

ORIGINAL RESEARCH

Evaluation of the quality of phantom limb pain information on YouTube

Padley TJ,¹ Tanqueray E,² Malhotra R³

1. Foundation Year 1 Doctor, London Deanery, UK
2. Anaesthetics ST5 Trainee, Mersey Deanery, UK
3. Consultant in Anaesthesia and Pain Management, Liverpool University Hospitals NHS Foundation Trust, UK

Corresponding author:

Thomas Joseph Padley
St George's Hospital, Blackshaw Road, London SW17 0QT, UK
Email: tom.padley@doctors.org.uk

Received: 27th June 2024

Accepted: 5th August 2024

Online: 29th August 2024

Plain English Summary

Why we undertook the work: Pain after the amputation of a limb is extremely common (95% of people). Of the different types of pain that patients can experience after amputation, 80% of patients experience phantom limb pain, which can have a significant effect on quality of life. Treatment options for phantom limb pain are expanding. Patients often access information about their health online, particularly through videos on YouTube. To our knowledge there has not been any previous study looking at the quality of information available to the public about phantom limb pain, and the authors wanted to find out more about how accurate and useful information available on YouTube is for patients suffering from this condition.

What we did: We identified the first 50 videos from a search on YouTube using the phrase 'pain after amputation', with 35 videos included for assessment. Two of the authors analysed these videos using five different measurement tools looking at various aspects of the quality, reliability and how understandable and actionable these videos are for patients.

What we found: Overall, the content of YouTube videos was of low to moderate quality, which is lower than studies looking at other pain-related conditions. There was no link between how high a particular video comes up on a search and its quality. There was also no strong link between the source of the information and quality; we found that even videos published by more reputable sources were not of higher quality.

What this means: The findings of this study highlight the challenges that patients suffering from pain after amputation may come up against when using online information to find out more about their condition. It may be difficult to find accurate, useful information about explanations and treatment options, and there is a potential to access incorrect information and advice. This is important for both patients to understand and also the healthcare professionals caring for them, and shows there is a need for easily accessible high quality health-related information. The quality assessment tools currently available are readily accessible and easy to use, so there is a potential for healthcare professionals to independently assess online resources.

Abstract

Background: Patients more frequently seek health-related information from online sources, including YouTube. Phantom limb pain (PLP) is a complex problem, with evolving research into its pathophysiology and management. As healthcare professionals, it is important to be aware of the quality of publicly accessible information. This study aimed to investigate the reliability and quality of YouTube videos about PLP.

Methods: 50 videos were identified from YouTube using the search term 'pain after amputation'. Sources and video parameters were documented. Two assessors examined the videos independently using five scoring systems including the *Journal of American Medical Association (JAMA)* Benchmark criteria, the Global Quality Score (GQS), a Subjective score, and the Patient Education Materials Assessment Tool for Audio-Visual Materials (PEMAT-A/V). At the time of video identification, the assessors included one final year medical student with an interest in anaesthetics and one anaesthetic speciality trainee doctor.

Results: Our study indicates that the overall quality and reliability of YouTube videos covering PLP is poor, with mean *JAMA*, GQS, Subjective, PEMAT-A/V Understandability, and PEMAT-A/V Actionability scores of 2.07 (minimum and maximum scores of 0 and 4), 2.66 (minimum and maximum scores of 1 and 5), 3.76 (minimum and maximum scores of 0 and 10), 16.67% (minimum and maximum scores of 0% and 100%) and 40.88% (minimum and maximum scores of 0% and 100%) respectively, demonstrating the inadequacy of currently available

online information. The percentage range of videos that were deemed high quality was 9–26%. We also found that videos that patients access more readily (calculated using an interaction index) are not necessarily of a higher quality, and that the publisher (ie, the professional, patient, independent academic or company who uploaded the video) of the content has no significant effect on the quality of the video ($p=0.704$, $p=0.580$, $p=0.086$, $p=0.432$, $p=0.364$).

Conclusions: Online audiovisual PLP-related information is of poor quality. When patients are searching for information online they are more likely to be directed to content that is inadequate and of poor quality. Clinicians should be aware of the quality of information that is available to patients. Higher quality videos are essential to aid patient understanding of PLP.

Key words: YouTube, phantom limb pain, health information, information quality, patient education

Introduction

Pain after amputation is an almost universal symptom in amputee patients with 95% reporting amputation-related pain.¹ Of these, phantom limb pain (PLP) is the most prevalent at 80%.¹ Increasing numbers of patients are undergoing amputations; an estimated prevalence rate in the UK is 26.3 per 100,000.² PLP significantly reduces quality of life³ and has a large impact on the workforce and economic society.⁴

Research around amputation has been named as a priority area by The Vascular Priority Setting Partnership⁵ in conjunction with the James Lind Alliance who play a pivotal role in directing the national research agenda based on workshops involving patients, carers and healthcare professionals. This study looks at one of the specific questions within this agenda: “How do we improve the information provided to patients undergoing amputation?”. This highlights the importance of healthcare providers understanding the information that is accessible by the public.

Patients often access health-related information related to their existing conditions or symptoms, with online video-streaming sites a popular method.⁶ Of the video-streaming sites, YouTube is the most commonly accessed in the world.⁷ YouTube is an extremely popular and accessible hub of information with users simply requiring an internet connection and an audiovisual device such as a PC or mobile phone to engage with content on the site.⁸ However, validity of information on the internet cannot be guaranteed⁹ as there are no required standards for medical information that is published online and there are no restrictions for who can publish and upload YouTube videos regardless of qualifications or profession.¹⁰ Information surrounding the mechanisms and treatment options for PLP is a rapidly developing field of medical research,¹¹ indicating that this may be a particularly relevant problem in publicly available information on PLP.

Multiple studies have assessed patient-centred health information on the internet for other conditions^{12–14} and, to the best of our knowledge, this is the first time it has been done for PLP. In a study by Kwan *et al*¹³ the authors investigated the reliability of internet-based information about statin therapy using the Global Quality Score (GQS) and *Journal of American Medical Association* (JAMA) Benchmark criteria and found that on the whole there was

no significant correlation between video characteristics and content quality other than number of days since publication ($p=0.022$). The overall content quality on YouTube about hip osteoarthritis was poor in a number of studies with the range of videos of poor educational quality between 64% and 91%.^{14–16}

The discourse varies about the reliability of video resources regarding chronic pain and chronic pain syndromes. Altun *et al*¹² evaluated video sources on YouTube covering complex regional pain syndrome using the GQS and JAMA scores and found that the majority of content was of intermediate to high quality. Furthermore, higher quality content achieved higher interaction indexes than lower quality videos ($p=0.010$), with patient sources being of a lower quality than information from health professionals ($p<0.001$). Other studies have investigated the online available content surrounding different types of pain such as inflammatory back pain, post-COVID pain and neck pain. The overall quality of videos were poor, with only 19–21% of high quality and 35% of moderate to high quality.^{17,18} Authors have also found statistically significant differences between the source type and content quality, with 57.9–79.2% of high quality videos published by academics, professional organisations and healthcare sources.^{17–19}

This study aims to evaluate the available online information from video sources publicly available on the topic of PLP.

Methods

This descriptive research evaluates information that is publicly accessible therefore does not require ethical committee approval.

Video identification

Videos were identified on YouTube (<https://www.youtube.com>) from a single IP address in Liverpool, UK without signing into a Google account using the search term ‘pain after amputation’. Video identification took place between 12 November 2023 and 30 November 2023. The search results were sorted by relevance. Videos were first screened for duplicates and for those with a duration of less than 1 minute including ‘YouTube Shorts’ content; these videos were not assessed. After this screening process, the first 50 videos were recorded.

Exclusion criteria consisted of: (1) videos not in English; (2)

videos with a lack of relevance for patients (ie, intended for healthcare staff); (3) videos that did not focus on the search topic; (4) videos that were inaccessible (requiring sign-in, age restricted, region blocked); and (5) videos without a focus on PLP.

Video evaluation

Videos were assessed separately by two of the researchers independently. The following video characteristics were recorded: (1) title; (2) duration; (3) number of views; (4) number of likes; (5) number of dislikes; (6) source type; (7) number of days since upload; (8) viewing rate (number of views/number of days since upload \times 100%); and (9) interaction index (number of likes and dislikes/total number of views \times 100%).¹² Source type was defined by the authors and videos were grouped into assigned categories. Viewing rate was intended to estimate the number of views, irrespective of the time that the video has been available on YouTube. Similarly, the interaction index was intended to determine the number of interactions (both positive and negative) that a video has per view, indicating increased levels of viewer engagement. These metrics were used in this study as the authors did not have access to more in-depth video characteristics such as average view time.

All 35 included videos were evaluated by the two researchers separately. Assessment of the quality and the ability of the videos to provide better education to the viewer was evaluated using the following scores/tools: (1) Global Quality Score (GQS); (2) Subjective score; (3) Patient Education Materials Assessment Tool for Audio-Visual Materials (PEMAT A/V) Understandability tool; (4) PEMAT A/V Actionability tool. Assessment of the reliability of the videos was evaluated using the *JAMA* Benchmark criteria.

The GQS grading system devised by Bernard *et al*²⁰ provides a score of 1–5 based on the quality of the videos, with 1 being the lowest quality and 5 being the highest quality: (1) low quality, video information flow weak, most information is missing, not beneficial for patients; (2) low quality, low flow of information, some listed information and many important issues are missing, very limited use for patients; (3) moderate quality, insufficient flow of information, some important information is sufficiently discussed, but some are poorly discussed and somewhat useful for patients; (4) good quality and generally good information flow, most relevant information is listed but some topics are not covered, useful for patients; (5) excellent quality and information flow, very useful for patients.²¹

The Subjective score was developed by the authors and again provides a score of 0–10 based on the quality of the videos, with 1 being the lowest quality and 5 being the highest quality. A score of 0–2 (0: not mentioned, 1: mentioned with little detail, 2: mentioned with good detail) is provided for each of the following points: (1) possible causes of pain after amputation; (2) symptoms of PLP; (3) pharmacological options for PLP; (4) non-pharmacological options for PLP; (5) mention of psychological/MDT support as part of holistic management.

The PEMAT A/V tools are widely accepted methods for

assessing the understandability and actionability of audiovisual materials. The extensive list of the items is available at <https://www.ahrq.gov/health-literacy/patient-education/pemat-av.html>, where each item is given a rating of either 0 ('disagree') or 1 ('agree'), or sometimes not given a ranking ('not applicable') depending on what is being assessed.²²

The *JAMA* Benchmark criteria use four core standards to grade the reliability of each video on a scale of 0–4 based on the following criteria, where each is given a score of 0 or 1: authorship, attribution, disclosure, currency.²³

Statistics

All statistical analyses were conducted using SPSS v.29.0.1.0 (171).

The interobserver correlation coefficient (ICC) was calculated between the two researchers for the *JAMA* criteria, GQS scores, Subjective scores, PEMAT A/V Actionability scores and PEMAT A/V Understandability scores (see Appendix 1 online at www.jvsngbi.com). The Shapiro–Wilks test was used to determine normality for all variables as our sample number was <50 .^{24,25} Either Spearman's or Pearson's correlation coefficients were used to investigate statistical significance between the video characteristics and the quality scores individually. Either one-way ANOVA or Kruskal–Wallis tests were then performed to investigate statistical significance between the source type and the quality scores.

Results

Video assessment

The YouTube search was performed with the term 'pain after amputation', allowing for the identification of the first 50 videos. Videos were included if they were not duplicates and they did not fall into the 'YouTube Shorts' content category or have a duration of less than 1 minute. The process of video selection is shown in the flow diagram in Figure 1.

The 50 videos were then screened for our exclusion criteria, with full details presented in Figure 1. The included videos ($n=35$) were categorised by source type into the following four main source types: independent academic channels, healthcare providers, doctors with independent channels, and patient testimonies. A representation of the different source types is shown in Figure 2.

The characteristics of all included videos are shown in Table 1, with the mean values displayed. The mean viewing rate was 3681.36 and the mean interaction index was 1.80.

The mean and median scores of the included videos are shown in Table 2. The quality scores were low, in particular the subjective score with a median of 3 out of 10. In addition, the PEMAT A/V Actionability score was very low with a mean score of only 16.67%.

Interobserver agreements were calculated and demonstrated that all quality and reliability scores showed excellent interrater reliability between the two researchers. This information is presented in the Appendix. The parameters for quality and reliability were then sub-categorised into low, moderate and high content

Figure 1 Flow diagram of the video selection process.

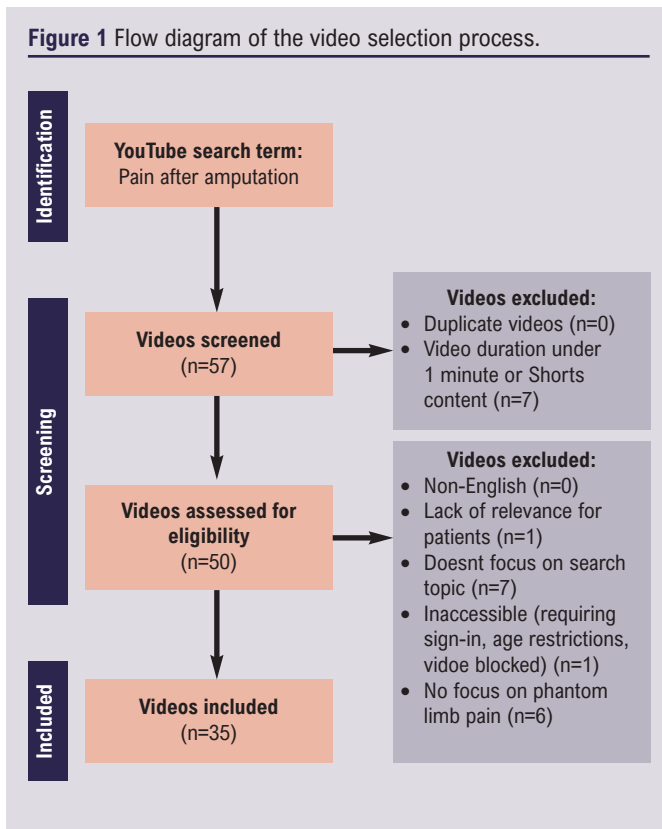


Table 1 Basic video characteristics.

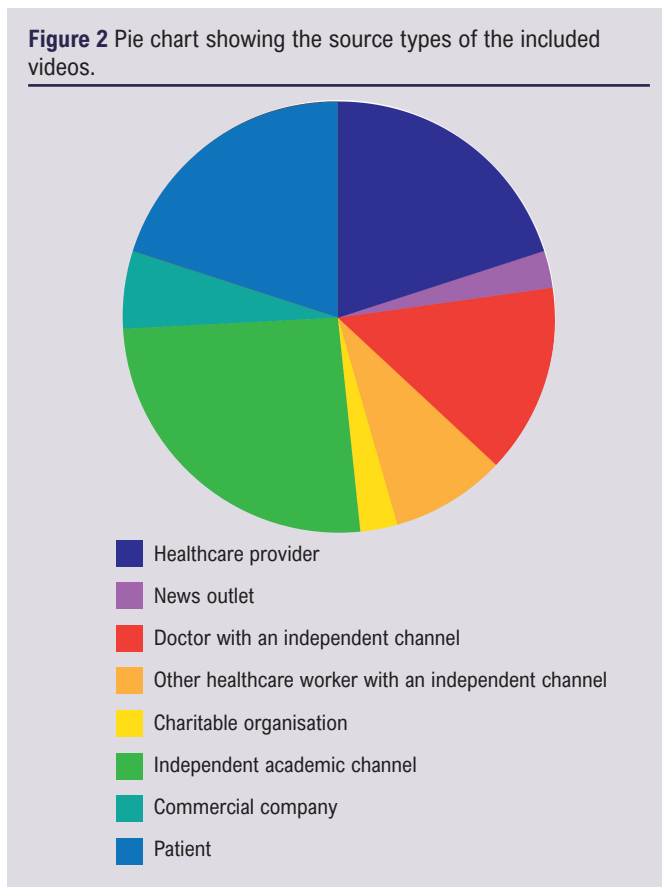
Characteristics	Mean (minimum–maximum)
Video duration (min)	6.32 (1.28–20.40)
Number of views	55057 (242–774,769)
Number of likes	1207 (0–20,792)
Number of dislikes	18 (0–257)
Number of days since publication	1774 (9–4657)
Viewing rate	3681.36 (18–41,454)
Interaction index	1.80 (0.13–6.19)

Table 2 Overall quality and reliability scores of the videos including the minimum-maximum possible scores of each scoring system

	Possible (minimum–maximum)	Mean	Median
GQS score	1–5	2.66	2.5
Subjective score	0–10	3.76	3
PEMAT A/V Understandability score	0–100	40.88	–
PEMAT A/V Actionability score	0–100	16.67	–
JAMA Benchmark criteria	0–4	2.07	2

GQS, Global Quality Score; JAMA, Journal of American Medical Association; PEMAT, Patient Education Materials Assessment Tool for Audio-Visual Materials.

Figure 2 Pie chart showing the source types of the included videos.



quality. The number of videos that fit into each quality and reliability score is shown in Table 3.

The Shapiro–Wilks test for normality was performed for all variables and showed that ‘number of days since publication’, ‘GQS score’ and the ‘PEMAT Understandability score’ were normally distributed, while all other variables were not normally distributed.

We investigated the correlation between the different quality and reliability scores using the Spearman’s correlation coefficient when one or neither variable was normally distributed or the Pearson’s correlation coefficient when both variables were normally distributed. From this analysis we found that many of the quality scores were statistically different from one another including the comparison between the GQS and the JAMA criteria ($p=0.039$), the Subjective score ($p<0.001$) and the PEMAT Understandability score ($p<0.001$). There were further statistically significant differences between the Subjective score and the PEMAT Understandability score ($p=0.021$) and the PEMAT Actionability score ($p=0.011$). Finally, the difference between the PEMAT Understandability and Actionability scores was statistically significant ($p=0.001$). These data are shown in Table 4.

We compared video characteristics data and quality scores

Table 3 Assessment parameter sub-categorisation into low, moderate and high content quality

	Quality scores	Number of videos (n=35)
GQS score (1–5 points)	Low content quality (1–2 points)	15 (43)
	Moderate content quality (2.5–3.5 points)	11 (31)
	High content quality (4–5 points)	9 (26)
Subjective score (0–10 points)	Low content quality (0–3 points)	19 (54)
	Moderate content quality (3.5–6.5 points)	10 (29)
	High content quality (7–10 points)	6 (17)
PEMAT AV Understandability score (%)	Low content quality (0–33.32%)	14 (40)
	Moderate content quality (33.33–66.66%)	18 (51)
	High content quality (66.67–100%)	3 (9)
PEMAT AV Actionability score (%)	Low content quality (0–33.32%)	24 (69)
	Moderate content quality (33.33–66.66%)	7 (20)
	High content quality (66.67–100%)	4 (11)
JAMA Benchmark criteria (0–4 points)	Low content quality (0–1 points)	7 (20)
	Moderate content quality (1.5–2.5 points)	20 (57)
	High content quality (3–4 points)	8 (23)

Values are presented as number (%).

GQS, Global Quality Score; JAMA, *Journal of American Medical Association*; PEMAT AV, Patient Education Materials Assessment Tool for Audio-Visual Materials.

Table 4 Significance between quality scores

	JAMA	GQS	Subjective score	PEMAT Understandability score	PEMAT Actionability score
JAMA	–	–	–	–	–
GQS	0.039* (0.350)	–	–	–	–
Subjective score	0.232 (0.207)	<0.001* (0.845)	–	–	–
PEMAT Understandability score	0.602 (0.091)	<0.001* (0.871)	0.021* (0.390)	–	–
PEMAT Actionability score	0.215 (0.215)	0.068 (0.312)	0.011* (0.424)	0.001* (0.531)	–

Values are presented as p values (Pearson's/Spearman's correlation coefficient).

*Statistically significant difference.

GQS, Global Quality Score; JAMA, *Journal of American Medical Association*; PEMAT AV, Patient Education Materials Assessment Tool for Audio-Visual Materials.

using the Spearman's correlation coefficient when one or neither variable was normally distributed or the Pearson's correlation coefficient when both variables were normally distributed. Statistical significance was found between 'Duration' and the GQS ($p=0.022$) and Subjective ($p=0.044$) quality scores, as well as between 'View number' and the PEMAT Actionability score ($p=0.014$). These data are shown in Table 5.

We then compared the source type to the quality scores using the one-way ANOVA test for the normally distributed GQS and PEMAT Understandability scores and the Kruskal–Wallis test for the remaining non-normally distributed scores. No statistical significance was found between source type and all of the quality scores individually. These data are shown in Table 6.

Discussion

The internet has become a regular source of healthcare-related information due to its highly accessible and inexpensive nature; often it is far quicker for patients to search the internet than to seek advice from their own healthcare professional. It is, however, an unregulated landscape which requires thorough evaluation in order to determine the overall quality and reliability of the information that patients are most likely to access.

The overall quality of the videos that we evaluated in this study was low to moderate, with the mean proportions of the low, moderate and high groups across all scoring systems of 45%, 40% and 17%, respectively. Furthermore, the largest proportion of videos fell into the low quality category when using the GQS,

Table 5 Significance between quality scores and video characteristics

	JAMA	GQS	Subjective score	PEMAT Understandability score	PEMAT Actionability score
Duration	0.086 (0.294)	0.022* (0.392)	0.044* (0.343)	0.222 (-0.212)	0.635 (0.083)
View number	0.889 (-0.024)	0.156 (0.245)	0.112 (0.274)	0.089 (0.292)	0.014* (0.410)
Like number	0.888 (0.025)	0.056 (0.326)	0.188 (0.228)	0.124 (0.265)	0.037 (0.354)
Dislike number	0.945 (-0.012)	0.370 (0.159)	0.317 (0.177)	0.615 (0.089)	0.149 (0.253)
Days since publication	0.174 (-0.232)	0.141 (-0.254)	0.777 (-0.050)	0.208 (0.231)	0.170 (0.237)
Interaction index	0.863 (0.031)	0.061 (0.324)	0.550 (0.106)	0.861 (0.031)	0.836 (-0.037)
Viewing rate	0.331 (0.169)	0.051 (0.338)	0.132 (0.259)	0.466 (0.127)	0.121 (0.267)

Values are presented as p values (correlation coefficient).

*Statistically significant difference.

GQS, Global Quality Score; JAMA, *Journal of American Medical Association*; PEMAT A/V, Patient Education Materials Assessment Tool for Audio-Visual Materials.

Table 6 Significance between quality score and source type (p values)

Quality score	p-value
GQS score	0.704
Subjective score	0.580
PEMAT A/V Understandability score	0.086
PEMAT A/V Actionability score	0.432
JAMA Benchmark criteria	0.364

GQS, Global Quality Score; JAMA, *Journal of American Medical Association*; PEMAT A/V, Patient Education Materials Assessment Tool for Audio-Visual Materials.

Subjective and PEMAT A/V Actionability score (43%, 54%, 69%), while moderate quality content was the largest when using the PEMAT A/V Understandability score and the JAMA Benchmark criteria (51%, 57%). The particularly high prevalence of low quality content when using the PEMAT A/V Actionability score could be due to a number of reasons. First, the videos that we analysed were generally intended to demystify the phenomenon of PLP and not designed to provide actionable avenues for patients to explore. The 3–4 criteria (depending on if there are charts, graphs, tables or diagrams) are also extremely specific and challenging to meet, making it difficult to score highly. This indicates that the audiovisual materials for PLP available on YouTube may not be of an adequate quality to provide patients with accurate and important information about this highly prevalent symptom. This in turn may lead to the spread of misinformation, causing patients to misunderstand and potentially doubt the advice they have been given by specialists.

Altun *et al*¹² investigated the quality of YouTube resources for complex regional pain syndrome and found that the information was of a much higher quality than in our study, and that this content was interacted with much more than videos of poorer quality. This

could indicate that publicly available audiovisual materials about complex regional pain syndrome are of much higher quality than those covering PLP. We propose that alternative explanations of this could be that PLP is a less well understood condition or possibly that the algorithm is less well suited to the terminology used around pain after amputation. However, our study is not the first to indicate that health-related YouTube content is not of a satisfactory quality. Studies that analysed the content covering other pain-related topics found that only 19–21% of videos were of high quality,^{17,18} which is similar to our findings (GQS: 26%, Subjective score: 17%, PEMAT A/V Understandability score: 9%, PEMAT A/V Actionability score: 11%, JAMA Benchmark criteria: 23%) and that 64–91% were of poor quality,^{14–16} which overall was higher than the results of our study (GQS: 43%, Subjective score: 54%, PEMAT A/V Understandability score: 40%, PEMAT A/V Actionability Score: 69%, JAMA Benchmark criteria: 20%). These findings may indicate a general lack of quality in healthcare-related audiovisual materials on YouTube that is not limited to just PLP but also to other pain-related conditions.

Of the first 50 videos (after exclusion of short-form content <1 minute), 14 were deemed irrelevant to the search topic or not pitched at a patient level by the two researchers. This could indicate that, when patients seek health-related information on these sites, they could find it challenging to identify suitable sources of information that are relevant to them. The promotion of irrelevant information may cause patients to gain a false understanding of symptoms and diseases which could have a number of consequences such as following incorrect advice and misunderstanding essential information about their condition, which could be damaging to both physical and psychological health.

During our initial search, seven videos were excluded as 'YouTube Shorts' content as the length of these videos would likely negatively influence their quality scores due to a simple lack of time to convey enough information, and the algorithm is different from the traditional YouTube algorithm meaning it would be unfair to

compare long-form and short-form content together. It is important to note, however, the high number of views that these videos had at the time of the researchers' initial video search, with an average view count of 1,142,857. As short-form content becomes more prevalent with the rise in popularity of features and applications such as 'YouTube Shorts', 'Instagram Reels' and 'TikTok'; more research is also needed to evaluate the quality and reliability, particularly with the higher engagement levels that this type of content achieves.

The *JAMA* Benchmark criteria indicate the reliability of the source depending on how much information is disclosed to the viewer. In our study, 77% of included videos were classified into the 'low' or 'moderate' quality measures based on these criteria, which may indicate that highly recommended videos may have a high risk of publisher bias which in turn damages the credibility of the source. This is made more apparent considering that all videos scored one point for currency, as all videos uploaded on YouTube are required to publish an upload date.

Currently the literature does not present a clear picture as to whether higher quality videos covering health-related topics are more readily recommended to patients and are accessed more frequently. In our study there was no overall correlation between the degree of interaction with the videos and the quality, with the only significant difference being between the view number and the PEMAT A/V Actionability score ($p=0.014$). However, Altun *et al*¹² found that, in videos covering complex regional pain syndrome, higher quality content achieved higher interaction indexes than lower quality videos ($p=0.010$). This could indicate that patients who access YouTube in search of PLP-related educational content will predominantly access low quality and potentially misleading videos, which could be damaging for the patient population. It is unclear why differences are apparent when analysing complex regional pain syndrome versus PLP, and more research needs to be done to analyse whether this is the case for other health conditions.

It is expected that seeking information from sources such as healthcare professionals or academic channels would produce higher quality and more reliable information; however, our study has shown that there is no statistically significant difference between the quality of the information provided and the source type. This does not agree with some of the findings in the literature which have previously found statistically significant differences between the quality and the source type, particularly that videos published by healthcare professionals and academics are of a higher quality than those published by patients. These findings could indicate that more regulated and higher-quality content should be published on YouTube by healthcare professionals or healthcare organisations regarding PLP to allow patients to access more accurate information.

This research has significant implications for patients who suffer with pain after amputation and the clinicians who care for them. This study alters perceptions about how patients access information, the quality of the content that they are accessing, and

gives an indication of how misinformation can spread within patient communities. Future research should focus on more specific reasons as to why patients access the information that they do, and how we can improve the online information landscape for PLP and other conditions. Furthermore, our understanding about online information sharing must improve in order to determine the optimal way to distribute accurate health-related online content and how such videos should be developed.

Strengths

To our knowledge, this is the first study to assess the quality and reliability of YouTube sources covering the topic of PLP, meaning that our study can aid patients to make decisions about where they seek health-related information on the internet. Our study also identified sources using popular search strategies in line with how a lay person would access information, increasing the reliability and accuracy of our results. To reduce subjectivity, two evaluators independently evaluated each source and used non-subjective grading criteria as well as subjective grading criteria. This use of multiple grading criteria also decreased the potential for inaccuracies to arise.

Limitations

The primary limitation of our study is that it is cross-sectional and only represents a snapshot in time of the ever-changing landscape of online content, and there is a possibility that the same study would produce different results if it was repeated over a different time period. During the initial video search the two assessors were not signed in; however, YouTube still uses a personalised algorithm meaning that patients may not be recommended the same first 50 videos as the assessors accessed. For patients who viewed the content, they may have been signed in and so this algorithm may in effect be potentially artificially inflating view numbers (including of 'poor quality' content) and is a confounder when investigating the relationship between video characteristics and quality. It is also important to note that there was no consideration of 'cookies' when the assessors completed their video identification, which may have affected the search results.

There is a debate over what constitutes a high quality source, meaning the use of scoring systems will never represent a completely accurate analysis of quality.²⁶ The subjective scoring criteria relies upon researchers creating domain-specific instruments based on medical guidelines, textbooks, literature and medical expertise to guide the evaluation of the quality of the content, which is something that the validated tools used do not address.²⁶ The lack of correlation between the different scoring systems makes this apparent, and emphasises that no single scoring system can provide a definition of quality. Other studies have previously developed new grading systems for the audiovisual content, which is something that this study does not explore and could be a better determinant of quality than the existing systems.²⁷

KEY MESSAGES

- The public is increasingly using YouTube videos to gain knowledge on health-related issues as well as other online content platforms such as 'Instagram' and 'TikTok'.
- Healthcare providers have a responsibility to understand the information that is accessible by patients.
- There is a need for improved quality scoring systems for health-related audiovisual online information.
- High quality online educational videos are required to effectively guide patients.

Recommendations

The quality and reliability of videos related to PLP on YouTube is insufficient, which might lead to the spread of misinformation within the patient population. The systems for grading these parameters are not adequate, so a generalised method of analysis is needed for future researchers. To address the issues that this paper identifies, we recommend a higher level of regulation by video publishing platforms to limit the spread of poor quality health-related information or, alternatively, to provide better support to professionals involved in patient care to evaluate sources that their patients are reviewing to ensure they are viewing high quality content. Furthermore, to ensure that future research studies can assess audiovisual materials accurately, it is necessary to increase education and training for researchers with online tutorials about health information evaluation. It may also be possible to design tools that automatically detect quality indicators such as the tool created by Griffiths *et al.*²⁸

Conclusion

This study demonstrates that, when patients seek health-related information from YouTube, they are likely to be presented with inadequate and poor quality information. It also shows that YouTube's independent engagement statistics such as likes and views should not be considered indicators of quality. This is highly relevant to clinicians, as they must understand what information patients are likely to come across from their independent research in order to tailor their own communication to patients. It is also important to understand which sources of information patients value the most, including whether they value content from other patients or non-patient sources. Furthermore, patients will continue to access these sources due to their easily accessible nature, so high quality educational videos are needed to effectively guide patients on the complex condition that is PLP.

Conflict of Interest: The authors declare that there is no conflict of interest.

Funding: None.

Reviewer acknowledgement: *JVSGBI* thanks Louise Tisdale, Clinical Specialist Physiotherapist Amputee Rehab, Royal Wolverhampton NHS Trust and Brenig Gwilym, South East Wales Vascular Network, for their contribution to the peer review of this work.

References

1. Ephraim PL, Wegener ST, MacKenzie EJ, Dillingham TR, Pezzin LE. Phantom pain, residual limb pain, and back pain in amputees: results of a national survey. *Arch Phys Med Rehabil* 2005;**86**(10):1910–9. <https://doi.org/10.1016/j.apmr.2005.03.031>
2. Ahmad N, Thomas GN, Gill P, Chan C, Torella F. Lower limb amputation in England: prevalence, regional variation and relationship with revascularisation, deprivation and risk factors. A retrospective review of hospital data. *J R Soc Med* 2014;**107**(12):483–9. <https://doi.org/10.1177/0141076814557301>
3. Nikolajsen L, Jensen TS. Phantom limb pain. *Curr Rev Pain* 2000;**4**(2):166–70. <https://doi.org/10.1007/s11916-000-0052-0>
4. Whyte A, Carroll L. A preliminary examination of the relationship between employment, pain and disability in an amputee population. *Disabil Rehabil* 2002;**24**(9):462–70. <https://doi.org/10.1080/09638280110105213>
5. The Vascular Priority Setting Partnership. *J Vasc Soc GB Irel* 2021; **1**(Suppl1):S1–S31 <http://doi.org/10.54522/jvsgbi.2021.005> Available from: <https://jvsgbi.com/wp-content/uploads/2021/11/FINAL-WEB-FILE-Vascular-PSP-Report-27.11.21.pdf> [cited 2024 Apr 26]
6. Choi H, Ahn S. Classifications, changes, and challenges of online health information seekers during COVID-19. *Int J Environ Res Public Health* 2021; **18**(18):9495. <https://doi.org/10.3390/ijerph18189495>
7. Statista. Most popular websites worldwide as of November 2023, by total visits. <https://www.statista.com/statistics/1201880/most-visited-websites-worldwide/>
8. Osman W, Mohamed F, Elhassan M, Shoufan A. Is YouTube a reliable source of health-related information? A systematic review. *BMC Med Educ* 2022; **22**(1):382. <https://doi.org/10.1186/s12909-022-03446-z>
9. Haslam K, Doucette H, Hachey S, *et al.* YouTube videos as health decision aids for the public: an integrative review. *Can J Dent Hyg* 2019;**53**(1):53–66.
10. Eysenbach G. How do consumers search for and appraise health information on the world wide web? Qualitative study using focus groups, usability tests, and in-depth interviews. *BMJ* 2002;**324**(7337):573–7. <https://doi.org/10.1136/bmj.324.7337.573>
11. Kaur A, Guan Y. Phantom limb pain: a literature review. *Chin J Traumatol* 2018; **21**(6):366–8. <https://doi.org/10.1016/j.cjtee.2018.04.006>
12. Altun A, Askin A, Sengul I, Aghazada N, Aydin Y. Evaluation of YouTube videos as sources of information about complex regional pain syndrome. *Korean J Pain* 2022;**35**(3):319–26. <https://doi.org/10.3344/kjp.2022.35.3.319>
13. Kwan JY, Stocco F, Scott DJA, Bailey MA, Coughlin PA. Assessment of internet-based information on statin therapy. *Eur J Cardiovasc Nurs* 2024; **23**(2):115–21. <https://doi.org/10.1093/eurjcn/zvad061>
14. Koller U, Waldstein W, Schatz KD, Windhager R. YouTube provides irrelevant information for the diagnosis and treatment of hip arthritis. *Int Orthopaed (SICOT)* 2016;**40**(10):1995–2002. <https://doi.org/10.1007/s00264-016-3174-7>
15. Wong M, Desai B, Bautista M, Kwon O, Kolodychuk N, Chimento G. YouTube is a poor source of patient information for knee arthroplasty and knee osteoarthritis. *Arthroplasty Today* 2019;**5**(1):78–82. <https://doi.org/10.1016/j.artd.2018.09.010>
16. Yadav AK, Prasanna Kumar GS, Khanna V. YouTube – an unreliable source of information for total hip replacement. *J Clin Orthopaed Trauma* 2021;**13**:82–4. <https://doi.org/10.1016/j.jcot.2020.08.010>
17. Kara M, Ozduran E, Mercan Kara M, Hanci V, Erkin Y. Assessing the quality and reliability of YouTube videos as a source of information on inflammatory back pain. *PeerJ* 2024;**12**:e17215. <https://doi.org/10.7717/peerj.17215>
18. Ozduran E, Büyükcoban S. A content analysis of the reliability and quality of YouTube videos as a source of information on health-related post-COVID pain. *PeerJ* 2022;**10**:e14089. <https://doi.org/10.7717/peerj.14089>
19. Zhang X, Yang Y, Shen YW, *et al.* Quality of online video resources concerning patient education for neck pain: a YouTube-based quality-control study. *Front Public Health* 2022;**10**:972348. <https://doi.org/10.3389/fpubh.2022.972348>
20. Bernard A, Langille M, Hughes S, Rose C, Leddin D, Veldhuyzen Van Zanten S. A systematic review of patient inflammatory bowel disease information resources on the world wide web. *Am J Gastroenterol* 2007;**102**(9):2070–7.

- <https://doi.org/10.1111/j.1572-0241.2007.01325.x>
21. Cakmak F, Ozkan S, Ipekci A, *et al*. Transition from pandemic to infodemic: an analysis of Turkish language COVID-19 YouTube videos. *East Mediterr Health J* 2021;**27**(5):443–51. <https://doi.org/10.26719/emhj.21.011>
 22. Agency for Healthcare Research and Quality. The Patient Education Materials Assessment Tool (PEMAT) and User's Guide. <https://www.ahrq.gov/health-literacy/patient-education/pemat-av.html>
 23. Silberg WM, Lundberg GD, Musacchio RA. Assessing, controlling, and assuring the quality of medical information on the Internet: Caveant Lector et Viewor—Let the Reader and Viewer Beware. *JAMA* 1997;**277**(15):1244–5. <https://doi.org/10.1001/jama.1997.03540390074039>
 24. Patricio M, Ferreira F, Oliveiros B, Caramelo F. Comparing the performance of normality tests with ROC analysis and confidence intervals. *Communications in Statistics - Simulation and Computation* 2017;**46**(10):7535–51. <https://doi.org/10.1080/03610918.2016.1241410>
 25. Mishra P, Pandey C, Singh U, Gupta A, Sahu C, Keshri A. Descriptive statistics and normality tests for statistical data. *Ann Cardiac Anaesth* 2019;**22**(1):67–72. https://doi.org/10.4103/aca.ACA_157_18
 26. 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 Info Sci Technol* 2015;**66**(10):2071–84. <https://doi.org/10.1002/asi.23311>
 27. Lee JS. YouTube as a source of patient information on gallstone disease. *World J Gastroenterol* 2014;**20**(14):4066–70. <https://doi.org/10.3748/wjg.v20.i14.4066>
 28. Griffiths KM, Tang TT, Hawking D, Christensen H. Automated assessment of the quality of depression websites. *J Med Internet Res* 2005;**7**(5):e59. <https://doi.org/10.2196/jmir.7.5.e59>

Appendix 1 Table presenting the intraclass correlation coefficients

	Intraclass correlation
GQS score	0.925 (95% CI 0.852-0.962)
Subjective score	0.936 (95% CI 0.875-0.968)
PEMAT A/V Understandability score	0.933 (95% CI 0.868-0.966)
PEMAT A/V Actionability score	0.977 (95% CI 0.954-0.988)
JAMA Benchmark criteria	0.936 (95% CI 0.875-0.968)

Correlation coefficient of >0.7 is considered acceptable. >0.8 is considered good. >0.9 is considered excellent
GQS, Global Quality Score; *JAMA*, *Journal of American Medical Association*; PEMAT A/V, Patient Education Materials Assessment Tool for Audio-Visual Materials.