

Information Quality of Nutrition Videos on YouTube

Karen Malin Krüger, Joachim Griesbaum, Linda Achilles

University of Hildesheim, Germany

{krueger003, griesbau, achilles}@uni-hildesheim.de

Abstract

This paper investigates the quality of German-language nutrition videos on YouTube. For that purpose, a qualitative analytical scheme is developed that combines the measurement of a) the inherent quality of 32 videos using the PRinciples for Health-related Information on Social Media (PRHISM) Scoring Tool and the Patient Education Materials Assessment Tool for Audiovisual Materials (PEMAT-A/V) with b) information provided by 800 user comments that indicate perceived helpfulness on the basis of a categorization scheme that was developed deductively and inductively. The selected sample reflects typical user behavior by selecting the highest ranked videos and videos recommended by YouTube. The results show that the videos have a high average information quality regarding formal quality criteria (e.g., authorship, attribution, balance and justifiability), understandability, and actionability. The analysis of user comments reveals that appreciation and/or acknowledgement, and positive feedback are among the most used types of comments, which implies that the videos are assessed as helpful. Furthermore, comments are valuable knowledge additions to the videos as they provide suggestions for improving the videos and descriptions of personal experiences. Overall, the value of the investigation is twofold. First, it shows that in contrast to the widespread criticism of the quality of health information on YouTube, in the investigated topical area of nutrition videos the information quality can rather be assessed as good. Further, user-generated content in form of comments provides additional informational value. From a user perspective, especially channels with the YouTube Health label seem to be a safe bet regarding information quality.

Keywords: health information; information quality; HIQ; understandability; actionability; YouTube; nutrition videos; dietary recommendations; obesity; YouTube Health; qualitative analysis; comment analysis

1 Introduction

The video platform YouTube is widely used as a digital health service in Germany (Baumann et al., 2021; YouTube, n.d. a). At the same time, studies on a wide range of health topics have repeatedly shown that YouTube is not a reliable source of health information (Madathil et al., 2015; Osman et al., 2022). As a response to the criticism that health-related misinformation is being spread on the platform, YouTube Health was launched in the United States in 2021 and in Germany in 2023 (Graham, 2021, 2022; Focken, 2023). YouTube Health comprises two functions, a) a Health Label, which is an information panel below the videos that provides context information on credible health sources, and b) a Health Shelf, which highlights videos from credible health sources in the search results (Graham, 2021; Focken, 2023; YouTube, n.d. b).

One health issue that is playing an increasingly important role in society is obesity (Müller et al., 2022; Schneider & Holzwarth, 2022). As of 2022, over one billion people worldwide are considered obese (NCD-RisC, 2024; WHO, 2024). Around half of adults (53.5%) aged 18 and over in Germany are overweight, including obese, and 19.0% are obese (Schienkiewitz et al., 2022). An important component in the treatment of obesity is a change in diet (Kasper & Burghardt, 2021: 275; Lautenbach & Aberle, 2022; Oberländer & Weimann, 2022; Hahn, 2023). Hence, this study focuses on the information quality of German-language YouTube videos on dietary recommendations for overweight or pre-obesity and obesity. To the authors' knowledge, no such study exists in German-speaking countries.

This leads to the question of how information quality can be measured. Sun et al. (2019) state that the evaluation of health information quality (HIQ) can be summarized as a process of applying criteria to assess information. A basic differentiation can be made between objective and subjective evaluation of information (Zhang & Kim, 2022). The subjective assessment of HIQ by users can vary depending on age, gender, level of education, health status, level of (digital) health literacy and attitude to and knowledge about the respective health topic (e.g., Diviani et al., 2015; Tao et al., 2017; Sun et al., 2019; Zhang & Kim, 2022). Hence, a popular conceptualization of HIQ from the user perspective is *fitness for use* (Stvilia et al., 2009; Tao et al., 2017; Sun et al., 2019; Zhang & Kim, 2022; Afful-Dadzie et al., 2023). Consequently, Zhang and Kim (2022) fundamentally understand HIQ as the user's percep-

tion of the quality of health information. This concept of perceived HIQ overlaps with related concepts such as credibility, trust and usefulness (Sbaffi & Rowley, 2017; Sun et al., 2019; Zhang & Kim, 2022; Liu et al., 2023).

On the other hand, Denniss et al. (2022: 2) define HIQ in a comparatively objective manner as “the reliability of information compared against a set of defined quality criteria”. A wide variety of criteria for evaluating HIQ can be found in literature, as well as numerous ways of describing, categorizing and weighting them (Eysenbach, 2005; Zhang et al., 2015; Tao et al., 2017; Afful-Dadzie et al., 2023). As stated by Zhang et al. (2015) and Tao et al. (2017), this reflects the complexity of the multidimensional construct of HIQ. This lack of conceptual clarity also contributes to a lack of instruments for the effective assessment of HIQ, as Zhang et al. (2015) note. Although numerous evaluation instruments and quality seals exist in practice, many have questionable validity and reliability (Zhang et al., 2015).

Aspects that have become increasingly important in recent years regarding health information aimed at laypersons are understandability and actionability (Shoemaker et al., 2014; Beaunoyer et al., 2017; Vishnevetsky et al., 2018; Denniss et al., 2022). Understandability and actionability are important to make content accessible to broad sections of the population (Shoemaker et al., 2014; Beaunoyer et al., 2017). According to Shoemaker et al. (2014: 396), health information materials are understandable “when consumers of diverse backgrounds and varying levels of health literacy can process and explain key messages” and actionable when they “can identify what they can do based on the information presented”. In this context, the Health Literacy Survey Germany 2 conducted in 2019/2020 found that both health literacy and digital health literacy are low among the majority of the population in Germany (Hurrelmann et al., 2020; Schaeffer et al., 2021).

The research interest of this investigation is on the information quality of nutrition videos on YouTube. To integrate objective and subjective perspectives of information quality, the following research questions (RQ) are covered.

RQ1: What is the inherent information quality of the nutrition videos provided by YouTube?

RQ2: How understandable and actionable are the nutrition videos provided by YouTube?

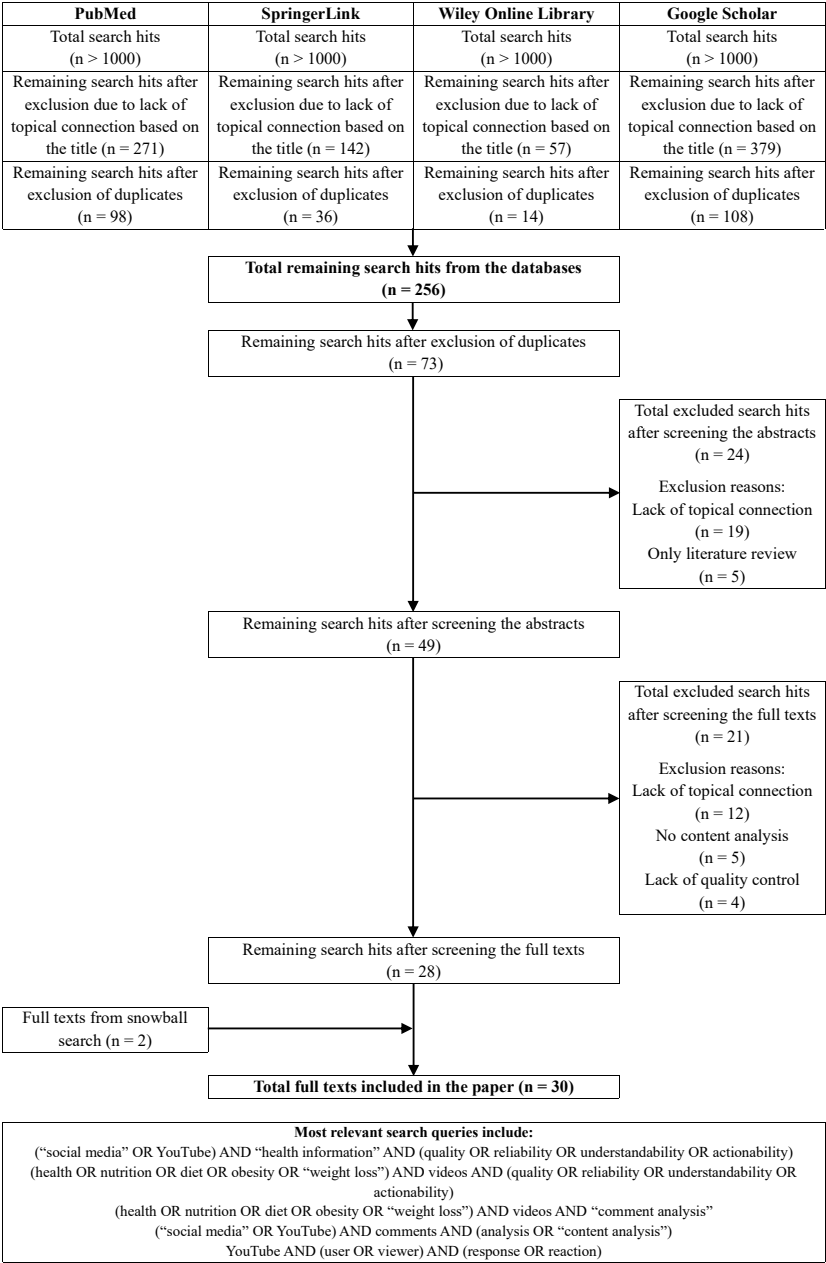
RQ3: Do users perceive the nutrition videos provided by YouTube as helpful?

RQ1 and RQ2 are concerned with objective information quality. Whereas RQ1 captures the inherent quality, RQ2 shows the consideration of laypersons as an audience. RQ3 tries to capture the assessment of the users by analyzing the comments. In addition, we are interested in whether quality indicators such as labels (YouTube Health) can serve as reliable differentiators of information quality.

To answer these questions, the paper is structured as follows: First, an overview of the current state of research on the analysis of health-related videos and associated comments is provided. This is followed by a description of the methodological approach of the empirical study conducted. Afterwards, the data collection and analysis results are presented. The paper concludes with a discussion of the results, the limitations of the analysis and an outlook.

2 Research Background

The literature section aims to uncover which instruments are available to analyze videos and associated comments. Figure 1 delineates the literature research process, which took place in December 2023. Out of the 30 studies selected for this paper, 18 relate to the current state of research on video analysis and twelve on comment analysis. Only studies published in English or German and from 2010 onwards were considered. The restriction regarding the publication date was set to ensure relevance of the content. Additionally, the studies should have been subjected to some form of quality control, such as peer review. Furthermore, the studies should not only consist of a literature review but also include an empirical analysis. For the video analysis, only studies that deal with videos on overweight and weight reduction, dietary recommendations for other diseases or other nutritional topics were selected. However, as no relevant studies on these topics could be identified for the comment analysis, studies on any health-related topic were considered relevant in this case. Anyhow, only studies in which a content analysis of the comments was conducted were included.



2.1 Research focusing on videos

Overall, the literature analysis shows that there is a multitude of instruments to measure the inherent quality of health information. The finding that the DISCERN, Global Quality Score (GQS) and Journal of the American Medical Association (JAMA) benchmarks appear to be the most used cross-topic instruments for evaluating the HIQ of YouTube videos (see Table 1) is consistent with other literature reviews such as Drozd et al. (2018), Osman et al. (2022) and Afful-Dadzie et al. (2023). In addition, according to Drozd et al. (2018), Osman et al. (2022) and Afful-Dadzie et al. (2023), the comparison against topic-specific guidelines or the use of specially developed topic-specific scoring systems are not uncommon (Table 1). Afful-Dadzie et al. (2023) criticize that frequently more than one cross-topic evaluation instrument is applied (Table 1), without explaining why this is deemed necessary.

Moreover, the use of these evaluation instruments is in itself not unproblematic: The DISCERN by Charnock et al. (1999) is an instrument for assessing the quality of written health information directed at consumers about treatment decisions. The JAMA benchmarks for websites were published by Silberg et al. (1997) and utilize the criteria authorship, attribution, disclosure and currency as quality criteria. Azer (2020), Denniss et al. (2022) and Guler and Aydin (2022) state that a major point of critique regarding the use of these two instruments is that they were developed at the end of the last century for rather static information environments such as brochures and websites, not for social media environments, and have hardly been revised since then. In some studies considered in the current state of research an adapted version of the DISCERN is used (Table 1), referred to as modified DISCERN (mDISCERN), though only the number of items was reduced to five and the scoring system was adapted (Singh et al., 2012). The GQS, developed by Bernard et al. (2007), was originally intended for evaluating the quality of medical websites, which Boté (2019) criticizes as well. In contrast, Lee et al. (2014) developed the Usefulness Score for videos. However, both instruments are symptomatic of the problem described by Zhang et al. (2015) that instruments often consist of items that cannot be assessed objectively, which leads to low reliability.

Due to the resulting lack of suitable instruments for evaluating videos, Guler and Aydın (2022) developed the Medical Quality Video Evaluation Tool (MQ-VET). Nonetheless, the MQ-VET exhibits the same subjectivity problem as the two previously mentioned instruments. As videos on the topic of nutrition are the subject of this study, instruments for nutrition-related health information could be considered instead: One would be the Quality of World Wide WEB articles on nutrition (QWEB) tool by Kriz et al. (2019) and another the Online Quality Assessment Tool (OQAT) by Ellis et al. (2023). However, both instruments, QWEB and OQAT, were not developed for videos, but exclusively for online articles and blog entries, as both Kriz et al. (2019) and Ellis et al. (2023) point out.

This narrows the selection process down to a tool developed by Denniss et al. (2022) based on 13 principles for HIQ on social media, namely authorship, authoritativeness, action orientation, financial disclosure, attribution, balance and justifiability, risks and benefits, privacy, complementary information, referrals and support, readability and comprehensibility, accessibility and images. After Denniss et al. (2022), the PRHISM Scoring Tool is designed to be applied by researchers and health professionals to content on social media of all kinds, including videos. The tool itself is, as stated by Denniss et al. (2022), based on the DISCERN, JAMA, Health On the Net Foundation code of conduct (HONcode) for websites (Boyer et al., 1998) and Quality Evaluation Scoring Tool (QUEST), developed and validated by Robillard et al. (2018) for the evaluation of online articles. The PRHISM Scoring Tool was developed by means of a Delphi study (Denniss et al., 2022). Yet, Denniss et al. (2022) do not mention a review of the reliability and validity of the tool otherwise. However, Denniss et al. (2024) used the PRHISM tool to evaluate the information quality of 676 Instagram posts with nutrition-related health information, achieving an acceptable inter-rater agreement.

In contrast to HIQ, all studies that examine the understandability and actionability of health videos use the same instrument for evaluation (see Table 1), the PEMAT-A/V by Shoemaker et al. (2014) for use by researchers and health professionals. This can be attributed to the fact that, to the best of the authors' knowledge, there is no other instrument to date that primarily serves to assess the understandability and actionability of health-related videos, as Beaunoyer et al. (2017) and Vishnevetsky et al. (2018) have determined as well. Shoemaker et al. (2014) state that a panel of experts confirmed the content validity of both the understandability and the actionability scale. Both

PEMAT scales also exhibit a high level of internal consistency (Shoemaker et al., 2014). Furthermore, the inter-rater reliability of the PEMAT is moderate to high (Shoemaker et al., 2014; Vishnevetsky et al., 2018).

2.2 Research focusing on user comments

Regarding the comment analysis, twelve studies were selected: Frohlich and Zmyslinski-Seelig (2012), Lewis et al. (2012), Basch et al. (2015), Godwin et al. (2017), Marcon and Caulfield (2017), Meldrum et al. (2017), Loeb et al. (2019), Dubovi and Tabak (2020), Loeb et al. (2021), Shukla (2021), Li et al. (2022) and Bakombo et al. (2023). Overall, the studies indicate that comments are often used to share personal information and experiences, but also to express emotions and opinions and to give feedback on the videos. Moreover, both personal and content-related questions are asked and answered in the comments and situational advice is requested. Further, Marcon and Caulfield (2017) as well as Dubovi and Tabak (2020) found that the comment section is not only used for feedback, sharing experiences and social support, but can also be a place for factual discussions and knowledge sharing on certain topics. In terms of methodology, content analysis category systems differ depending on the research objective and topic, but deductive and inductive approaches were often combined. In addition, coding was carried out manually in all studies, usually with several coders.

3 Methods and Research Design

For RQ1, concerning the HIQ of the nutrition videos, the PRHISM tool by Denniss et al. (2022) is used. The PRHISM tool consists of 13 items (see Table 2), each of which is rated on a five-point Likert scale (Denniss et al., 2022). The evaluation is carried out in accordance with the PRHISM Guide provided by Denniss et al. (2022). For both RQ1 and RQ2, the video descriptions are included in the analysis alongside the videos. One exception is item 11 of the PRHISM because the readability of the video descriptions is not considered relevant to the study.

Table 2: PRHISM Items (Denniss et al., 2022)

No.	Item
1	Principle 1: Authorship When providing health-related information on social media, the authors and contributors, their credentials, affiliations, and contact information should be clearly stated on the social media profile.
2	Principle 2: Authoritative Health-related information provided on social media should be given by qualified professionals, including health and medical scientists, and information should be within the scope of practice of the author's qualifications. If information is provided by an unqualified person, this should be clearly indicated.
3	Principle 3: Action-orientated Health-related information provided on social media should be action-orientated and include clear, succinct messages to support decision making and provide context for the consumer.
4	Principle 4: Financial Disclosure Sponsorship, advertising, funding arrangements, financial support or any potential conflicts of interest should be fully disclosed in a prominent and clear manner.
5	Principle 5: Attribution Health-related information on social media should include clear references and hyperlinks to the original source(s) of information used to compile the post.
6	Principle 6: Balance & Justifiability Health-related information provided on social media that includes claims relating to the benefits/performance of a particular treatment, product, service or behaviour should be balanced, unbiased and supported by appropriate and quality evidence.
7	Principle 7: Risks & Benefits Health-related information provided on social media about a particular treatment, product, service or behaviour should clearly outline associated risks and benefits.
8	Principle 8: Privacy Health-related information on social media should respect principles of privacy and confidentiality.
9	Principle 9: Complementary Information Health-related information provided on social media should provide support for individuals' relationships with their doctor and other professional healthcare providers and should not be designed to replace such relationships. Support for discussion of options with the individuals' healthcare provider should be included in posts containing health-related information.
10	Principle 10: Referrals & Support Health-related information provided on social media should include referrals to additional sources of support and information.
11	Principle 11: Readability & Comprehensibility Health-related information on social media should avoid the use of technical language and medical jargon. Plain language should be used and information should be easily understandable by the general public.
12	Principle 12: Accessibility Health-related information provided on social media should be accessible to vision- and hearing-impaired individuals.
13	Principle 13: Images Images included in health-related social media posts should be visually appealing and reflect rather than contradict the information provided in the post.
Response options per item: Completely unmet = 0; Partially met = 1, 2 or 3; Completely met = 4; Not applicable = N/A* * only for items 5, 6, 7, 8, 9, 12 and 13	

In terms of RQ2, regarding the understandability and actionability of the videos, the PEMAT-A/V by Shoemaker et al. (2014) is applied. The PEMAT-A/V consists of two scales (Table 3): an understandability scale with 13 items and an actionability scale with four items, which are rated on a binary scale (Shoemaker et al., 2014). The PEMAT User Guide by Shoemaker et al. (2013) contains specific recommendations for the scoring of the items.

Table 3: PEMAT-A/V Items (Shoemaker et al., 2013)

Understandability Scale	
No.	Item
1	The material makes its purpose completely evident.
3	The material uses common, everyday language.
4	Medical terms are used only to familiarize audience with the terms. When used, medical terms are defined.
5	The material uses the active voice.
8	The material breaks or “chunks” information into short sections.
9	The material’s sections have informative headers.
10	The material presents information in a logical sequence.
11	The material provides a summary.
12	The material uses visual cues (e.g., arrows, boxes, bullets, bold, larger font, highlighting) to draw attention to key points.
13	Text on the screen is easy to read.
14	The material allows the user to hear the words clearly (e.g., not too fast, not garbled).
18	The material uses illustrations and photographs that are clear and uncluttered.
19	The material uses simple tables with short and clear row and column headings.
Actionability Scale	
No.	Item
20	The material clearly identifies at least one action the user can take.
21	The material addresses the user directly when describing actions.
22	The material breaks down any action into manageable, explicit steps.
25	The material explains how to use the charts, graphs, tables, or diagrams to take actions.
Response options per item: Disagree = 0; Agree = 1; Not applicable = N/A*	
* only for items 8, 9, 11, 12, 13, 14, 18, 19 and 25	

It should be noted that the first author of this paper and evaluator of the videos has no medical education. Accordingly, the PRHISM tool and the PEMAT-A/V are chosen as existing evaluation instruments rather than developing an own topic-specific scoring system. This is also why the criterion

of accuracy of health information is not considered in this paper despite its importance for HIQ (Zhang et al., 2015), which could be evaluated using topic-specific guidelines for instance.

To answer RQ3, regarding the perceived helpfulness of the nutrition videos, a qualitative content analysis of the comments using content-related categories is conducted (Scheufele & Engelmann, 2009, pp. 165–169). In Mayring's established approach (Döring, 2023: 533), qualitative content analysis comprises not only content aspects but also formal ones and can include quantitative analysis parts, for example, determining category frequencies (Mayring & Fenzl, 2022). The structuring technique is used to analyze the comments (Mayring, 2022, pp. 66–67). Further, a coding guide with category definitions, anchor examples and coding rules is developed (Mayring, 2022: 96–99).

The category system is mainly based on Meldrum et al. (2017), as they pursued the same research question, albeit on a different health topic, and achieved a high intercoder agreement (Meldrum et al., 2017). The category system of Meldrum et al. (2017) is primarily based on that of Madden et al. (2013). The latter developed a comprehensive, cross-topic content-analytical classification scheme for comments on YouTube, which has a high intercoder reliability (Madden et al., 2013). The main categories of Meldrum et al. (2017) form the starting point of the category system of this paper, which consists of main categories and subcategories. Additionally, the comprehensive category system of Madden et al. (2013) is used, especially concerning the subcategories. However, the category systems of the other eleven studies mentioned in the current state of research (see Section 2) are also considered in the development of the category system. The main categories for RQ3 are presented in Table 4. Figure 2 provides an overview of the research design.

Table 4: Main categories for comment analysis

Main Category*	Definition*	Example*
Question	A comment in which a person asks for or requests information, ideas or support. This does not have to be in question format.	<i>In what time period, if I may ask? :D</i>
Response	A comment in which a person offers information, ideas or support in response to a question or request.	<i>I think in about two and a half months.</i>
General feedback	A comment in which a person expresses a reaction or opinion about the video, the topic, the channel or another person. This can also be unsolicited. The comment is of general or neutral nature.	<i>Thank you very much for the video. I think it's also very important to mention that it's essential to keep your everyday life as "stress-free" as possible and to relax regularly. For example, I always put on weight in stressful times because I get a craving for sweets and fatty foods.</i>
Positive feedback	A comment in which a person expresses an opinion, reaction or feelings about the video, the channel or another person who follows or confirms the recommendations mentioned in the video. This can also be unsolicited. The comment is positive in nature, the person expresses agreement or disagrees with another person who makes a negative comment about the video, the channel or a person who follows or confirms the recommendations mentioned in the video.	<i>What I find extremely positive here is that the psyche and habits from childhood are included and no radical change is demanded from one day to the next. The fact that carbohydrates are not completely demonized is also very pleasant. I wish them both continued success.</i>
Negative feedback	A comment in which a person expresses an opinion, reaction or feelings about the video, the channel or another person who follows or confirms the recommendations mentioned in the	<i>Hello to the directors and nutritionists. I think it's a bit of a shame that carbohydrates are being lumped together. Potatoes, for example, have far fewer carbohydrates/kg</i>

Main Category*	Definition*	Example*
	video. This can also be unsolicited. The comment is negative in nature, the person expresses disagreement or agrees with another person who makes a negative comment about the video, the channel or a person who follows or confirms the recommendations mentioned in the video.	<i>than white flour. And even white flour is not fundamentally wrong if the amount is greatly reduced.</i>
Appreciation and/or acknowledgement	A comment in which a person expresses appreciation and/or acknowledgement for the video, the channel or another person.	<i>Thank you for the great video</i> 👍
		<i>That's a really great achievement, you can be rightly proud of yourself!</i>
Personal experiences and/or information	A comment in which a person describes personal experiences or the experiences of people close to them and/or which contains information about themselves or people close to them.	<i>I have been living on low carb for 5 years with intermittent fasting and HIIT for 25 minutes almost every day. I have my diabetes 2 (for 20 years) under control to the extent that one pill a day is enough. I have reduced my weight from 100kg to 88kg and have kept it stable for a long time. This is the lifestyle that I will maintain.</i>
Personal action	A comment in which a person states what they will or will not do or have already done or not done after watching the video or reading other comments.	<i>After your video, I started intermittent fasting, followed your tips, feel good and am already seeing the first positive results</i> 👍
Advertisement	A comment in which a person promotes a product or service related to the video or topic.	— **

* The category names, category definitions and examples have been translated from German into English.

** No example is given for this category because it would allow conclusions to be drawn about specific individuals.

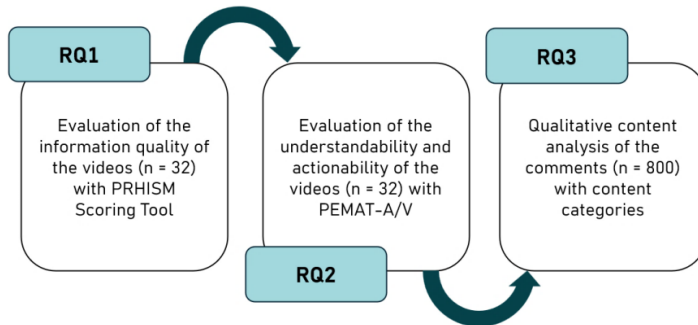


Fig. 2 Overview of research design

4 Data Analysis and Results

This section starts with the description of the sample selection process. For the query and video selection, the Google Chrome browser was used. The user logged out of their Google account, cleared the cache and cookies and deactivated the search and playback history to reduce personalization effects (Boté 2019). The keyword research tool Google Trends and YouTube’s auto-complete function were used to identify suitable search queries. As per literature reviews by Sampson et al. (2013) and Boté (2019), Google Trends is used in studies to compare the usage frequency of different search queries on YouTube over time. The YouTube query autocomplete feature is mentioned by Drozd et al. (2018) and Boté (2019) as a function used in comparable studies to determine search queries. As is also common in other studies according to Drozd et al. (2018), not only one but several search queries are selected in this study, four in total.

The four search queries were selected on 29 January 2024. To this end, the terms *Übergewicht* (overweight), *Präadipositas* (pre-obesity) and *Adipositas* (obesity) and its synonyms *Fettleibigkeit* (obeseness), *Fettsucht* (adiposeness) and *Obesitas* (adiposity) were entered into Google Trends, with *Übergewicht* (overweight) and *Adipositas* (obesity) proving to be the most frequently used terms. The suggestions from YouTube’s query autocomplete feature were therefore checked for these two terms. This was also carried out with the terms *Ernährung* (diet) and *Bauchfett* (belly fat) because visceral

abdominal fat plays a key role in the health risk of obesity as a nutrition-associated disease (Kasper & Burghardt, 2021: 272–273; Nimptsch & Pischon, 2022; Hahn, 2023). The search queries that were categorized as thematically appropriate were then compared using Google Trends to determine their frequency. Ultimately, the following search queries were selected: *Adipositas Ernährung* (obesity diet), *Übergewicht abnehmen* (lose overweight), *Ernährung abnehmen* (lose weight through diet) and *Bauchfett verlieren* (lose belly fat).

For the video selection, primarily the search bar integrated into the YouTube platform was used. For each search query, the first ten videos in the search results and, if applicable, the videos in the Health Shelf were checked to see whether they meet the selection criteria. The default settings for sorting the search results, i.e., by relevance (Sui et al., 2022), were retained and no search filters were set, which is typically handled the same way in comparable studies (Madathil, 2015; Drozd et al., 2018; Boté, 2019). Furthermore, for each video that met the selection criteria, the first three videos recommended by YouTube were checked. After Sampson et al. (2013), Madathil et al. (2015) and Boté (2019), this strategy simulates real user behavior. In addition, the first 25 comments for each selected video were used for the analysis, whereby the default settings were retained, i.e., sorting by top comments (Sui et al., 2022). These 25 comments per video also include responses to other comments.

4.1 Video analysis

The four search queries were entered in YouTube on 30 January 2024. Videos were selected from the search results and video recommendations and saved locally. The video selection process is shown in Figure 3.

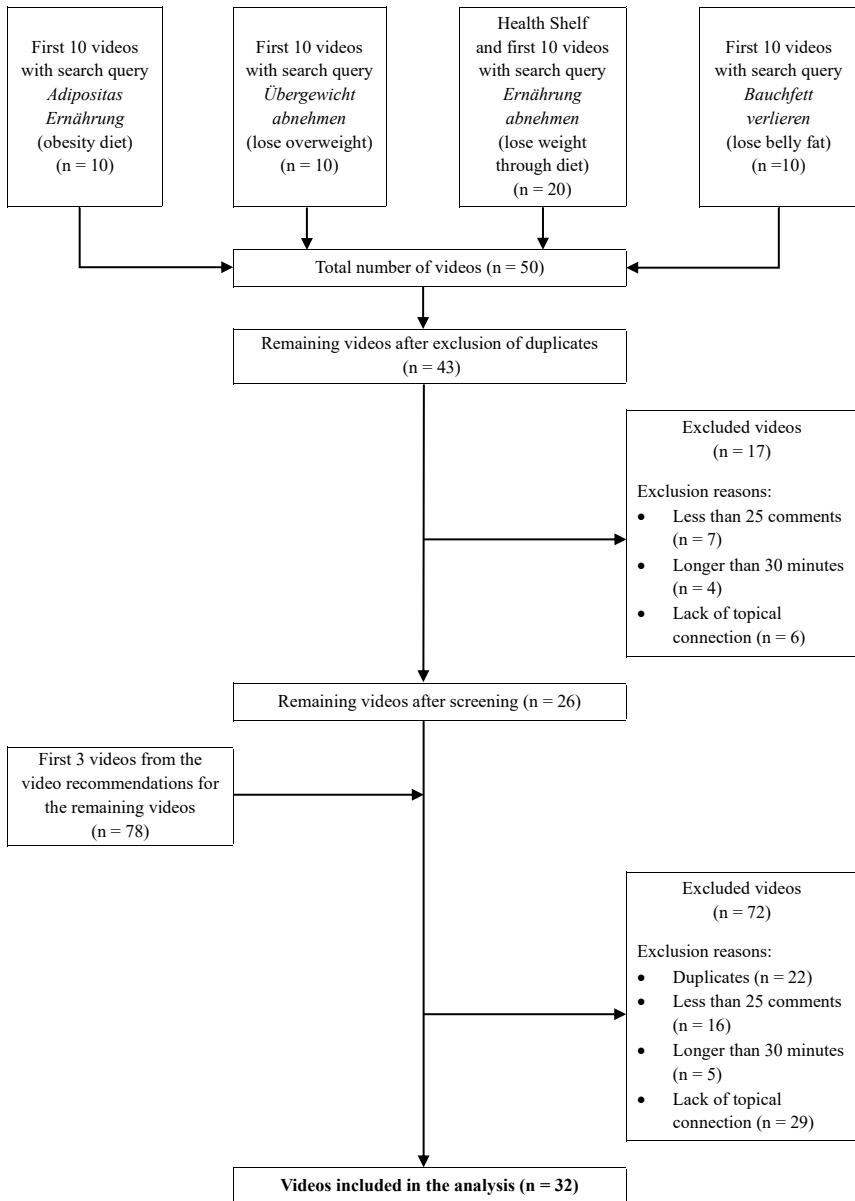


Fig. 3 Flowchart of the video selection process

To describe the sample in more detail, it was examined who the information providers are. Regarding the kinds of channels that uploaded the nutrition videos (Fig. 4), it can be stated that more than a third of the videos are from the media (37.50%; 12/32). Around a fifth of the videos were uploaded by physicians (21.88%; 7/32), followed by laypersons (18.75%; 6/32). Only 9.38% (3/32) of the videos were uploaded by non-physician health professionals, 6.25% (2/32) by health organizations and medical institutions, 3.13% (1/32) by government institutions and organizations and 3.13% (1/32) by commercial companies.

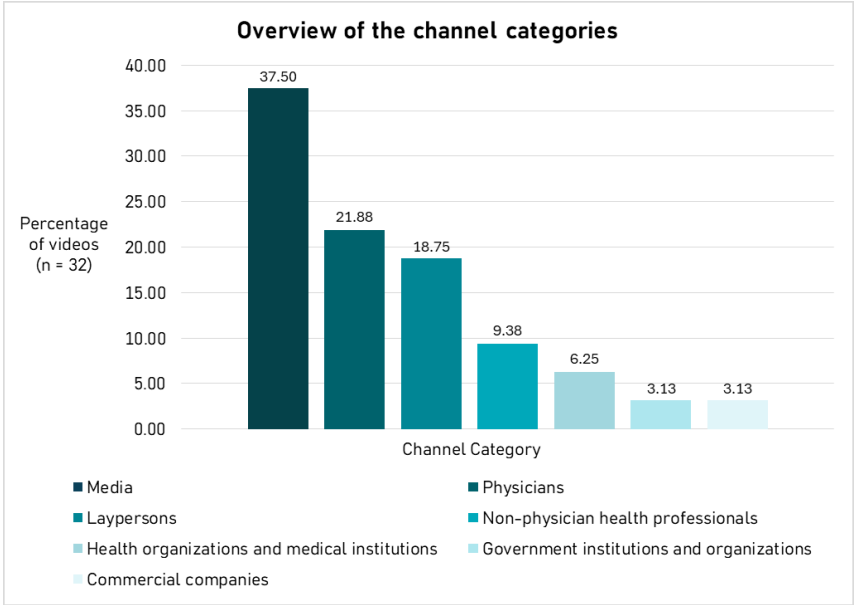


Fig. 4 Overview of the channel categories

In terms of the speakers communicating health information in the videos (see Fig. 5), the results are as follows: In nearly two thirds of the 32 videos (62.50%; 20/32) physicians share the health information, followed by off-speakers (34.38%; 11/32), laypersons (18.75%; 6/32) and finally non-physician health professionals (15.63%; 5/32). In 23 of the 32 videos (71.88%) there are only speakers of one category, in eight videos (25.00%) of two categories and in one video (3.13%) of three categories. Exclusively in the videos from the media there are speakers from more than one category. In eleven (91.67%) of the twelve videos assigned to the media category, physicians provide the

health information, in ten (83.33%) off-screen speakers and in one (8.33%) a non-physician health professional.

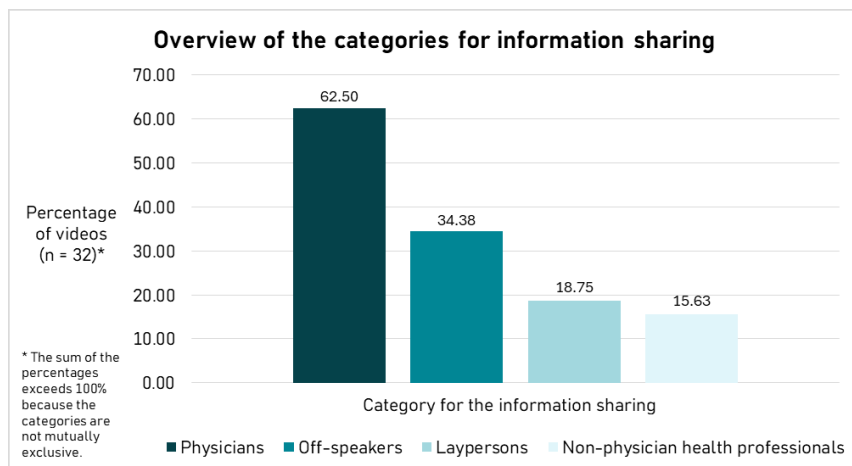


Fig. 5 Overview of the categories for information sharing

The average PRHISM score of the 32 videos is 63.68 (SD 15.02; minimum 25.00; maximum 89.58). Following the score classification by Denniss et al. (2022), the HIQ of the videos is therefore good on average. The average PEMAT-A/V understandability score is 84.56% (SD 12.11%; minimum 60.00%; maximum 100.00%) and the average PEMAT-A/V actionability score is 89.84% (SD 14.47%; minimum 66.67%; maximum 100.00%). Both scores are above the threshold value of 70% set by Shoemaker et al. (2014), which means the videos are understandable and actionable on average.

Moreover, almost two thirds of the videos (62.50%; 20/32) are from channels with the YouTube Health label. These 20 videos received a good or excellent rating for their HIQ using the PRHISM tool, which is significantly higher than the PRHISM scores of the other videos ($p < 0.001$). Additionally, these videos have significantly higher points for the quality criteria authorship ($p = 0.032$), authoritativeness ($p < 0.001$), action orientation ($p = 0.002$), attribution ($p = 0.029$), balance and justifiability ($p = 0.003$), risks and benefits ($p < 0.001$) as well as referrals and support ($p < 0.001$). For the PEMAT-A/V understandability score and the PEMAT-A/V actionability score there is no statistically significant difference between videos with and without the YouTube Health label. Regarding the speakers, it was found that the videos in which physicians or non-physician health professionals share the health

information have a statistically significant higher PRHISM score than those with laypersons ($p = 0.004$). All statistical tests were conducted in SPSS using the Mann-Whitney U test.

4.2 Comment analysis

On 31 January 2024, the first 25 comments from each of the 32 videos were selected and saved manually, with a restriction of a maximum of five responses per thread. 46 comments were excluded because they were either non-German comments, consisting solely of emojis or comments with content that was incomprehensible or nonsensical. Instead, the subsequent comment was selected, yet the rule of a maximum of five responses per thread still applied. A total of 800 comments are therefore included in the content analysis. 398 comments (49.75%) are responses to other comments. 89 comments (11.13%) are written by content creators.

The category system was continuously reviewed throughout the coding process and revised considering the current state of research as soon as it was deemed necessary. The final category system consists of nine main categories and 13 subcategories. In general, the categories are not mutually exclusive, unless a specific coding rule precludes it. Once the category system had been finalized, all the codes assigned were reviewed once again and adjusted if necessary. A total of 1079 codes were assigned for the main categories and 327 for the subcategories. The results regarding the comment categories can be seen in Table 5 and an overview of the main comment categories in Figure 6. All coding was done by the first author of this paper. Intracoder reliability was determined for 20% of the comments in November 2024 by re-coding the first five comments of every video. To determine the intracoder reliability, the presence (1 = presence, 0 = absence) of each category within a comment was counted. The average Cohen's kappa of the main categories is $\kappa = 0.96$ and of the subcategories $\kappa = 0.97$, which represents an almost perfect intracoder agreement in each case (Landis & Koch, 1977).

Table 5: Overview of the results for the comment categories

Category**	Code frequency of main categories ($n = 1079$), n (%)	Number of codes in subcategories ($n = 327$)	Percentage of subcategory in main category	Percentage of comments ($n = 800$)*
Question	74 (6.86)	–	–	9.25
Video-related question	–	20	27.03	2.50
Request for situation-specific advice	–	11	14.86	1.38
Question about personal experiences	–	35	47.30	4.38
Further inquiry	–	8	10.81	1.00
Response	23 (2.13)	–	–	2.88
Response to video-related question	–	7	30.43	0.88
Situation-specific advice	–	7	30.43	0.88
Response to question about personal experiences	–	8	34.78	1.00
Response to further inquiry	–	1	4.35	0.13
General feedback	230 (21.32)	–	–	28.75
Suggestion regarding the video	–	25	10.87	3.13
Suggestion for future content	–	26	11.30	3.25
Unsolicited advice on the topic	–	39	16.96	4.88
Opinion on the topic	–	71	30.87	8.88
Other remark	–	69	30.00	8.63
Positive feedback	159 (14.74)	–	–	19.88
Negative feedback	37 (3.43)	–	–	4.63
Appreciation and/or acknowledgement	284 (26.32)	–	–	35.50
Personal experiences and/or information	221 (20.48)	–	–	27.63
Personal action	45 (4.17)	–	–	5.63
Advertisement	6 (0.56)	–	–	0.75
* The sum of the percentages exceeds 100% because the categories are not mutually exclusive.				
** The category names of the main categories and the subcategories have been translated from German into English.				

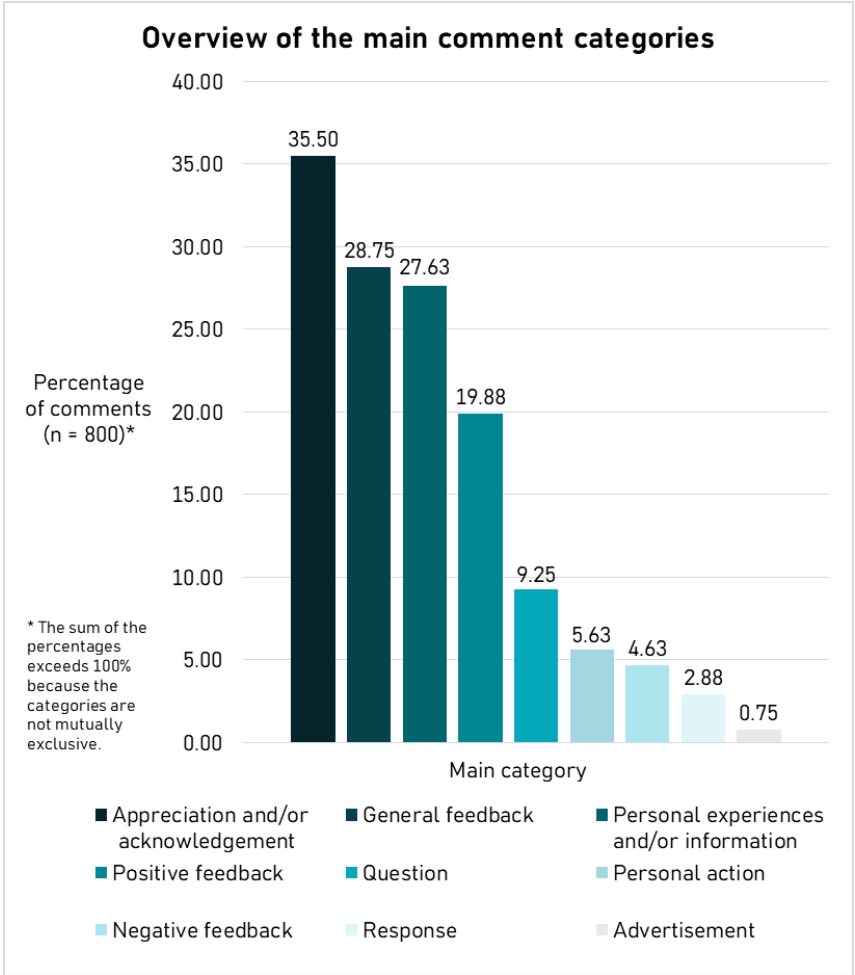


Fig. 6 Overview of the main comment categories

5 Discussion and Conclusion

5.1 Video Analysis

The consensus in the studies included in the current state of research (see Table 1) is, with a few exceptions, that the HIQ of YouTube videos with a

nutritional focus is low on average. However, the HIQ of the YouTube nutrition videos analyzed for RQ1 in this study is good on average, as indicated by the PRHISM score. With regard to the comparability of the results, the following aspects should be considered: Firstly, although all studies are related to nutrition or nutrition-associated diseases, the specific topics differ from each other. Also, none of the studies directly addressed videos on dietary recommendations for pre-obesity and obesity like this study. Secondly, videos in different languages, mostly English, were analyzed in the studies. Thirdly, in this study, the PRHISM tool by Denniss et al. (2022) was used, a new instrument for evaluating HIQ on social media, which was not used by any of the studies from the current state of research, meaning that the results on HIQ are only comparable to a limited extent.

Moreover, the higher HIQ can possibly be attributed to the channels and speakers: As Batar et al. (2020), Batar et al. (2021), Calisir and Ece (2022), Kiss et al. (2023), Rana and Arora (2023) and Segado-Fernández et al. (2023), among others, have found, the HIQ of videos uploaded by health organizations, institutions and professionals is generally significantly higher than that of channels without a professional background in the health sector. A total of 37.50% (12/32) of the videos analyzed in this study originate from physicians, non-physician health professionals or health organizations and medical institutions. Still, at 37.50%, a considerable proportion of the videos analyzed is from the media sector. However, judging by the PRHISM score, the quality of these twelve videos, all of which were uploaded by channels operated by public broadcasters, is good or even excellent in some cases. This can presumably be explained by the fact that in 91.67% of the videos originating from the media, physicians communicate health information. As stated by Kiss et al. (2023) as well as Şahin and Bayır Garbioğlu (2023), the HIQ of videos is significantly higher if the speakers are health professionals, which is confirmed in this study. Overall, 78.13% (25/32) of the videos feature physicians or non-physician health professionals providing health information.

For RQ2, it can be stated that the analyzed videos are understandable and actionable on average, as the PEMAT-A/V scores indicate. However, as with HIQ, most of the studies included in the current state of research come to the opposite conclusion. One exception is Lambert et al. (2017), whose analyzed videos also show a high level of understandability and actionability. This could be related to the channels the videos originate from and the speakers in the videos. Most of the videos in the study by Lambert et al. (2017) were uploaded by commercial companies. In the present study a large proportion

are from the media sector, which could be an explanation for the overall high level of understandability and actionability, as all these videos are considered understandable and actionable based on the PEMAT-A/V.

5.2 Comment analysis

In terms of RQ3, it can first be noted that around one in five comments (19.88%) contains positive feedback and only 4.63% of the comments contain negative feedback, which, according to Meldrum et al. (2017), indicates usefulness of the videos for the users. Moreover, only 3.13% of the comments are suggestions regarding the videos, such as additions to the content, which nevertheless could be useful information for others. Around a quarter of the positive feedback (26.42%; 42/159) overlaps with the category *Personal experiences and/or information*. For example, users justify their positive feedback with their experiences and thus confirm the video content.

Furthermore, around one in three comments (35.50%) is of an appreciative or acknowledging nature. Studies concerning videos on other health topics such as Lewis et al. (2012), Meldrum et al. (2017), Shukla (2021) and Li et al. (2022) found many comments of this nature as well. In addition to appreciative and acknowledging comments towards the content creator for the video, some users express appreciation and/or acknowledgement towards other users who describe their experiences with obesity and weight loss, which goes in the direction of emotional support in the comments described by Frohlich and Zmyslinski-Seelig (2012). Notably, about one in four comments (27.63%) contains personal experiences and/or information. Similarly, Meldrum et al. (2017), Lewis et al. (2012), Shukla (2021), Li et al. (2022) and Bakombo et al. (2023) found that users often use the comments to share their own experiences or those of people close to them. As per Meldrum et al. (2017) and Bakombo et al. (2023), this behavior can be appealing in terms of a sense of community to share experiences with others who have similar health problems to feel less alone, gain hope and pass on coping strategies.

The overall proportion of comments containing questions is relatively low (9.25%) compared to Meldrum et al. (2017) and Li et al. (2022). Although around one in four questions (27.03%) is video-related, the proportion of these in relation to all comments is only 2.50% and only two questions in total are comprehension questions regarding the content of the video. Questions about the personal experiences of other people account for almost half of all questions (47.30%). While some users ask for situation-specific advice,

the proportion of such requests in the main category *Question* is at 14.86% much lower than in Meldrum et al. (2017), where it is around 50%. Meldrum et al. (2017) conclude that the health information on knee pain provided in the videos is probably not specific enough for the users, as many have questions about the applicability of the video content to their specific knee joint condition. As this problem does not exist for the videos analyzed in this paper, the topic difference probably plays a crucial role.

In addition, a few comments (5.63%), contain information about a personal action the users intend to do or have already taken. According to Meldrum et al. (2017), this can be seen as an indicator that a video left a lasting impression on users and influenced their health behavior. Ultimately, from the users' point of view, the videos still seem to need improvement in some aspects, but they appear to be quite helpful overall.

5.3 Implications

Since, as per Meldrum et al. (2017), users hardly have any means of verifying the credibility of the health information in the videos, the comment section offers them the opportunity to discuss the information and thus in some way confirm the usefulness of the videos. Following the analysis of this study, this view can only be shared to a limited extent: It is true that the users point out shortcomings of the videos in the comments. However, the seven videos uploaded by non-professionals, which mostly received only mediocre PRHISM scores, nevertheless received a lot of positive feedback, and mostly no negative feedback at all. One exception is the video uploaded by a commercial company, which was also rated poorly with the PRHISM tool. However, as stated in Section 1, the users' perspective on HIQ is subjective and the aspects relevant to the users in this regard can therefore differ from the evaluation criteria of the PRHISM tool. Furthermore, the understandability of most videos made by laypersons is high, which presumably influences the perception of the videos by the users. Tao et al. (2017), Al-Jefri et al. (2018) and Sun et al. (2019) state that for laypersons understandability plays an important role regarding the perceived HIQ.

According to Osman et al. (2022), it is advisable for users to generally exercise caution when searching for health information on YouTube and to always consult their physician before making health-related decisions. However, Mohamed and Shoufan (2024) found that users do make health decisions based on health videos on YouTube, including whether to see a physi-

cian at all. Instruments for evaluating the HIQ such as DISCERN, JAMA or the PRHISM tool could potentially be used by laypersons (Bernstam et al., 2005; Eysenbach, 2005; Denniss et al., 2022) but are generally not practical to use in everyday life due to their complexity (Bernstam et al., 2005). Additionally, due to the way YouTube algorithms work, it cannot be guaranteed that videos with high HIQ originating from health organizations, medical institutions and health professionals will appear at the top of the search results and video recommendations (Osman et al., 2022). Hence, Osman et al. (2022) suggest that an efficient way to direct users to high-quality health videos could be to identify and recommend high-quality channels instead of individual videos.

This proposal basically corresponds to the concept of YouTube Health mentioned in Section 1, consisting of the Health Label and the Health Shelf. In this study, the Health Shelf was only displayed for one search query (see Fig. 3). As per Google (n.d.), however, this function is not yet reliably available for all health conditions and is to be expanded in the future. Yet, the videos from channels with the YouTube Health label exhibit a significantly higher HIQ based on the PRHISM score in this study. Despite criticism of the reliability of the verification process (see Focken, 2023), the YouTube Health label appears to fulfil its purpose, at least at first glance, as likewise indicated by a study by Lindenmayr and Foerderer (2023). Ultimately, YouTube Health is still a relatively new function, and it therefore remains to be seen to what extent it can fulfil the expectations placed on it in terms of high-quality health information. However, YouTube Health and a new framework on medical misinformation (Graham & Halprin, 2023) indicate that YouTube has seemingly recognized that the quality of the health information disseminated on its platform is part of its responsibility as the platform operator.

From the users' perspective, the YouTube Health label therefore appears to be a practical indicator of information quality. In case a channel is not verified with the YouTube Health label, users could still check whether the video in question is from a channel operated by a health organization, medical institution or a qualified health professional. Additionally, it seems to be worthwhile for users to consider whether the speaker communicating the health information in a video is a qualified health professional. Furthermore, channels run by public broadcasters appear to be a safe bet, particularly if the speaker is a qualified health professional. Nevertheless, users should be cautious, since it cannot be ruled out with absolute certainty that there are not

also health professionals disseminating misinformation (e.g., Sule et al., 2023).

5.4 Limitations and future research

The results of this study are subject to some limitations: First, this is a cross-sectional study. However, the search results and video recommendations on YouTube as well as the comments on the videos are constantly changing, sometimes within days or hours (Sampson et al., 2013; Sui et al., 2022). Additionally, the search results and video recommendations are determined by algorithms that are only known in detail to YouTube itself (Arthurs et al., 2018; Kirdemir et al., 2021; Sui et al., 2022), meaning that it is ultimately almost impossible to replicate previous searches, as Sui et al. (2022) conclude.

Furthermore, some restrictions were made when selecting the videos and comments for the analysis (see Section 4), which might have influenced the results to some degree. These restrictions are arbitrary in nature and are not based on specific previous research. Besides this, the opinions of viewers without an account and/or users who have not commented could lead to different results regarding the perception of the videos.

Another important limitation is that the rating and coding carried out during the analysis always remain subjective to some degree, despite the detailed user guides provided for the PRHISM tool and PEMAT-A/V as well as the self-created coding guide. Moreover, only one person, the first author of this paper, coded the videos and comments, and not several people who could have discussed their results and/or determined the inter-rater or intercoder reliability. In addition, the evaluator of the videos does not have a medical education as mentioned in Section 3.

Finally, it should be emphasized that the results are not necessarily transferable to videos in other languages, videos on other topics or on other social media platforms. In this respect, a research approach could be to conduct a study on health information on another social media platform, such as Facebook, Instagram or TikTok, and to compare the results with those of this study. Another option would be to conduct a comparative study of health information on social media and on websites. Beyond this, the effectiveness of the YouTube Health program could be investigated in greater depth, whereby the perspective of the users could be included with user studies. Another research idea is to investigate users' motivations for watching health

videos on YouTube and writing different types of comments as well as the factors that lead users to make health decisions based on these videos.

References

- Ab Hamid, M. R., Mohd Yusof, N. D. B., & Buhari, S. S. (2022). Understandability, actionability and suitability of educational videos on dietary management for hypertension. *Health Education Journal*, 81(2), 238–247. <https://doi.org/10.1177/00178969211072275>.
- Afful-Dadzie, E., Afful-Dadzie, A., & Egala, S. B. (2023). Social media in health communication: A literature review of information quality. *Health Information Management Journal*, 52(1), 3–17. <https://doi.org/10.1177/1833358321992683>.
- Al-Jefri, M., Evans, R., Uchyigit, G., & Ghezzi, P. (2018). What Is Health Information Quality? Ethical Dimension and Perception by Users. *Frontiers in Medicine*, 5, 260. <https://doi.org/10.3389/fmed.2018.00260>.
- Arthurs, J., Drakopoulou, S., & Gandini, A. (2018). Researching YouTube. *Convergence: The International Journal of Research into New Media Technologies*, 24(1), 3–15. <https://doi.org/10.1177/1354856517737222>.
- Azer, S. A. (2020). Are DISCERN and JAMA Suitable Instruments for Assessing YouTube Videos on Thyroid Cancer? Methodological Concerns. *Journal of Cancer Education*, 35(6), 1267–1277. <https://doi.org/10.1007/s13187-020-01763-9>.
- Bakombo, S., Ewalefo, P., & Konkole, A. T. M. (2023). The Influence of Social Media on the Perception of Autism Spectrum Disorders: Content Analysis of Public Discourse on YouTube Videos. *International Journal of Environmental Research and Public Health*, 20(4), 3246. <https://doi.org/10.3390/ijerph20043246>.
- Basch, C. H., Hillyer, G. C., MacDonald, Z. L., Reeves, R., & Basch, C. E. (2015). Characteristics of YouTube™ Videos Related to Mammography. *Journal of Cancer Education*, 30(4), 699–703. <https://doi.org/10.1007/s13187-014-0769-9>.
- Basch, C. H., Fung, I. C.-H., Menafro, A., Mo, C., & Yin, J. (2017). An exploratory assessment of weight loss videos on YouTube™. *Public Health*, 151, 31–38. <https://doi.org/10.1016/j.puhe.2017.06.016>.
- Batar, N., Kermen, S., Sevdin, S., Yıldız, N., & Güçlü, D. (2020). Assessment of the Quality and Reliability of Information on Nutrition After Bariatric Surgery on YouTube. *Obesity Surgery*, 30(12), 4905–4910. <https://doi.org/10.1007/s11695-020-05015-z>.

- Batar, N., Aşkin, G., İpek, E. B., Sevdin, S., & Özçalkap, R. (2021). Assessment of the Quality and Reliability of Information on Nutrition for Patients with Diabetes on YouTube. *European Journal of Science and Technology*, 31, 268–274. <https://doi.org/10.31590/ejosat.992770>.
- Baumann, E., Czerwinski, F., Großmann, U., & Calhoun, K. (2021). Ist Gesundheit schon digitaler Alltag? Teilergebnisse der Studie „HINTS Germany“ zur Nutzung digitaler Gesundheitsangebote. *SGW-Trendmonitor*, 04/2021, 1–6. https://www.stiftung-gesundheitswissen.de/sites/default/files/pdf/2021_07_08_trendmonitor_Digitalisierung_AktFass_vf_o.pdf.
- Beaunoyer, E., Arsenault, M., Lomanowska, A. M., & Guitton, M. J. (2017). Understanding online health information: Evaluation, tools, and strategies. *Patient Education and Counseling*, 100(2), 183–189. <https://doi.org/10.1016/j.pec.2016.08.028>.
- Benajiba, N., Alhomidi, M., Alsunaid, F., Alabdulkarim, A., Dodge, E., Chavarria, E. A., & Aboul-Enein, B. H. (2023). Video clips of the Mediterranean Diet on YouTube™: A social Media Content Analysis. *American Journal of Health Promotion*, 37(3), 366–374. <https://doi.org/10.1177/08901171221132113>.
- Bernard, A., Langille, M., Hughes, S., Rose, C., Leddin, D., & van Zanten, S. V. (2007). A Systematic Review of Patient Inflammatory Bowel Disease Information Resources on the World Wide Web. *American Journal of Gastroenterology*, 102(9), 2070–2077. <https://doi.org/10.1111/j.1572-0241.2007.01325.x>.
- Bernstam, E. V., Shelton, D. M., Walji, M., & Meric-Bernstam, F. (2005). Instruments to assess the quality of health information on the World Wide Web: What can our patients actually use? *International Journal of Medical Informatics*, 74(1), 13–19. <https://doi.org/10.1016/j.ijmedinf.2004.10.001>.
- Boté, J.-J. (2019). Lack of standards in evaluating YouTube health videos. *Revista Cubana de Información en Ciencias de la Salud*, 30(2), e1357. <https://acimed.sld.cu/index.php/acimed/article/view/1357>.
- Boyer, C., Selby, M., Scherrer, J.-R., & Appel, R. D. (1998). The Health On the Net Code of Conduct for medical and health Websites. *Computers in Biology and Medicine*, 28(5), 603–610. [https://doi.org/10.1016/S0010-4825\(98\)00037-7](https://doi.org/10.1016/S0010-4825(98)00037-7).
- Calisir, A., & Ece, I. (2022). Assessment of the Quality and Reliability of Intragastric Balloon Videos on YouTube. *Obesity Surgery*, 32(4), 1157–1163. <https://doi.org/10.1007/s11695-022-05911-6>.
- Charnock, D., Shepperd, S., Needham, G., & Gann, R. (1999). DISCERN: An instrument for judging the quality of written consumer health information on treatment choices. *Journal of Epidemiology & Community Health*, 53(2), 105–111. <https://doi.org/10.1136/jech.53.2.105>.

- Denniss, E., Lindberg, R., & McNaughton, S. A. (2022). Development of Principles for Health-Related Information on Social Media: Delphi Study. *Journal of Medical Internet Research*, 24(9), e37337. <https://doi.org/10.2196/37337>.
- Denniss, E., Lindberg, R., Marchese, L. E., & McNaughton, S. A. (2024). #Fail: The quality and accuracy of nutrition-related information by influential Australian Instagram accounts. *International Journal of Behavioral Nutrition and Physical Activity*, 21, 16. <https://doi.org/10.1186/s12966-024-01565-y>.
- Diviani, N., van den Putte, B., Giani, S., & van Weert, J. C. (2015). Low Health Literacy and Evaluation of Online Health Information: A Systematic Review of the Literature. *Journal of Medical Internet Research*, 17(5), e4018. <https://doi.org/10.2196/jmir.4018>.
- Döring, N. (2023). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften* (6th ed.). Berlin: Springer. <https://doi.org/10.1007/978-3-662-64762-2>.
- Drozdz, B., Couvillon, E., & Suarez, A. (2018). Medical YouTube Videos and Methods of Evaluation: Literature Review. *JMIR Medical Education*, 4(1), e8527. <https://doi.org/10.2196/mededu.8527>.
- Dubovi, I., & Tabak, I. (2020). An empirical analysis of knowledge co-construction in YouTube comments. *Computers & Education*, 156, 103939. <https://doi.org/10.1016/j.compedu.2020.103939>.
- Ellis, C. H., Moore, J. B., Ho, P., & Evans, C. E. (2023). Development and validation of a quality assessment tool to assess online nutrition information. *Digital Health*, 9, 1–13. <https://doi.org/10.1177/20552076231187249>.
- Eysenbach, G. (2005). Design and Evaluation of Consumer Health Information Web Sites. In D. Lewis, G. Eysenbach, R. Kukafka, P. Z. Stavri, & H. B. Jimison (Eds.), *Consumer Health Informatics: Informing Consumers and Improving Health Care* (pp. 34–60). New York: Springer. https://doi.org/10.1007/0-387-27652-1_4.
- Focken, C. (2023). YouTube Health: Ein Label für seriöse Gesundheitsinfos? Warum YouTube nicht hält, was es verspricht. *MedWatch*. <https://medwatch.de/weitere-artikel/youtube-health/>.
- Frohlich, D. O., & Zmyslinski-Seelig, A. (2012). The Presence of Social Support Messages on YouTube Videos About Inflammatory Bowel Disease and Ostomies. *Health Communication*, 27(5), 421–428. <https://doi.org/10.1080/10410236.2011.606524>.
- Gimenez-Perez, G., Robert-Vila, N., Tomé-Guerreiro, M., Castells, I., & Mauricio, D. (2020). Are YouTube videos useful for patient self-education in type 2 diabetes? *Health Informatics Journal*, 26(1), 45–55. <https://doi.org/10.1177/1460458218813632>.

- Godwin, H. T., Khan, M., & Yellowlees, P. (2017). The Educational Potential of YouTube. *Academic Psychiatry*, 41(6), 823–827. <https://doi.org/10.1007/s40596-017-0809-y>.
- Google. (n.d.). Get health info on YouTube. Get info on health-related content. *YouTube Help*. <https://support.google.com/youtube/answer/9795167?hl=en>.
- Graham, G. (2021). Introducing new ways to help you find answers to your health questions. *YouTube Official Blog*. <https://blog.youtube/news-and-events/introducing-new-ways-help-you-find-answers-your-health-questions/>.
- Graham, G. (2022). YouTube Health: Zugang zu zuverlässigen Gesundheitsinformationen. *YouTube Official Blog*. <https://blog.youtube/intl/de-de/news-and-events/youtube-health-zugang-zu-zuverlassigen-gesundheitsinformationen/>.
- Graham, G., & Halprin, M. (2023). A long term vision for YouTube’s medical misinformation policies. *YouTube Official Blog*. <https://blog.youtube/inside-youtube/a-long-term-vision-for-medical-misinformation-policies/>.
- Guler, M. A., & Aydın, E. O. (2022). Development and validation of a tool for evaluating YouTube-based medical videos. *Irish Journal of Medical Science*, 191(5), 1985–1990. <https://doi.org/10.1007/s11845-021-02864-0>.
- Hahn, A. (2023). Adipositas und metabolisches Syndrom. In A. Hahn, A. Ströhle, & M. Wolters (Eds.), *Ernährung. Physiologische Grundlagen, Prävention, Therapie* (4th ed., pp. 785–831). Stuttgart: WVG.
- Høj, S., Meteran, H., Thomsen, S. F., Sigsgaard, T., & Meteran, H. (2023). Nutritional treatment of atopic diseases according to YouTube videos. *The Journal of Allergy and Clinical Immunology: In Practice*, 11(5), 1552–1553. <https://doi.org/10.1016/j.jaip.2023.01.055>.
- Hurrelmann, K., Klinger, J., & Schaeffer, D. (2020). *Gesundheitskompetenz der Bevölkerung in Deutschland: Vergleich der Erhebungen 2014 und 2020*. Bielefeld: Interdisziplinäres Zentrum für Gesundheitskompetenzforschung, Universität Bielefeld. <https://doi.org/10.4119/unibi/2950303>.
- Inan-Eroglu, E., & Buyuktuncer, Z. (2022). YouTube: Is It a Reliable Source of Nutrition Information on COVID-19 Pandemic? *Healthcare*, 10(10), 1911. <https://doi.org/10.3390/healthcare10101911>.
- Kasper, H., & Burghardt, W. (2021). *Ernährungsmedizin und Diätetik* (13th ed.). München: Urban & Fischer.
- Kharbat, F. F., & Abu Daabes, A. (2023). Assessing Arabic youtube videos on herbal cancer treatment: Absence of health information quality. *Health Informatics Journal*, 29(3), 1–17. <https://doi.org/10.1177/14604582231198022>.
- Kirdemir, B., Kready, J., Mead, E., Hussain, M. N., Agarwal, N., & Adjero, D. (2021). Assessing Bias in YouTube’s Video Recommendation Algorithm in a

- Cross-lingual and Cross-topical Context. In R. Thomson, M. N. Hussain, C. Dancy, & A. Pyke (Eds.), *Social, Cultural, and Behavioral Modeling. 14th International Conference, SBP-BRIMS 2021 Virtual Event, July 6–9, 2021, Proceedings* (pp. 71–80). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-80387-2_7.
- Kiss, A., Soós, S., Temesi, Á., Unger-Plasek, B., Lakner, Z., & Tompa, O. (2023). Evaluation of the reliability and educational quality of YouTube™ videos on sport nutrition topics. *Journal of the International Society of Sports Nutrition*, 20(1), 2278632. <https://doi.org/10.1080/15502783.2023.2278632>.
- Kriz, M., Möseneder, J. M., & Leitner, G. (2019). The QWEB tool: Evaluation sheet for the quality of nutrition articles on the World Wide Web. Using the example of articles on the effect of coconut oil for weight loss. *Ernährungs Umschau*, 66(4), 68–74. <https://doi.org/10.4455/eu.2019.014>.
- Lamb, K. L., Barker, M. E., & Lynn, A. (2023). A content analysis of online videos containing dietary recommendations for gout and their alignment with evidence-based dietary guidelines. *Public Health Nutrition*, 26(10), 2014–2025. <https://doi.org/10.1017/S136898002300160X>.
- Lambert, K., Mullan, J., Mansfield, K., Koukoulos, A., & Mesiti, L. (2017). Evaluation of the quality and health literacy demand of online renal diet information. *Journal of Human Nutrition and Dietetics*, 30(5), 634–645. <https://doi.org/10.1111/jhn.12466>.
- Lambert, K., Miglioretto, C., & Javadpour, A. (2023). Update on the Quality and Health Literacy Demand of Diet-Related Videos on YouTube for People With Polycystic Kidney Disease. *Journal of Renal Nutrition*, 33(3), 495–498. <https://doi.org/10.1053/j.jrn.2022.10.006>.
- Landis, J. R., & Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, 33(1), 159–174. <https://doi.org/10.2307/2529310>.
- Lautenbach, A., & Aberle, J. (2022). Adipositas und Ernährung. *Die Diabetologie*, 18(5), 542–548. <https://doi.org/10.1007/s11428-022-00890-9>.
- Lee, J. S., Seo, H. S., & Hong, T. H. (2014). YouTube as a source of patient information on gallstone disease. *World Journal of Gastroenterology*, 20(14), 4066–4070. <https://doi.org/10.3748/wjg.v20.i14.4066>.
- Lewis, S. P., Heath, N. L., Sornberger, M. J., & Arbuthnott, A. E. (2012). Helpful or Harmful? An Examination of Viewers' Responses to Nonsuicidal Self-Injury Videos on YouTube. *Journal of Adolescent Health*, 51(4), 380–385. <https://doi.org/10.1016/j.jadohealth.2012.01.013>.
- Li, Y., Ashley, D. L., & Popova, L. (2022). Users' Modifications to Electronic Nicotine Delivery Systems: Content Analysis of YouTube Video Comments. *JMIR Infodemiology*, 2(2), e38268. <https://doi.org/10.2196/38268>.

- Lindenmayr, M., & Foerderer, J. (2023). Is Labeling of Authorized Sources Effective? Evidence from YouTube Health. *DIGIT 2023 Proceedings. Diffusion Interest Group In Information Technology Annual Workshop, Hyderabad, India, December 10, 2023* (Paper 12). Atlanta: Association for Information Systems. <https://aisel.aisnet.org/digit2023/12>.
- Liu, J., Zhang, Y., & Kim, Y. (2023). Consumer Health Information Quality, Credibility, and Trust: An Analysis of Definitions, Measures, and Conceptual Dimensions. In J. Gwizdka & S. Y. Rieh (Eds.), *CHIIR '23: Proceedings of the 2023 Conference on Human Information Interaction and Retrieval. ACM SIGIR Conference on Human Information Interaction and Retrieval, Austin, TX, USA, March 19–23, 2023* (pp. 197–210). New York: Association for Computing Machinery. <https://doi.org/10.1145/3576840.3578331>.
- Loeb, S., Sengupta, S., Butaney, M., Macaluso, J. N., Czarniecki, S. W., Robbins, R., Braithwaite, R. S., Gao, L., Byrne, N., Walter, D., & Langford, A. (2019). Dissemination of Misinformative and Biased Information about Prostate Cancer on YouTube. *European Urology*, 75(4), 564–567. <https://doi.org/10.1016/j.eururo.2018.10.056>.
- Loeb, S., Reines, K., Abu-Salha, Y., French, W., Butaney, M., Macaluso, J. N., Steinberg, G. D., Walter, D., Byrne, N., de la Garza, D., & Smith, A. B. (2021). Quality of Bladder Cancer Information on YouTube. *European Urology*, 79(1), 56–59. <https://doi.org/10.1016/j.eururo.2020.09.014>.
- Madathil, K. C., Rivera-Rodriguez, A. J., Greenstein, J. S., & Gramopadhye, A. K. (2015). Healthcare information on YouTube: A systematic review. *Health Informatics Journal*, 21(3), 173–194. <https://doi.org/10.1177/1460458213512220>.
- Madden, A., Ruthven, I., & McMenemy, D. (2013). A classification scheme for content analyses of YouTube video comments. *Journal of Documentation*, 69(5), 693–714. <https://doi.org/10.1108/JD-06-2012-0078>.
- Marcon, A. R., & Caulfield, T. (2017). Commenting on chiropractic: A YouTube analysis. *Cogent Medicine*, 4(1), 1277450. <https://doi.org/10.1080/2331205X.2016.1277450>.
- Mayring, P. (2022). *Qualitative Inhaltsanalyse. Grundlagen und Techniken* (13th ed.). Weinheim: Beltz.
- Mayring, P., & Fenzl, T. (2022). Qualitative Inhaltsanalyse. In N. Baur & J. Blasius (Eds.), *Handbuch Methoden der empirischen Sozialforschung* (3rd ed., pp. 691–706). Wiesbaden: Springer Fachmedien. https://doi.org/10.1007/978-3-658-37985-8_43.
- Meldrum, S., Savarimuthu, B. T., Licorish, S., Tahir, A., Bosu, M., & Jayakaran, P. (2017). Is knee pain information on YouTube videos perceived to be helpful? An

- analysis of user comments and implications for dissemination on social media. *Digital Health*, 3, 1–18. <https://doi.org/10.1177/2055207617698908>.
- Mohamed, F., & Shoufan, A. (2024). Users' experience with health-related content on YouTube: An exploratory study. *BMC Public Health*, 24, 86. <https://doi.org/10.1186/s12889-023-17585-5>.
- Müller, M. J., Gaetjens, I., & Bösby-Westphal, A. (2022). Prävention der Adipositas. In S. Herpertz, M. de Zwaan, & S. Zipfel (Eds.), *Handbuch Essstörungen und Adipositas* (3rd ed., pp. 533–542). Berlin: Springer. https://doi.org/10.1007/978-3-662-63544-5_68.
- Non-Communicable Diseases Risk Factor Collaboration (NCD-RisC). (2024). Worldwide trends in underweight and obesity from 1990 to 2022: A pooled analysis of 3663 population-representative studies with 222 million children, adolescents, and adults. *The Lancet*, 403(10431), 1027–1050. [https://doi.org/10.1016/S0140-6736\(23\)02750-2](https://doi.org/10.1016/S0140-6736(23)02750-2).
- Nimptsch, K., & Pischon, T. (2022). Epidemiologie der Adipositas. In S. Herpertz, M. de Zwaan, & S. Zipfel (Eds.), *Handbuch Essstörungen und Adipositas* (3rd ed., pp. 445–451). Berlin: Springer. https://doi.org/10.1007/978-3-662-63544-5_56.
- Oberänder, N., & Weimann, A. (2022). Bewegung und Ernährung bei multimodalen Adipositasprogrammen – wie unterscheiden sich die angebotenen Programme und welche Ergebnisse sind zu erwarten? *Aktuelle Ernährungsmedizin*, 47(3), 183–187. <https://doi.org/10.1055/a-1714-7978>.
- Osman, W., Mohamed, F., Elhassan, M., & Shoufan, A. (2022). Is YouTube a reliable source of health-related information? A systematic review. *BMC Medical Education*, 22, 382. <https://doi.org/10.1186/s12909-022-03446-z>.
- Rana, A., & Arora, M. (2023). Ketogenic diet: Assessing YouTube video information using quality, reliability, and text analytics methods. *Nutrition and Health* (OnlineFirst). <https://doi.org/10.1177/02601060231193789>.
- Robillard, J. M., Jun, J. H., Lai, J.-A., & Feng, T. L. (2018). The QUEST for quality online health information: Validation of a short quantitative tool. *BMC Medical Informatics and Decision Making*, 18, 87. <https://doi.org/10.1186/s12911-018-0668-9>.
- Şahin, E., & Bayır Garbioğlu, D. (2023). Scientific Quality and Reliability Analysis of Turkish YouTube Videos on Cancer and Nutrition. *Eurasian Journal of Medical Advances*, 3(2), 79–84. <https://doi.org/10.14744/ejma.2023.70298>.
- Sampson, M., Cumber, J., Li, C., Pound, C. M., Fuller, A., & Harrison, D. (2013). A systematic review of methods for studying consumer health YouTube videos, with implications for systematic reviews. *PeerJ*, 1, e147. <https://doi.org/10.7717/peerj.147>.

- Sbaffi, L., & Rowley, J. (2017). Trust and Credibility in Web-Based Health Information: A Review and Agenda for Future Research. *Journal of Medical Internet Research*, 19(6), e7579. <https://doi.org/10.2196/jmir.7579>.
- Schaeffer, D., Berens, E.-M., Gille, S., Griesse, L., Klinger, J., de Sombre, S., Vogt, D., & Hurrelmann, K. (2021). *Gesundheitskompetenz der Bevölkerung in Deutschland vor und während der Corona Pandemie: Ergebnisse des HLS-GER 2*. Bielefeld: Interdisziplinäres Zentrum für Gesundheitskompetenzforschung, Universität Bielefeld. <https://doi.org/10.4119/unibi/2950305>.
- Scheufele, B., & Engelmann, I. (2009). *Empirische Kommunikationsforschung*. Konstanz: UVK.
- Schienkiewitz, A., Kuhnert, R., Blume, M., & Mensink, G. B. M. (2022). Übergewicht und Adipositas bei Erwachsenen in Deutschland – Ergebnisse der Studie GEDA 2019/2020-EHIS. *Journal of Health Monitoring*, 7(3), 23–31. <https://doi.org/10.25646/10292>.
- Schneider, S., & Holzwarth, B. (2022). Ansätze zur Beseitigung adipogener Umwelten. In S. Herpertz, M. de Zwaan, & S. Zipfel (Eds.), *Handbuch Essstörungen und Adipositas* (3rd ed., pp. 587–593). Berlin: Springer. https://doi.org/10.1007/978-3-662-63544-5_74.
- Segado-Fernández, S., Herrera-Peco, I., Jiménez-Gómez, B., Ruiz Núñez, C., Jiménez-Hidalgo, P. J., Benítez de Gracia, E., González-Rodríguez, L. G., Torres-Ramírez, C., & Lozano-Estevan, M. del C. (2023). Realfood and Cancer: Analysis of the Reliability and Quality of YouTube Content. *International Journal of Environmental Research and Public Health*, 20(6), 5046. <https://doi.org/10.3390/ijerph20065046>.
- Shoemaker, S. J., Wolf, M. S., & Brach, C. (2013). *The Patient Education Materials Assessment Tool (PEMAT) and User's Guide. An Instrument to Assess the Understandability and Actionability of Print and Audiovisual Education Materials (Version 1.0)* (No. 14-0002-EF). Rockville: Agency for Healthcare Research and Quality. https://www.ahrq.gov/sites/default/files/publications2/files/pemat_guide_o.pdf.
- Shoemaker, S. J., Wolf, M. S., & Brach, C. (2014). Development of the Patient Education Materials Assessment Tool (PEMAT): A new measure of understandability and actionability for print and audiovisual patient information. *Patient Education and Counseling*, 96(3), 395–403. <https://doi.org/10.1016/j.pec.2014.05.027>.
- Shukla, A. (2021). COVID-19 pandemic: An analysis of popular YouTube videos as an alternative health information platform. *Health Informatics Journal*, 27(2), 1–13. <https://doi.org/10.1177/1460458221994878>.
- Silberg, W. M., Lundberg, G. D., & Musacchio, R. A. (1997). Assessing, Controlling, and Assuring the Quality of Medical Information on the Internet: Caveant Lector

- et Viewer – Let the Reader and Viewer Beware. *Journal of the American Medical Association*, 277(15), 1244–1245. <https://doi.org/10.1001/jama.1997.03540390074039>.
- Singh, A. G., Singh, S., & Singh, P. P. (2012). YouTube for Information on Rheumatoid Arthritis—A Wakeup Call? *The Journal of Rheumatology*, 39(5), 899–903. <https://doi.org/10.3899/jrheum.111114>.
- Stvilia, B., Mon, L., & Yi, Y. J. (2009). A model for online consumer health information quality. *Journal of the American Society for Information Science and Technology*, 60(9), 1781–1791. <https://doi.org/10.1002/asi.21115>.
- Sui, W., Sui, A., & Rhodes, R. E. (2022). What to watch: Practical considerations and strategies for using YouTube for research. *Digital Health*, 8, 1–13. <https://doi.org/10.1177/20552076221123707>.
- Sule, S., DaCosta, M. C., DeCou, E., Gilson, C., Wallace, K., & Goff, S. L. (2023). Communication of COVID-19 Misinformation on Social Media by Physicians in the US. *JAMA Network Open*, 6(8), e2328928. <https://doi.org/10.1001/jama-networkopen.2023.28928>.
- Sun, Y., Zhang, Y., Gwizdka, J., & Trace, C. B. (2019). Consumer Evaluation of the Quality of Online Health Information: Systematic Literature Review of Relevant Criteria and Indicators. *Journal of Medical Internet Research*, 21(5), e12522. <https://doi.org/10.2196/12522>.
- Sütçüoğlu, O., Özay, Z. İ., Özet, A., Yazıcı, O., & Özdemir, N. (2023). Evaluation of scientific reliability and quality of YouTube videos on cancer and nutrition. *Nutrition*, 108, 111933. <https://doi.org/10.1016/j.nut.2022.111933>.
- Tao, D., LeRouge, C., Smith, K. J., & Leo, G. D. (2017). Defining Information Quality Into Health Websites: A Conceptual Framework of Health Website Information Quality for Educated Young Adults. *JMIR Human Factors*, 4(4), e6455. <https://doi.org/10.2196/humanfactors.6455>.
- Vishnevetsky, J., Walters, C. B., & Tan, K. S. (2018). Interrater reliability of the Patient Education Materials Assessment Tool (PEMAT). *Patient Education and Counseling*, 101(3), 490–496. <https://doi.org/10.1016/j.pec.2017.09.003>.
- World Health Organization (WHO). (2024). One in eight people are now living with obesity. <https://www.who.int/news/item/01-03-2024-one-in-eight-people-are-now-living-with-obesity>.
- YouTube. (n.d.a). Our global impact. *YouTube Health*. <https://health.youtube/our-impact/>.
- YouTube. (n.d.b). How We Provide Credible Online Health Information. *YouTube Health*. <https://health.youtube/how-we-help/>.

- Zhang, Y., & Kim, Y. (2022). Consumers' Evaluation of Web-Based Health Information Quality: Meta-analysis. *Journal of Medical Internet Research*, 24(4), e36463. <https://doi.org/10.2196/36463>.
- Zhang, Y., Sun, Y., & Xie, B. (2015). Quality of health information for consumers on the web: A systematic review of indicators, criteria, tools, and evaluation results. *Journal of the Association for Information Science and Technology*, 66(10), 2071–2084. <https://doi.org/10.1002/asi.23311>.