Patient Education and Counseling 113 (2023) 107762

Contents lists available at ScienceDirect

Patient Education and Counseling

journal homepage: www.journals.elsevier.com/patient-education-and-counseling

Development and testing of the health information website evaluation tool on neck pain websites – An analysis of reliability, validity, and utility

Luke Zubiena, Olivia Lewin, Robert Coleman, James Phezulu, Gbemisola Ogunfiditimi, Tiffany Blackburn, Leonard Joseph *

School of Sports and Health Science, University of Brighton, Eastbourne, East Sussex, United Kingdom

ARTICLE INFO

Keywords:
Online health information
Tool validation
Tool reliability
Tool utility
Quality evaluation

ABSTRACT

Objective: Online health information contributes to patient education and knowledge on disease management. The aims of this study were to design the Health Information Website Evaluation Tool (HIWET) to evaluate the quality of online information, and to investigate the reliability, validity, and utility of HIWET.
Methods: HIWET was developed by a literature search and small-scale pilot testing. Upon development, psychometric properties of HIWET were evaluated on 20 neck pain websites. Reliability was analysed using Intra class correlation coefficient (ICC). Validity was analysed using Pearson and Spearman correlation coefficients. Utility was analysed using an independent samples t-test.
Results: HIWET demonstrated excellent intra-rater reliability (0.94 (0.98–0.99), $p < .001$) and fair inter-rater reliability (0.55 (0.88–0.10), $p = .04$). HIWET demonstrated validity with strong correlation against DISCERN ($r = 0.656$, $n = 20$, $p = .002$) and LIDA ($r = 0.564$, $n = 20$, $p = 0.010$). HIWET was time-efficient when compared to three comparison tools combined.
Conclusion: HIWET is a reliable and valid tool for evaluating the qualities of online health information. *Practical implications:* HIWET has the advantages of being a simple, quick to use and freely accessible tool. It can be implemented into clinical practice, education, and research to evaluate quality of online health information.

1. Introduction

People commonly seek health information from online sources as part of patient education prior to seeking healthcare advice [1]. In 2018, approximately 61 % of people in the United Kingdom (UK) searched for online health information to help make treatment decisions, Supplement information from clinicians, and self-manage a condition [1]. Furthermore, online health information sources were recommended to 65.2 % of musculoskeletal physiotherapy patients during consultations [2]. Access to online health information can be linked to enhanced consumer engagement in healthcare and improved health outcomes, but may also be associated with adverse outcomes resulting from incorrect information which may lead to increased anxiety, and tension between patient and clinician [3]. A major component of online health information quality is whether it is evidence-based, however, much of the current online health information is not informed by good quality evidence [4]. With this variability in quality comes greater risk of health-related misinformation, which could have adverse effects on health and overall health outcomes [5].

The spectrum of quality evaluation of online health information can be divided into the following seven domains; accuracy, completeness and comprehensiveness, technical elements, design and aesthetics, usability, accessibility and lastly, readability [6]. Accuracy is the extent that information complies with clinical guidelines; Completeness and Comprehensiveness ensures websites provide information covering all the main points; Technical Elements refers to the trustworthiness of information, and includes who authored the content; Design and Aesthetics relates to visual elements that facilitate consumer understanding; Usability is defined as website ease of use; Accessibility judges if the website is appropriate for those who are visually, cognitively, or hearing impaired; and Readability assesses how easily the text can be read [6]. Establishing objectively what quality means in relation to online health

Abbreviations: FRES, Flesch-Kincaid Readability Ease Score; HIWET, Health Information Website Evaluation Tool; ICC, Intraclass Correlation Coefficient; LIDA, MinerLIDAvation.

* Correspondence to: School of Health Sciences, University of Brighton, Robert Dodd Building, 49 Darley Road, Eastbourne, East Sussex BN20 7UR, United Kingdom.
E-mail address: leonardjoseph85@hotmail.com (L. Joseph).

https://doi.org/10.1016/j.pec.2023.107762
Received 28 January 2022; Received in revised form 30 August 2022; Accepted 14 April 2023
Available online 15 April 2023

information is important because previous tools evaluate only some aspects of quality with varying description when creating criteria for quality assessment, which lead to differing views on how quality can be evaluated [7]. The seven quality domains used in the Health Information Website Evaluation Tool (HIWET) may provide a standardised approach to evaluating online health information [6].

The variable nature of online health information quality means that the development of tools that can evaluate health information is essential [7]. Website evaluation tools can be used as down-stream interventions that can help all stakeholders in healthcare evaluate online health information [8]. DISCERN (not an acronym), a valid and reliable 16-item questionnaire developed in partnership with the National Health Service [8,9], was designed to help users judge the quality of written health information. However, although it was designed to be used by all stakeholders, it can be too complex and time consuming to be used by both patients and clinicians [7]. DISCERN, which assesses the domains of Accuracy, and Completeness and Comprehensiveness, does not assess all seven domains of quality evaluation [6,10]. Other examples of website evaluation tools include MinervaLIDAtion (LIDA) [10], which is a valid appraisal instrument that uses a four-point scale, that assess Usability [11], and the Flesch reading-ease score (FRES), which evaluates Readability, is a valid formula that can help users determine the reading level of text [12,13]. Readability is important to consider when evaluating online health information because of the potential of a website to directly impact on consumer understanding [14]. Health literacy is a factor in consumer understanding, and is described as an individual's ability to attain, process, and understand the fundamental health information needed to make appropriate decisions regarding either their own health, or the health of loved ones [15]. Poor health literacy can result in sub-optimal disease knowledge and reduced compliance with treatment, and so online health information should be written in a simple and readable manner that can be understood by as much of the target audience as possible [16].

In contemporary practice, one has to use three different tools such as DISCERN, LIDA and FRES to appraise the different domains to evaluate the quality of online information. There are currently no existing valid and reliable tools that can evaluate all the seven identified spectrums of domains when appraising online health information quality [7]. This study looked to address this gap by developing a new tool, HIWET, that could evaluate a broad spectrum of online health information quality, and might be used by all stakeholders in healthcare. The main aim of this study was to develop a preliminary design of HIWET, and to investigate the reliability, validity, and utility of HIWET when evaluating the quality of health information websites.

2. Methods

2.1. Study design

This study comprised of two parts. The first part of the study involved the development phase of HIWET which included the literature review and designing of HIWET, as well as the small-scale pilot testing to refine the items of the HIWET tool as part of its developmental process. The second part of the study was an original study to investigate the reliability, validity, and utility of the HIWET, which had been developed in the first part of the study. Utility pertains to issues of functionality, and what the tool can do [33]. A way to measure utility is to assess the time it takes to use a particular tool to complete its intended task. This is relevant because users are less likely to change their working practices to incorporate a new tool, if that tool is too time consuming to adopt into their work effectively.

2.2. Development and small-scale pilot testing of HIWET

HIWET was developed from a literature review on topics related to website evaluation, health information quality, and internet evaluation tools in May 2020 conducted among various databases such as PubMed, ScienceDirect, One search, and Google scholar platform. The literature review was conducted on studies related to website evaluation, health information tools, and quality evaluation of online information. The following search terms (with Boolean operators AND, OR) were used to identify relevant literature; 'quality and health information and consumers and internet' and 'health information and internet or website and evaluating tool and quality'. No data restrictions were placed on the research publication. A comprehensive list of literature was retrieved and reviewed to design the structure and collate the contents for the development of HIWET. In addition, the questions from the existing tools such as DISCERN were reviewed as part of the literature review and development process of the HIWET. Similarities of the commonly reported contents were identified from the studies and duplication were removed, and a final list of items were identified. From the above explained process, a final list of 30 items commonly used in the literature [6] to evaluate the quality of online health information were identified to populate the seven domains of the HIWET tool. Each item which closely matched a specific domain was then grouped under one of the seven domains of the HIWET. Thus, an early instrument of HIWET was developed and then tested in a small-scale pilot study on five randomly selected health information websites related to knee and low back pain. The small-scale pilot test was conducted with the aim of ensuring that the layout, syntax, and wording of the HIWET was appropriate. Websites relating to knee and low back pain were chosen because of the related clinical speciality of the researchers in musculoskeletal health conditions. After adjusting the syntax and wording of the items, the researchers further completed additional small-scale pilot studies on three randomly selected neck pain websites. Websites related to neck pain were chosen for the second sets of pilot studies because of its significant worldwide prevalence, this being 288.7 million cases in 2017 [17], as well as being in line with the clinical speciality of the researchers. The websites were chosen from the first page of a Google search using a random number generator. The additional pilot study on neck pain websites was conducted to further clarify HIWET's scoring criteria, and also to test and refine the methodology to be used for the second part of the study which is detailed in Section 2.3. After the small-scale pilot studies, the original 30 items were reduced to 25 after removing repeated items. None of the data collected from these small-scale pilot studies were used in the original study of the HIWET.

2.2.1. Scoring of the HIWET tool

HIWET's seven scoring domains were adapted from a recent paper documenting the seven quality domains [6], and included: Accuracy, Completeness and Comprehensiveness, Technical Elements, Design and Aesthetics, Usability, Accessibility, and Readability. A Likert scale scoring system was adapted to give three possible answers to each question: "yes" scores 2 points; "partially" scores 1 point; "no" scores 0 points. Hints were developed for each item in HIWET to aid the users understanding of how to apply each criterion to the website. For example, to evaluate the domain 'Readability', a hint which reads "Look for clear headings, sub-headings, titles, font, colour, paragraphs, and reference style throughout" may help users identify what to look for. Further examples and clear descriptions of each hint for every domain has been presented in the HIWET tool, which has been shared to the readers as a Supplementary material. A percentage score was produced that provides a quantitative measure of information quality of the websites. Based on the percentage score [18], a label of "poor", "moderate", or "good" was assigned to the percentile score which then provides an interpretation of the overall quality status of the evaluated website.

2.3. Original study to assess reliability, validity, and utility of the HIWET

2.3.1. Website search extraction

Following the protocol of a previous study [8,19], Google.com was

used to search for neck pain websites. Google has the highest market share of existing internet search engines, and websites were selected from the first three pages of the Google search to replicate the search habits of internet users [20,21]. The Completeness and Comprehensiveness domain of HIWET is sub-divided into treatment and management, signs and symptoms, and diagnosis of health conditions. The internet search terms were devised to fit into these sub-divisions, and included:

"treatment of neck pain"; "management of neck pain"; "signs and symptoms of neck pain"; and "diagnosis of neck pain", with each search term allowing for the collection of separate samples of websites. Approximately 10 % of websites from each sample were chosen using a random number generator [20]. The inclusion and exclusion criteria for selection of websites are presented in Appendix A. The study selection criteria for selecting websites were adapted from a previous study [19].

2.3.2. Study instruments

For the validity and utility evaluations, as no tools currently exist that evaluate all domains of quality, three separate frequently used evaluation tools each assessing different domains were selected to be gold-standard comparisons in which to judge the performance of HIWET [7]. The number of gold-standard comparison tools were capped at three as per a previous study [8]. DISCERN was selected to judge the categories of Accuracy, and Completeness and Comprehensiveness, LIDA was selected to measure Usability, and FRES was chosen to evaluate Readability. "WebFX.com" [21] was used to calculate the FRES score.

2.3.3. Study procedure

The study protocol of the original study to evaluate reliability, validity, and utility of the HIWET tool was adapted from previously established methods [8,22,23,24]. To evaluate inter-rater reliability, HIWET was applied independently to each of the 20 websites by two separate researchers (OL and GO). To evaluate intra-rater reliability, HIWET was applied to each of the 20 websites by one researcher (RC) over two different points in time [23]. As per the study protocol [24], webpages were frozen using "Save as PDF" [24]. This facilitated the evaluation of websites over two points in time. Concurrent validity was determined by evaluating the correlation between the scores of HIWET and the scores of comparison gold-standard evaluation tools. To evaluate validity, HIWET, DISCSERN, LIDA, and FRES were applied to the 20 websites by two researchers (LZ and JP). HIWET and FRES both produce percentage scores, whilst LIDA and DISCERN produce numeric values. To ensure parity between all tools when evaluating validity, the scores for LIDA and DISCERN were converted to percentages, thus allowing comparison across tools [8]. The time taken to evaluate each website using each tool was recorded using a stopwatch, and this data was used to evaluate the utility of HIWET [22]. The psychometric testing took place concurrently. The data was recorded onto data-sheets for analysis.

2.3.4. Statistical analysis

The data was analysed using SPSS version 26.0. Before analysis the dataset was checked for missing data, and outliers were checked for using the Tukey method [25], but no outliers existed. Shapiro-Wilkes tests were computed to determine if the data was normally distributed. Inter- and intra-rater reliability was assessed using the Intraclass Correlation Coefficient (ICC). Reliability analysis was conducted for the overall score of HIWET, different domains of HIWET, and lastly, inter-item analysis. Four value ranges were used to interpret the ICCs: ICC values 0.75–1.00 were deemed "excellent"; values between 0.60 and 0.74 were considered "good"; values between 0.40 and 0.59 showed "fair" reliability; and values below 0.40 were deemed to have "poor" reliability [26].

For the validity analysis, correlational tests comparing HIWET to a different gold-standard tool were calculated, with HIWET assessing all seven domains, DISCERN assessing Accuracy, and Completeness and Comprehensiveness, LIDA assessing Usability, and FRES assessing Readability. A Pearson Correlation Coefficient was computed to measure concurrent validity between HIWET, and DISCERN and FRES. A Spearman Correlation Coefficient was computed to determine concurrent validity between HIWET and LIDA. Utility was evaluated by computing an independent-samples t-test comparing the average time taken to evaluate websites using HIWET and the average combined time taken to evaluate the same websites using the gold-standard tools.

3. Results

3.1. Descriptive characteristics

In total, 90 websites were retrieved during the website search, of which 65 met the criteria for inclusion. From those 65 websites, 20 were chosen at random to be included in the analysis, seven of which were included for "treatment and management", seven for "diagnosis", and six for "signs and symptoms". A complete version of HIWET is shown in Appendix B. Table 1 shows the overall website characteristics. Table 2 presents the individual categories and overall scores for each website evaluated using HIWET. A key point being that HIWET rated the overall quality of the websites as moderate to good.

3.2. Reliability analysis

Table 3 shows the results of intra-rater and inter-rater reliability of HIWET. The ICC for intra-rater reliability was excellent, being 0.94 (0.98–0.99) with $p < .001$. The ICC for interrater reliability was fair to excellent, being 0.55 (0.88–0.10) with $p = .04$. Table 4 presents the inter-rater reliability of the inter-item scores for HIWET. The ICC for Accuracy was 0.26 (0.13–0.53) with $p = .06$, and the ICC for Design and Aesthetics was 0.05 (−0.86 to 0.40) with $p = .57$, both of which are lower compared to the overall inter-rater reliability result. Table 5 reports on the intra-rater reliability scores for the various domains of HIWET. A key point being that the ICC was excellent for every domain.

Table 1
Neck pain websites characteristics evaluated by HIWET.

	Number of Websites	HIWET (Mean ± SD)	DISCERN (Mean ± SD)	LIDA (Median (IQR))	FRES (Mean ± SD)
Overall Score For The Websites Related to Treatment and Management of Neck Pain (Mean ± SD)	N = 7	63.96 ± 13.49	62.66 ± 15.55	70.37 (26.85)	60.09 ± 12.12
Overall Score For The Websites Related to Diagnosis of Neck Pain (Mean ± SD)	N = 7	71.36 ± 12.70	75.24 ± 13.08	72.22 (15.75)	44.59 ± 13.19
Overall Score For The Websites Related to Signs and Symptoms of Neck Pain (Mean ± SD)	N = 6	70.96 ± 9.40	59.33 ± 8.12	71.30 (6.94)	58.22 ± 10.32
Overall Score For The Neck Pain Websites (Mean ± SD)		68.65 ± 12.01	66.07 ± 14.05	71.30 (12.03)	54.10 ± 13.46

Note. HIWET = Health Information Website Evaluation Tool, LIDA = MinervaLIDAtion, FRES = Flesch-Kinkaid Readability Ease Score.

Table 2
Individual domains and overall scores for all the websites related to neck pain as evaluated by The HIWET tool.

Name of Websites	URL of Website	A	C and C	TE	D and A	U	A	R	Overall HIWET Score	Overall HIWET Grade
Management of Neck Pain Websites										
1. Saga	https://www.saga.co.uk/magazine/health-wellbeing/treatments/managing-neck-pain	0	5	3	5	6	4	6	29	Good
2. University of Maryland Medical System	https://www.umms.org/ummc/health-services/orthopedics/services/spine/patient-guides/neck-pain-overview	0	4	2	2	5	3	3	19	Moderate
3. Chartered Society of Physiotherapy	https://www.csp.org.uk/conditions/managing-pain-home/managing-your-neck-pain	1	7	5	6	6	3	6	34	Good
4. Medicine Net	https://www.medicinenet.com/neck_pain/article.htm	2	6	4	4	6	2	2	26	Moderate
5. Cedars Sinai	https://www.cedars-sinai.org/health-library/diseases-and-conditions/b/back-and-neck-pain	0	5	2	5	6	1	3	22	Moderate
6. Airedale NHS Foundation Trust	http://www.airedale-trust.nhs.uk	0	4	6	3	0	1	6	20	Moderate
7. Spine health	https://www.spine-health.comment-neck-pain	3	8	6	4	6	3	4	34	Good
Diagnosis of Neck Pain Websites										
1. Mayo clinic	https://www.mayoclinic.org/diseases-conditions/neck-pain	2	6	6	3	6	3	5	29	Good
2. Practical pain management	https://www.practicalpainmanagement.com/pain/spine	5	5	6	4	6	1	0	27	Good
3. National Institute for Health and Care Excellence	https://cks.nice.org.uk/topics/neck-pain-non-specific	4	8	6	4	6	2	3	33	Good
4. Spine Universe	https://www.spineuniverse.com/conditions/neck-pain	6	4	6	5	6	2	3	32	Good
5. New York University Langone Health	https://nyulangone.org/conditions/neck-pain-in-adults/diagnosis	0	3	2	2	5	2	3	17	Moderate
6. Uptodate	https://www.uptodate.com/contents/neck-pain-beyond-the-basics	3	2	6	2	6	2	3	24	Good
7. Very well health	https://www.verywellhealth.com/neck-pain	6	3	6	5	5	2	3	30	Good
Signs and Symptoms of Neck Pain Websites										
1. Healthline	https://www.healthline.com/symptom/neck-pain	4	2	6	3	6	3	5	29	Good
2. Everyday health	https://www.everydayhealth.com/neck-pain/	3	3	6	3	6	3	2	26	Good
3. National Health Service	https://www.nhs.uk/conditions/cervical-spondylosis/	1	2	2	3	5	3	6	22	Moderate
4. WEBMD	https://www.webmd.com/pain-management/why-does-my-neck-hurt	3	2	3	4	6	3	6	27	Good
5. Mayfield clinic	https://mayfieldclinic.com/pe-neckpain	2	4	4	5	4	1	2	22	Moderate
6. Patient info	https://patient.info/bones-joints-muscles/neck-pain	5	4	6	3	6	3	3	30	Good

Note. HIWET = Health Information Website Evaluation Tool, A = Accuracy, C and C = Completeness and Comprehensiveness, TE = Technical Elements, D and A = Design and Aesthetics, U = Usability, A = Accessibility, R = Readability.

Table 3
Summary of intra- and inter-rater reliability scores of the HIWET tool for the websites.

Neck Pain Websites	Intra-Rater Reliability ICC (95 % CI Lower Limit-Upper Limit)	*p*	Rater 1 Score %	Rater 2 Score %	Inter-Rater Reliability ICC (95 % CI Lower Limit-Upper Limit)	*p*
Treatment Websites	0.995 (0.972 – 0.999)	0.001	63.8	56.26	-0.262 (−2.62 to 0.720)	0.649
Diagnosis Websites	0.994 (0.967 – 0.999)	0.001	64.95	68.82	0.447 (−3.286 to 0.909)	0.261
Signs and Symptoms Websites	0.991 (0.948 – 0.999)	0.001	80.9	74.05	-0.318 (−1.869 to 0.734)	0.681
Overall Neck Pain Websites	0.994 (0.985 – 0.998)	0.001	69.33	65.99	0.554 (−0.102 to 0.882)	0.042

Note. HIWET = Health Information Website Evaluation Tool.

3.3. Validity analysis

The results of the Pearson Correlation Coefficient analysis are detailed in Table 6, which demonstrates a positive correlation between the scores of HIWET and DISCERN, r = 0.656, n = 20, p = .002. There was a negative correlation between the scores of HIWET and FRES, r = 0.261, n = 20, p = .267. The quality scoring of HIWET correlated with that of DISCERN, but did not correlate with that of FRES. The results of the Spearman Correlation Coefficient are detailed in Table 6, which demonstrates a positive correlation between the scores HIWET and LIDA, r_s = 0.564, p = .010, N = 20. The quality scoring of HIWET correlated with that of LIDA.

3.4. Utility analysis

The results of the independent-samples *t*-test are detailed in Table 7, which demonstrates a significant difference between the timings for HIWET (M=8.33, SD=1.80) and the gold-standard comparison tools (M=15.42, SD=3.03); t(38) = −9.02, p = .001. These results suggest that it takes significantly less time to evaluate websites using HIWET than compared to using the three comparison tools combined.

4. Disucssion and conclusion

4.1. Discussion

The current study designed HIWET, a new tool to evaluate the quality of online health information, and tested the psychometric properties of HIWET. The findings suggested that HIWET reported high ratings for reliability and concurrent validity when applied to samples of neck pain websites. The findings on the utility also demonstrated HIWET to be more time-efficient in evaluating the quality of online health information when compared to the comparison tools DISCERN, LIDA, and FRES combined.

Table 4
Inter-rater reliability scores of the HIWET tool for the neck pain websites.

Domains	Inter-Item Analysis	Inter-Rater Reliability ICC (95 % CI $_{\text{Lower Limit- Upper Limit}}$)	*p*
Accuracy	Citations clearly provided	-0.118 (−1.502 to 0.534)	0.603
	References clearly displayed	0.142 (−0.534 to 0.594)	0.323
	Procedure for content section clearly acknowledged	0.5 (−0.166 to 0.798)	0.017
	Overall section for Accuracy	**0.263 (−0.128 to 0.534)**	**0.063**
Completeness and Comprehensiveness	Condition fully explained	0.906 (0.434–0.986)	0.011
	Symptoms fully defined	0.390 (−0.201 to 0.869)	0.071
	Causes fully explained	-0.500 (−16.343 to 0.762)	0.672
	Process of diagnosis fully explained	0.211 (−8.128 to 0.875)	0.402
	Management and treatment fully explained	0.864 (0.109–0.977)	0.020
	Benefits and harms of treatment clearly explained	0.222 (−1.113 to 0.814)	0.342
	Information on prevention clearly stated	0.800 (−0.387 to 0.967)	0.047
	Details to avoid condition is stated	0.909 (0.534–0.984)	0.005
	Overall Section For Completeness and Comprehensiveness	**0.779 (0.621–0.871)**	**0.001**
Technical Elements	Authorship of contents clearly displayed	0.481 (−0.149 to 0.783)	0.036
	Ownership of the website disclosed	0.474 (−0.282 to 0.785)	0.080
	Last update of the contents provided	0.747 (0.899–0.378	0.001
	Overall Section For Technical Elements	**0.588 (0.311–0.754)**	**0.001**
Design and Aesthetics	Visual appearance of website	0.507 (−0.149 to 0.798)	0.053
	Navigation menu design	0.024 (−1.531 to 0.624)	0.480
	Graphics and images relate to information	-0.417 (−1.120 to 0.526)	0.864
	Overall Section For Design and Aesthetics	**-0.051 (−0.863 to 0.409)**	**0.568**
Usability	Is the website easy to navigate	0.651 (0.095–0.864)	0.016
	Is there a functioning search engine	-0.216 (−2.388 to 0.543)	0.653
	Is the partnering link or linking websites working	0.472 (−0.316 to 0.790)	0.086
	Overall Section For Usability	**0.402 (0.009–0.645)**	**0.027**
Accessibility	Ease of finding information	0.428 (−0.428 to 0.773)	0.116
	Suitability for the hearing/ visual impaired	0.345 (−0.247 to 0.709)	0.039
	Overall Section For Accessibility	**0.652 (0.350–0.815)**	**0.001**
Readability	Information given clear and easy to understand	0.617 (0.011–0.851)	0.005
	Writing style of information simple	-0.802 (0.457–0.924)	0.001
	Readability of the organisation of information	-0.462 (−3.068 to 0.442)	0.781
	Overall Section For Readability	**0.478 (0.143–0.684)**	**0.005**

Note. HIWET = Health Information Website Evaluation Tool.

Table 5
Intra-rater reliability scores of the HIWET Tool for the neck pain websites.

Domains	Intra-Rater Reliability ICC (95 % CI $_{\text{Lower Limit- Upper Limit}}$)	*p*
Accuracy	1.000 (1.000–1.000)	0.001
Completeness and Comprehensiveness	0.925 (0.824–0.969)	0.001
Technical Elements	1.000 (1.000–1.000)	0.001
Design and Aesthetics	0.970 (0.928–0.988)	0.001
Usability	1.000 (1.000–1.000)	0.001
Accessibility	1.000 (1.000–1.000)	0.001
Readability	0.936 (0.847–0.974)	0.001

Note. HIWET = Health Information Website Evaluation Tool.

Table 6
Validity of the HIWET Tool Against Other Tools.

Tools	Pearson Correlation (95 % CI)	*p*	*n*
HIWET vs DISCERN	0.656 (.408-.854)	0.002	20
HIWET vs LIDA	0.564 (.130-.879)*	0.010	20
HIWET vs FRES	-0.261 (−.582 to .026	0.267	20

Note. HIWET = Health Information Website Evaluation Tool, LIDA = MinervaLIDAtion, FRES = Flesch-Kinkaid Readability Ease Score. *Spearman Correlation.

This study demonstrated overall findings for inter-rater reliability to be fair across all seven of HIWETs scoring domains. However, when inter-item scores are considered, Accuracy, and Design and Aesthetics scores were lower overall than those of the other domains. A past study which evaluated DISCERN [9] reported that lower scores were generally associated with criteria that were more subjective in nature. It might be the case for lower scores in the Accuracy domain of HIWET, as some of the items could be subjective in nature, so this needs to be investigated in future studies which might then improve the reliability scores further. It could be argued that the Design and Aesthetics category, with questions such as "Does the visual appearance of the website look well designed?", is too subjective and open to interpretation, and therefore prone to large variations in results. A website that looks well designed to one person may not look so well designed to another. During the design and development stages of HIWET, descriptive hints to score each item were added in order to reduce subjectivity of the items. This might have contributed to an overall good reliability of HIWET when evaluating online health information. On the contrary, the hints provided for the items in the domains of Accuracy, and Design and Aesthetics need to be studied further in terms of making them less subjective. Perhaps, engaging with end users who refer to websites, such as patients and the general public, who can review the hints along with facilitating the hints rewording and improvement should be carried forward as a process of further development of the HIWET tool in future, which may improve the overall reliability of HIWET as well as engage data and representation of naive users which may also avoid any inherent bias.

There are few studies that have evaluated the intra-rater reliability of website evaluation tools. DISCERN however, was recently evaluated for intra-rater reliability [27]. Two raters evaluated 20 health information leaflets with DISCERN over two different points in time. The intra-rater reliability of DISCERN was demonstrated to be excellent (ICC=81 %). However, the authors did not make clear how much time was left between the two evaluations, which had made the study hard to compare to the present study evaluating HIWET. Although leaving a month between evaluations is a strength of the present study, future research into HIWET could look to evaluate intra-rater reliability over a greater period than one month to determine if this significantly effects HIWETs performance.

Regarding concurrent validity, the results demonstrated that overall quality scores positively correlated between HIWET and DISCERN, and HIWET and LIDA. DISCERN was designed to evaluate content quality [9], and LIDA was developed to assess website design (LIDA). A strength

Table 7
Time Required to Evaluate the Websites Using HIWET Tools Against Other Tools.

Tools	Total Samples	Time In Minutes Mean (SD)	Mean Difference	Standard Error Mean	t	*df*	*p*
HIWET	20	8.32 (1.80)	-7.100	0.40198	-9.02	38	0.01
Gold-Standard Tool (Combination of DISCERN, LIDA and FRES)	20	15.42 (3.03)	7.100	0.67645	-9.02	38	0.01

Note. HIWET = Health Information Website Evaluation Tool, LIDA = MinervaLIDAtion, FRES = Flesch-Kinkaid Readability Ease Score.

of HIWET is that it can evaluate both website content and website design. Website content and website design are equally important because high quality website content helps to gain readers trust, and good quality website design makes it easier to distribute content if that content is provided in an interactive way [6]. The readability of online information plays a key part in determining a website's accessibility [28]. Health information that is hard to read is reported to remain inaccessible to people with low health literacy [28]. Given the importance of health literacy and its relationship to health outcomes [29], it is important that HIWET evaluates the readability of health-related information on websites. The results demonstrated a negative correlation between HIWET and FRES. A key difference between the two tools is that HIWET asks the reader to not only judge if the information is written in an easy to understand way, but also if, for example, the colour of the font makes the writing easier to see against the background webpage. HIWET therefore assesses an extra level of readability which FRES does not. FRES involves an automated formula that provides a percentage rating pertaining to word and sentence length, but not if the font colour is suitable to be easily read. This could partially explain the negative correlation between HIWET and FRES. With such importance attached to health literacy, and the key role this plays in consumer understanding, future research should be aimed at evaluating HIWET against other readability tools to further strengthen the tools validity.

Utility was investigated by the time taken to complete HIWET against the gold standard tools combined, which is a similar protocol used in a previous study which investigated the development of the diabetes quality of internet information tool (DQIIT) [22]. Regarding utility, it took significantly less time to evaluate websites using HIWET (8.33 min) than compared to using the three gold-standard comparison tools combined (15.42 min) approximately. The average time reported to use DQIIT was 30.26 min, which is almost twice as long as the combined gold-standard tools in the present study (15.42 min). HIWET's strength is that the mean time taken to use it to evaluate websites was 8.33 min. With most evaluation tools being too time consuming to complete [7], HIWET demonstrates that it is possible to evaluate a broad range of information quality criteria in a concise and timely fashion.

In the current study, HIWET was tested around only one heath condition, which was neck pain. The decision to test the HIWET on neck pain websites was made in line with the authors' clinical speciality, and further prompted by their patient interactions about online information in their clinical area of practice. It is beyond the scope of this study to test HIWET in websites related to all other health conditions, especially with the current study focus on development and preliminary testing of the HIWET tool. However, future research should consider testing the HIWET tool on a greater number of websites related to health conditions, including diabetes, prostate and breast cancer, and more recent diseases such as Covid-19. Thus, the authors had shared the HIWET tool as a Supplementary material so that anyone interested in exploring the websites related to other health conditions are encouraged to apply testing of HIWET to evaluate other commonly used websites related to diabetes, prostate cancer and more recent diseases such as Covid-19 etc. One could argue that a patient may partly base their comprehension of the quality of online health information on the reputation of the person or company that created the website [30]. Therefore, in the current study, in an effort to reduce the risk of confounding variables such as website logos inducing bias, it was considered beneficial to blind researchers from website names to reduce the subjective impact of bias, thus producing results that were less influenced by the individual researchers past experiences or opinions. Conversely, users of HIWET in practice will not have website names blinded, and so there is a degree of realism in the non-blinding of website names in the research process. Future studies of HIWET might therefore focus on evaluating the websites without any need of blinding of the researchers as it might reflect testing the tool in real life setting.

There are some study limitations. Potential researcher bias needs to be acknowledged. The researchers developed the tool and were oriented with HIWET, and so might be familiar with using the tool. This study looked to evaluate utility, and the time taken to complete website evaluations using HIWET, and therefore familiarisation of the tool might have influenced the results related to utility. Perhaps, new users who are not familiar with HIWET might take a longer time using HIWET to complete the evaluations for these reasons. Therefore, future studies to evaluate the utility and reliability of HIWET by lay people involving end users and patient support groups are under planning, and it will need to be conducted on websites of other common health conditions such as diabetes and cancer. Website evaluation tools have been created for the benefit of clinicians, who are finding themselves under increasing workload pressures [31]. The average GP appointment time in the UK is 11.7 min [32], and the average time taken to evaluate a website using HIWET is 8.33 min, which in practice would leave GPs with precious little time during consultations to both be able to use HIWET and to consult patients. Therefore, a potential area of development for future studies could be to automate HIWET in smart phones and in web platforms to increase accessibility, not just for clinicians, but to all those who might benefit from it. An automated version of HIWET with an analytical equation to enhance faster performance could exist on a website that allows users to determine the quality of online health information with ease in future developmental stages of HIWET. Thus, the current study was necessary to investigate the preliminary functionality of the HIWET tool prior to its automation. Finally, because the main language of the researchers was English, HIWET was designed in the English language. Considering the multinational nature of the internet and of the people who use the it, there may be benefit in evaluating future versions of HIWET that are ported to different languages, which will further strengthen the aim of HIWET to support all stakeholders in healthcare evaluate the quality of online health information. Further developmental studies of HIWET incorporating public-patient engagement, factorial analysis of the items, and automation of the test output are warranted in future.

4.2. Conclusion

With the variable nature of online health information quality, and with clinicians directing patients to online resources during consultations, HIWET was designed to help all stakeholders in healthcare, including clinicians, patients, and researchers, evaluate a wide spectrum of quality in relation to online health information. The overall preliminary findings demonstrate HIWET's reliability, validity, and utility when evaluating online health information. HIWET builds on the existing strengths and foundations of current research. It incorporates existing frameworks to address the gaps within the available website evaluation tools by evaluating a wide spectrum of quality criteria in a timely fashion, and with similar efficacy, when compared to similar tools used for comparison, which were DISCERN, LIDA, and FRES.

4.3. Practice implications

HIWET provides a broad spectrum of domains to evaluate the quality of online heath information. The tool also provides additional criteria and domains when compared to what is currently available to evaluate the quality of online health information. There is potential for HIWT to be used as an educational tool that can be implemented into practice, education, and research relatively quickly. HIWET may help health information consumers and all Stakeholders in healthcare to be better educated about quality websites and online health information, and help them to understand websites with incorrect health information. Clinicians from all fields could use HIWET to screen for suitable health information websites to share with their patients, and health information website developers could use it to help them ensure that their websites are of good quality.

4.4. Policy and ethics

Not applicable. This study involved accessing webpages in the public domain.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

CRediT authorship contribution statement

Conceptualization – All authors contributed Data curation - All authors contributed Formal analysis - All authors contributed Funding acquisition – Not applicable as this study does not involve any funding. Investigation - All authors contributed Methodology - All authors contributed Project administration – Leonard Joseph and Tiffany Blackburn. Supervision and Validation – Leonard Joseph. Writing - original draft; - Luke Zubiena and Leonard Joseph Writing - review & editing - All authors contributed.

Declaration of Competing Interest

The authors have no conflict of interest to declare.

Acknowledgments

The author would like to express appreciation to everyone who had contributed to the study.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.pec.2023.107762.

References

[1] Ferreira G, Traeger A, Machado G, O'Keeffe M, Maher C. Credibility, accuracy, and comprehensiveness of internet-based information about low back pain: a systematic review. J Med Internet Res 2019;21(5):e13357. Available at: <https://www.jmir.org/2019/5/e13357/> [Accessed 8 May 2020].
[2] Ho D, Sullivan S, Schneiders A. The World Wide Web and clinical practice: do sports and musculoskeletal physiotherapists recommend Internet resources to their patients? Physiotherapy 2015;101. pp.e575-6. Available at: <https://www-sciencedirect-com.ezproxy.brighton.ac.uk/science/article/pii/S0031940615034306?via%3Dihub> [Accessed 11 May 2020].
[3] El Sherif R, Pluye P, Thoër C, Rodriguez C. Reducing negative outcomes of online consumer health information: qualitative interpretive study with clinicians, librarians, and consumers. J Med Internet Res 2018;20(5):e169. Available at: <https://www.jmir.org/2018/5/e169/> [Accessed 9 May 2020].
[4] Dobbins M, Watson S, Read K, Graham K, Yousefi Nooraie R, Levinson A. A Tool that assesses the evidence, transparency, and usability of online health information: development and reliability assessment. JMIR Aging 2018;1(1):e3. Available at: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6715399/> [Accessed 14 April 2021].
[5] Tan S, Goonawardene N. Internet health information seeking and the patient-physician relationship: a systematic review. J Med Internet Res 2017;19(1):e9. Available at: <https://www.jmir.org/2017/1/e9/> [Accessed 9 May 2020].
[6] Abdel-Wahab N, Rai D, Siddhanamatha H, Dodeja A, Suarez-Almazor M, Lopez-Olivo M. A comprehensive scoping review to identify standards for the development of health information resources on the internet. PLoS One 2019;14(6):e0218342. Available at: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0218342> [Accessed 11 May 2020].
[7] Zhang Y, Sun Y, Xie B. Quality of health information for consumers on the web: a systematic review of indicators, criteria, tools, and evaluation results. J Assoc Inf Sci Technol 2015;66(10):2071–84. Available at: <https://asistdl.onlinelibrary.wiley.com/doi/abs/10.1002/asi.23311> [Accessed 9 May 2020].
[8] Robillard J, Jun J, Lai J, Feng T. The QUEST for quality online health information: validation of a short quantitative tool. BMC Med. Inform Decis Mak 2018 [online] 18(1). Available at: <https://bmcmedinformdecismak.biomedcentral.com/articles/10.1186/s12911-018-0668-9> [Accessed 11 May 2020].
[9] Charnock D, Shepperd S, Needham G, Gann R. DISCERN: an instrument for judging the quality of written consumer health information on treatment choices. J Epidemiol Community Health 1999;53(2):105–11. Available at: <https://jech-bmj-com.ezproxy.brighton.ac.uk/content/53/2/105.long> [Accessed 11 May 2020].
[10] Minervation, 2008. Is the Lida website assessment tool valid?. [online] Available at: <https://www.minervation.com/does-lida-work/> [Accessed 14 April 2021].
[11] Borgmann H, Wölm J, Vallo S, Mager R, Huber J, Breyer J, Salem J, Loeb S, Haferkamp A, Tsaur I. Prostate cancer on the web—expedient tool for patients' decision-making? J Cancer Educ 2015;32(1):135–40. Available at: <https://search-proquest-com.ezproxy.brighton.ac.uk/docview/1865259163/fulltextPDF/9A6A8ACE1D014823PQ/1?accountid=9727> [Accessed 14 April 2021].
[12] Flesch R. A new readability yardstick. J Appl Psychol 1948;32(3):221–33.
[13] Leonard Grabeel K, Russomanno J, Oelschlegel S, Tester E, Heidel R. Computerized versus hand-scored health literacy tools: a comparison of Simple Measure of Gobbledygook (SMOG) and Flesch-Kincaid in printed patient education materials. J Med Libr Assoc 2018;106(1). Available at: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5764592/?tool=pmcentrez&report=abstract> [Accessed 14 April 2021].
[14] Diviani N, van den Putte B, Giani S, van Weert J. Low health literacy and evaluation of online health information: a systematic review of the literature. J Med Internet Res 2015;17(5):e112.
[15] Keselman A, Smith Arnott, Murcko, A C, Kaufman D. Evaluating the quality of health information in a changing digital ecosystem. J Med Internet Res 2019;21(2):e11129.
[16] Fitzsimmons P, Michael B, Hulley J, Scott G. A readability assessment of online Parkinson's disease information. J R Coll Physicians Edinb 2010;40(4):292–6.
[17] Safiri S, Kolahi A, Hoy D, Buchbinder R, Mansournia M, Bettampadi D, Ashrafi-Asgarabad A, Almasi-Hashiani A, Smith E, Sepidarkish M, Cross M, Qorbani M, Moradi-Lakeh M, Woolf A, March L, Collins G, Ferreira M. Global, regional, and national burden of neck pain in the general population, 1990-2017: systematic analysis of the Global Burden of Disease Study 2017. BMJ 2020:m791. Available at: <https://www.bmj.com/content/368/bmj.m791#:~:text=Globally%2C%20the%20number%20of%20prevalent,%E2%88%920.9%25%2C%2095%25%20uncertainty> [Accessed 14 April 2021].
[18] Fairbank J, Pynsent P. The oswestry disability index. Spine 2000;25(22):2940–53. Available at: <https://oce-ovid-com.ezproxy.brighton.ac.uk/article/00007632-200011150-00017/HTML> [Accessed 14 April 2021].
[19] Hirsch M, Aggarwal S, Barker C, Davis C, Duffy J. Googling endometriosis: a systematic review of information available on the Internet. Am J Obstet Gynecol 2017;216(5):451–8. Available at: <https://www-sciencedirect-com.ezproxy.brighton.ac.uk/science/article/pii/S0002937816319871?via%3Dihub> [Accessed 14 April 2021].
[20] Haahr, M., 2021. RANDOM.ORG - True Random Number Service. [online] Random.org. Available at: <https://www.random.org/> [Accessed 3 June 2021].
[21] WebFX.com, 2021. Readability Test Tool. Readability Test Results for painscience.com/articles/when-to-worry-about-neck-pain-and-when-not-to.php. Available at: https://www.webfx.com/tools/read-able/check.php?tab=Test%2BBy%2BUrl&uri=https%3A%2F%2Fwww.painscience.com%2Farticles%2Fwhen-to-worry-about-neck-pain-and-when-not-to.php [Accessed January 25, 2021].
[22] Seidman J, Steinwachs D, Rubin H. Conceptual framework for a new tool for evaluating the quality of diabetes consumer-information web sites. J Med Internet Res 2003;5(4):e29. Available at: <https://www.jmir.org/2003/4/e29/> [Accessed 12 May 2020].
[23] Rees C. Rapid research methods for nurses. Midwives, and health professionals. Chichester: Wiley Blackwell; 2016. p. 97.
[24] Pdfcrowd.com. 2021. Convert HTML to PDF online, Web/HTML to PDF API. [online] Available at: <https://pdfcrowd.com/> [Accessed 14 April 2021].
[25] Zijlstra W, van der Ark L, Sijtsma K. Outlier detection in test and questionnaire data. Multivar Behav Res 2007;42(3):531–55. Available at: <https://www-tandfonline-com.ezproxy.brighton.ac.uk/doi/full/10.1080/00273170701384340> [Accessed 24 February 2021].
[26] Hernaez R. Reliability and agreement studies: a guide for clinical investigators. Gut 2015;64(7):1018–27. Available at: <https://gut-bmj-com.ezproxy.brighton.ac.uk/content/64/7/1018> [Accessed 14 April 2021].
[27] McCool M, Wahl J, Schlecht I, Apfelbacher C. Evaluating written patient information for eczema in German: comparing the reliability of two instruments,

DISCERN and EQIP. PLOS ONE 2015;10(10):e0139895. Available at: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0139895>.

[28] McInnes N, Haglund BJ. Readability of online health information: implications for health literacy. Inform Health Soc Care 2011;36(4):173–89. https://doi.org/10.3109/17538157.2010.542529.

[29] Oliffe M, Thompson E, Johnston J, Freeman D, Bagga H, Wong P. Assessing the readability and patient comprehension of rheumatology medicine information sheets: a cross-sectional Health Literacy Study. BMJ Open 2019;9(2):e024582. https://doi.org/10.1136/bmjopen-2018-024582.

[30] Kim J, Lennon S. Effects of reputation and website quality on online consumers' emotion, perceived risk and purchase intention. J Res Interact Mark 2013;7(1):33–56. Available at: <https://www.emerald.com/insight/content/doi/10.1108/17505931311316734/full/html> [Accessed 25 October 2021].

[31] Beccy, B., 2021. Understanding pressures in general practice. [online] KingsFund.org.uk. Available at: <https://www.kingsfund.org.uk/publications/pressures-in-general-practice> [Accessed 25 October 2021].

[32] Irving G, Neves A, Dambha-Miller H, Oishi A, Tagashira H, Verho A, Holden J. International variations in primary care physician consultation time: a systematic review of 67 countries. BMJ Open 2017;7(10):e017902. Available at: <https://bmjopen.bmj.com/content/7/10/e017902> [Accessed 25 October 2021].

[33] Tarkkanen K, Harkke V, Reijonen P. Are we testing utility? Analysis of usability problem types. Des Use Exp Usability: Des Discourse 2015:269–80.