. This process was repeated for each webpage search result in turn, until 100 webpages containing consumer-orientated healthcare information were identified. A sample size of 100 webpages was selected as it was felt unlikely that most patients would click on more than 100 consecutive links from a Google search.

Where duplicate webpages were identified by the search engine, only the highest ranked webpage was included in the analysis. Unique webpages from the same website were included, providing the information contained in the webpage did not duplicate previously analysed text content.

A second researcher (PR Fitzsimmons) independently reviewed the webpages to confirm consumer orientation and screen for duplicate results.

Readability assessment

Full webpage body text content was extracted into Microsoft® Word 2007 and readability assessed using two validated measures: the Flesch-Kincaid Grade Level and SMOG formulae. These represent the most commonly used readability measure (Flesch-Kincaid) and the most exacting (SMOG). Titles, subtitles, references, weblinks and advertising text were excluded from the analysis, with only body text and bullet point text included.

TABLE 1 Distribution of webpages by USDHHS classification and grade level

USDHHS classification	Grade level	Flesch-Kincaid (%Articles)	SMOG (%Articles)
Easy	4th–6th grade	4%	0%
Average	7th–9th grade	9%	1%
Difficult	10th–12th grade	27%	10%
	>12th grade	60%	89%

Note: 4–6th grade = British school age equivalent of 9–12 years.
 7–9th grade = British school age equivalent of 12–15 years.
 10–12th grade = British school age equivalent of 15–17 years.
 >12th grade = British school age equivalent of 17 years and over.

The Flesch-Kincaid Grade Level formula is well established and conveniently is often included as a function in word-processing packages such as MS Word. The Flesch-Kincaid formula produces a score that estimates the grade level (number of years of education) required to understand the text.¹⁶ Despite the formula's widespread use, validation studies have suggested it may be of limited accuracy when grading low-level literacy materials.¹⁷ The formula is expressed as $[(0.39 \times \text{ASL}) + (11.8 \times \text{ASW}) - 15.59]$, where ASL is the average sentence length (number of words divided by number of sentences) and ASW is the average syllables per word (number of syllables divided by number of words).¹⁶ Previous versions of MS Word limited the maximum calculated Flesch-Kincaid grade to 12; this formula error has been corrected in the 2007 version.

Simple Measure Of Gobbledygook is a more exacting measure of readability, and accurately grades for the grade level required for complete text comprehension (whereas the Flesch-Kincaid formula grades for less than complete comprehension), demonstrating strong correlation ($r=0.985$) with the required reading level in validation studies.¹⁸ The SMOG formula, in brief terms, consists of counting the words of three or more syllables in three ten-sentence samples, calculating the count's square root and adding three to obtain the grade level.¹⁸

The SMOG level was calculated using an online calculator produced by Professor Harry McLaughlin, the original developer of the SMOG formula (<http://www.harrymclaughlin.com/SMOG.htm>).

For the purposes of this study the reading grade level required for the complete comprehension of text material (i.e. the SMOG grade level) was taken as the gold standard readability measure.

Statistical analysis

Reading level data were categorised in line with USDHHS standards and analysed in GraphPad Prism 4.0 (GraphPad Software, San Diego, USA). The D'Agostino & Pearson omnibus normality test demonstrated that SMOG and Flesch-Kincaid Grade Level data followed a non-parametric distribution. As such, Wilcoxon's signed rank test was used to compare paired SMOG and Flesch-Kincaid Grade Level values. Mann-Whitney and Kruskal-Wallis tests were used to compare SMOG grade levels in webpages of varying commercial nature (commercial and not-for profit) and nationality (UK, USA and other) respectively. Relationships between webpage ranking, word count and readability were analysed using Spearman's correlation. Statistical significance was set at 0.05.

RESULTS

Our analysis included 100 unique webpages. The mean webpage word count was 1,450 (95% CI 1,081–1,818, max. 11,148, min. 175). Of the 100 webpages analysed, 45% were run as commercial interests and 55% on a not-for-profit basis. More than half (57%) of websites were based in the USA, with 34% based in the UK and 9% based in other countries. Table 1 shows the distribution of webpages by USDHHS classification and grade level for the Flesch-Kincaid and SMOG formulae.

Mean Flesch-Kincaid grade level was 12.1 (95% CI 11.5–12.7), mean SMOG grade 14.6 (95% CI 14.2–15.1). A significant correlation was seen between the grade levels produced by the formulae ($r=0.52$, $p<0.0001$). The Flesch-Kincaid formula significantly underestimated reading difficulty compared with the gold standard SMOG formula, with a mean underestimation of 2.52 grades (95% CI 1.97–3.07), $p<0.0001$.

A weak but significant correlation was observed between SMOG readability and search engine ranking ($r=0.35$, $p=0.0004$).

No significant correlation was observed between SMOG readability and webpage word count ($r=0.004$, $p=0.96$).

Commercial websites were significantly easier to read than not-for-profit websites, with a mean commercial SMOG of 14.1 (95% CI 13.46–14.73) and a mean non-commercial SMOG of 15.1 (95% CI 14.5–15.65), $p=0.035$.

No significant difference in SMOG grade level was observed between webpages taken from websites based in different countries. The mean SMOG for US webpages was 14.3 (95% CI 13.7–14.8), the mean SMOG for UK webpages was 15.3 (95% CI 14.4–16.1) and the mean SMOG for webpages of other nationality was 14.5 (95% CI 13.5–15.5), $p=0.25$.

CONCLUSIONS

While the internet is a rapidly growing and highly convenient source of information for patients, irrespective of its quality, this information can only be helpful if patients are able to understand it.

We found that the majority (60–89%) of consumer-orientated PD information webpages are written at above the 12th grade level and that none complied with the USDHHS maximum recommended sixth-grade level when readability was measured with SMOG. Our findings are in keeping with a previous smaller study of readability in online PD literature.¹⁴ Excessive levels of reading difficulty have long been noted in printed patient information.¹⁹ More recently, similar levels of poor readability have been demonstrated in websites concerning other areas of medicine, including mental health,²⁰ epilepsy,²¹ general internal medicine²² and orthopaedic surgery.²³

The readability levels of the webpages analysed contrast sharply with the observed reading abilities of patients, with the average adult reading between the seventh and eighth grade,²⁴ inferring only 1% of the webpages included in this study would be fully comprehensible to the average adult when measured by SMOG.

We found a small but significant difference in readability of commercial and not-for-profit websites. The reason for this is unclear, but we hypothesise that this may be due to different writing styles in each of these sectors, with commercial authors and editors aiming to ensure the widest possible commercial audience by minimising reading difficulty.

Readability levels were similarly poor in webpages produced in the USA, UK and other countries, suggesting that readability is likely to limit the ability of patients to comprehend the contents of locally produced webpages in most developed countries.

Webpage length was not significantly associated with readability, suggesting that reading difficulty of these websites is independent of word count and that webpages with low word counts were of a similar level of reading difficulty to longer webpages.

We observed a weak but significant correlation between reading difficulty and Google search engine ranking. The reason for this is unclear but suggests that higher ranked webpages may be easier to read.

This is the first study of readability in consumer-orientated PD literature to use the SMOG formula which accurately grades for complete comprehension of reading materials. The Flesch-Kincaid formula may appear to be an attractive method of determining readability, primarily because of its incorporation in commercial word-processing software. However, compared with the gold standard readability measure grading for complete comprehension of text, Flesch-Kincaid significantly underestimated reading difficulty, presumably due to its grading for less than complete comprehension of written material. As such, we recommend that SMOG should be the preferred measure of readability when evaluating consumer-orientated healthcare material (several online SMOG calculators are now available).

Our study is limited in that we did not undertake an assessment of the quality, accuracy or relevance of the information presented. As such, we are unable to comment as to whether information quality is preserved in webpages written at low literacy levels. However, while providing good quality information is of critical importance, its positive impact is negligible if that information is presented in an unintelligible format.

By reporting the discrepancy between levels of website readability and population estimates of literacy we hope to encourage compliance with readability guidelines and thus improve accessibility to online information for patients with PD.

Website editors should consider introducing minimum readability policies based on USDHHS guidelines and also consider routine monitoring of content readability with SMOG to increase the accessibility and ease of comprehension of online consumer-orientated healthcare information.

In summary, the majority of consumer-orientated PD information websites require major text revision to comply with readability guidelines and to be comprehensible to the average patient.

REFERENCES

- 1 Sultz HA. *Health care USA: understanding its organization and delivery*. 5th ed. Sudbury, MA: Jones and Bartlett; 2006. p.64–5.
- 2 Diaz JA, Griffith RA, Ng JJ et al. Patients' use of the internet for medical information. *J Gen Intern Med* 2002; 17:180–5. doi:10.1046/j.1525-1497.2002.10603.x
- 3 Baker L, Wagner TH, Singer S et al. Use of the internet and e-mail for health care information results from a national survey. *JAMA* 2003; 289:2400–6.
- 4 Flynn KE, Smith MA, Freese J. When do older adults turn to the internet for health information? Findings from the Wisconsin Longitudinal Study. *J Gen Intern Med* 2006; 21:295–301. doi:10.1111/j.1525-1497.2006.00622.x
- 5 Wong SH, Cowen Z, Allen EA et al. Internet gambling and other pathological gambling in Parkinson's disease: a case series. *Mov Disord* 2007; 22:591–3. doi:10.1002/mds.21334

- 6 Larner AJ. Medical hazards of the internet: gambling in Parkinson's disease. *Mov Disord* 2006; 21:1789. doi:10.1002/mds.21057
- 7 Ad Hoc Committee on Health Literacy for the Council on Scientific Affairs, American Medical Association. Health literacy: report of the Council on Scientific Affairs. *JAMA* 1999; 281:552-7.
- 8 Gazmararian JA, Baker DW, Williams MV et al. Health literacy among medicare enrollees in a managed care organization. *JAMA* 1999; 281:545-51.
- 9 US Department of Education, National Center for Education Statistics. *The condition of education 2007*. Washington, DC: US Government Printing Office; 2007.
- 10 Williams J, Clemens S, Oleinikova J et al. *The skills for life survey: a national needs and impact survey of literacy, numeracy and ICT skills*. London: Department for Education and Skills; 2003.
- 11 Ibrahim SY, Reid F, Shaw A et al. Validation of a health literacy screening tool (REALM) in a UK population with coronary heart disease. *J Public Health (Oxf)* 2008; 30:449-55. doi:10.1093/pubmed/fdn059
- 12 The National Work Group on Literacy and Health. Communicating with patients who have limited literacy skills. *J Fam Pract* 1998; 46:168-76.
- 13 US Department of Health and Human Services. *Saying it clearly*. Washington, DC: US Government Printing Office; 2000.
- 14 Barton BR. Parkinson's disease on the internet: an evaluation of the readability, quality, and technical content of patient information. *Mov Disord* 2009; 24(Suppl 1):S119-S120.
- 15 Private Healthcare UK. *WebWatch 2009: an analysis of web activity in the health and medical sector*. Berkhamsted: Intuition Communication; 2009. Available from: <http://www.privatehealth.co.uk/aboutus/webwatch-2009>
- 16 Flesch R. *How to write plain English*. New York: Harper and Row; 1979.
- 17 McLaughlin GH. Temptations of the Flesch. *Instr Sci* 1974; 2:367-84. doi:10.1007/BF00123459
- 18 McLaughlin GH. SMOG grading: a new readability formula. *J Reading* 1969; 12:639-46.
- 19 Davis TC, Crouch MA, Wills G et al. The gap between patient reading comprehension and the readability of patient education materials. *J Fam Pract* 1990; 31:533-8.
- 20 Kalk NJ, Pothier DD. Patient information on schizophrenia on the internet. *Psychiatrist* 2008; 32:409-11. doi: 10.1192/pb.bp.107.017004
- 21 Elliott JO, Shneker BF. A health literacy assessment of the epilepsy.com website. *Seizure* 2009; 18:434-9. doi:10.1016/j.seizure.2009.04.003
- 22 Walsh TM, Volsko TA. Readability assessment of internet-based consumer health information. *Respir Care* 2008; 53:1310-5.
- 23 Sabharwal S, Badarudeen S, Unes Kunju S. Readability of online patient education materials from the AAOS web site. *Clin Orthop Relat Res* 2008; 466:1245-50. doi:10.1007/s11999-008-0193-8
- 24 National Center for Education Statistics. National Assessment of Adult Literacy (NAAL). Washington, DC: National Center for Education Statistics; 2003. Available from: <http://nces.ed.gov/naal>

UPCOMING SYMPOSIA

RESPIRATORY MEDICINE

Friday 4 March

- Challenges in the clinic
- The state of the art
- The science behind respiratory disease
- Respiratory problems in acute medicine

PALLIATIVE CARE

(Joint RCPE/RCGP symposium)

Monday 28 March

- Paediatric palliative care
- Inflammation and cancer symptoms
- Non-malignant disease
- Integration of non-physical and physical symptoms

PUBLIC HEALTH POLICY DEVELOPMENT AND IMPLEMENTATION: COMPETING PRESSURE IN THE FORMULATION OF STATE HEALTH POLICY

Thursday 10 March

- Evidence and policy: general principles
- Policy area – cardiovascular
- Policy area – alcohol
- Policy area – obesity
- Policy area – public mental health
- Plenary discussion on links between evidence and policy in Scottish health policy and the extent to which public health policymaking has diverged in the UK

For more information, please visit: <http://events.rcpe.ac.uk>