Online content on eating disorders: a natural language processing study

Livio Tarchi, Tommaso Mario Buonocore, Giulia Selvi, Valdo Ricca & Giovanni Castellini

---

📄 View supplementary material 🗗

📅 Published online: 23 Jul 2024.

✎ Submit your article to this journal 🗗

🔍 View related articles 🗗

CrossMark View Crossmark data 🗗

Taylor & Francis
Taylor & Francis Group

Check for updates

Online content on eating disorders: a natural language processing study

Livio Tarchi [a], Tommaso Mario Buonocore [b], Giulia Selvi [a], Valdo Ricca [a] and Giovanni Castellini [a]

aPsychiatry Unit, Department of Health Sciences, University of Florence, Florence, Italy; bDepartment of Electrical, Computer and Biomedical Engineering, University of Pavia, Pavia, Italy

ABSTRACT

Background: Online content can inform the personal risk of developing an eating disorder, and it can influence the time and motivation to seek treatment. Patients routinely seek information online, and access to information is crucial for both prevention and treatment. The primary aim of the current study was to quantify the readability scores of online content on eating disorders using natural language processing algorithms, across two languages: English and Italian.

Methods: Unique terms related to single diagnoses were searched using Google®. The content available on Wikipedia was also assessed. Readability was defined according to the Flesch Readability Ease (FRE) and the Rate Readability Index (RIX). The scientific support of retrieved content and the authoritativeness of sources were measured through standardized variables.

Results: In Italian, online content was more likely published by private psychotherapy institutes or by websites that promote diet-advice or weight-loss. In both languages, the most readable content was on Anorexia Nervosa (RIX 4.18, FRE-en 59.6, FRE-it 41.69), Bulimia Nervosa (RIX 3.99, FRE-en 66.27, FRE-it 39.66) or Binge Eating (RIX 4.01, FRE-en 68.10, FRE-it 38.62). English sources consistently had more references than Italian pages (range 35–182, vs 1–163, respectively). and had a higher percentage of citations available in the target language. The content of these references was mainly reflective of peer-reviewed or clinical manuals.

Conclusion: Attention should be given to developing online content for Muscle Dysmorphia and Orthorexia Nervosa, as well as improving the overall readability of online content on eating disorders, especially for languages other than English.

1. Introduction

Eating Disorders (EDs) are defined as persistent alterations in eating behaviors, that result in an altered food consumption which significantly affects the general functioning of an individual [1]. The prevalence of EDs is increasing over time at the global level, partly due to changes in diagnostic criteria [2], but also due to socio-cultural and geographic factors contributing to an individual's lifetime risk and severity [3,4]. Among socio-cultural factors, a prominent role has been observed for media representations of thin bodies [5–7], as well as 'pro-anorexia' or 'pro-eating disorders' websites. This socio-cultural influence seems particularly impactful on adolescent women [8], and can be conceptualized as two-fold [9]. First, as preventive and informative on EDs, describing associated risk factors and clinical presentations. Second, as potentially detrimental to mental health, promoting pathological eating behaviors [10].

In recent decades, the search for health information changed from a qualified and physical domain (e.g. the general practitioner's office, the local pharmacist, a physician of choice) to online content. However, which online sources are considered reliable, authoritative, or accessible by an individual depends on several factors, among which: the relative access to healthcare services, the level of education, age, or income, and the individual degree of competence in the digital domain [11–13]. An important factor in the search for online information has been the recent decrease in social media engagement, not widely deemed as a reliable source of health information [14], mainly due to concerns about information quality and authoritativeness. Users commonly use search engines to seek health information, rarely going beyond the first page of their results, avoiding overt commercialism, but not paying attention to the credibility of online sources, nor noting their potential conflicts of interest [15].

Online information-seeking can change healthcare interventions. The digital space can ultimately respond to individual requests differently than the analogic one. For instance, it can address the desire for further information on home remedies [16]. The

digital space can also offer information on first-line diagnostics widely sold over-the-counter or point toward accessible resources offered by private enterprises [17]. Finally, it can integrate with the traditional health domain, recommending a specific doctor aligned with the patient's requests, influenced by the previous experiences of other patients or by the offer of a specific expertise (such as the offer of advanced techniques or diagnostics to which the patient is interested).

Health information-seeking increases the perceived self-efficacy of an individual, as most users look for information on specific conditions when they or someone they know is diagnosed by a health professional [15,18]. This observation is of primary importance for the field of EDs, as there could be an incongruence between the perception of how weight is assessed in the clinical setting and how, on the other hand, patients report their needs should be addressed [19]. Additionally, while internet sources can be an effective tool for communicating nutritional information, a lack of authoritative information on online media risks promoting controversial weight-loss practices, at the expense of evidence-based clinical recommendations [20].

In general, online information on mental health has been characterized by inappropriate readability across the globe [21,22], even though institutional and governmental agencies have suggested readability as a key goal for evaluating health content published on the internet [23,24]. Contemporary efforts have therefore focused on measuring and assessing the quality of online information for EDs, first as a means of prevention, and second to address the potential neglected needs of patients, with the objective of improving patients' insight and quality of life through guided self-education [25,26]. However, most efforts evaluated online information according to its content or formality, not by a user-center perspective of appropriateness and accessibility [27]. The lack of assessment for the readability of online information on health seems problematic when considering that around 35% of citizens in the United States [28] and more than 50% of citizens in Italy [27] exhibit basic or below basic health literacy competencies (that is, the capacity to search, understand, and consume information in order to develop an educated decision for the self).

In light of these challenges, contemporary research converged on measuring the ease of access to online information, as a crucial component for information-seeking behaviors and as a better representation of user-center demands [23]. These attempts aimed to quantify the readability of online content, here defined as the ease by which a written passage can be understood, due to the style of writing [29]. For instance, Arts and colleagues [10] have examined the readability of online information for Anorexia Nervosa in the English language. The resulting body of content was found to be too complex for general purposes, and the quality of information not satisfactory across all sources (mean information quality 'fair' or lower; [10]). In other languages, Italian for example, up to 36% of online content related to Anorexia Nervosa was found to be actually devoted to pro-anorexia websites [30], with an overall poor quality of language.

Beyond issues of readability, conflicts of interest seem to be particularly present in online media covering information on EDs. In fact, Guardiola-Wanden-Berghe and colleagues [31] systematically reviewed websites related to diet and/or offering information on Eds. This research found that no site satisfied quality criteria such as stating authorship, affiliation, and endorsement by professional or academic institutions. Moreover, no explicit acknowledgement of economic conflict of interest was offered [31]. Although most of the websites described symptoms or treatment, only about 10% of them were found to fully describe diagnostic criteria based on the Diagnostic and Statistical Manual of Mental Disorders. Of particular interest, treatment options were seldom adequately reported [31]. Therefore, a review that includes sources in other languages than English is warranted, answering the call for the exploration of cross-linguistic comparisons in readability scores for online mental health information [32], in order to assess the generalizability of previous results. In the current work, Italian sources were compared to online information on eating disorders available in English. This choice was motivated by potential differences in cultural norms on eating behaviors across Italy and the Anglosphere [33,34], as well as divergent health literacy competencies [27,28].

An objective assessment of the scientific support cited in online content is needed, considering that online information that describes treatments or symptoms can inform both patients and the general public about the time to seek treatment [18,35,36]. As previously mentioned, online content on EDs has been described as suffering from a low engagement by professionals in their writing [31]. Moreover, online content has been found to be characterized by inaccuracies in the presentation of symptoms or treatment options for EDs, while also being at potential risk for conflicts of interest [31]. For these reasons, online content on EDs should be systematically evaluated and reviewed in light of potential conflicts of interest, such as those arising from the sponsorship of online content by commercial entities. Finally, since Muscle Dysmorphia Disorder and Orthorexia Nervosa have gained prominence among EDs [37,38], while potentially being less commonly known among the general public, online content about these disorders should be formally addressed.

1.1. Aims

The primary aim of the present work was to quantify the readability of online content on EDs categories: Eating Disorders in general, Anorexia Nervosa, Bulimia Nervosa, Binge Eating Disorder, Muscle Dysmorphia, Orthorexia Nervosa. The secondary aim was to describe eventual differences between the retrieved content based on the language of content (Italian or English), as well as to systematically review the retrieved content in respect to the scientific support of its cited references, the authoritativeness of its sources, and potential conflicts of interest arising from sponsorship biases.

2. Materials & methods

2.1. Sample description and pre-processing

The following diagnoses were included: Eating Disorders, Anorexia Nervosa, Bulimia Nervosa, Binge Eating Disorder, Muscle Dysmorphia, Orthorexia Nervosa. Each term was searched through a private search using Google® search, executed in December 2021. The private search was executed by obfuscating the country of origin, internet provider, previous internet cache or cookies, and using a public DNS. The first 10 results of each search were saved. This selection of material was chosen as the most likely to reflect those encountered by the typical user [15,39–41]. Duplicates (among the same term) and websites that were portals to content from other sites were excluded. If Wikipedia.org was not among the first 10 results, the specific Wikipedia page for the term was searched, and the landing page included. This procedure was implemented to reflect page traffic not directly coming from search engines, but specific and targeted by users exploring health information on the internet [42,43].

Only the main text of the landing pages was saved, social links, indices, 'read more' were excluded. References were saved in a separate corpus for each landing page in order to evaluate the quality of information of online content. The written corpus of each landing page was pre-processed in order to remove upper case, punctuation, control characters, URLs, line breaks, e-mails, phone numbers, numbers, digits, currency symbols [44]. The overall written corpus was divided into words and sentences [45]. All analyses were conducted in Python 3.8.13, with the support of the following libraries: *pandas* [46]; *numpy* [47]; *scipy* [48]; *ntlk* [45]; *pyphen* [49]. The quality of information for online content was operatively defined by two indices: (i) its readability, given by natural-language-processing algorithms; (ii) the scientific support of its cited references, the authoritativeness of its sources, and potential conflicts of interest arising from sponsorship biases

(see section 2.3 for how this information was assessed).

2.2. Readability

The readability of content was defined by three scores: (i) syllables per word; (ii) rate readability index; (iii) Flesch reading ease score. The results were averaged by diagnosis. Means for the top 10 results by Google® search were computed.

2.2.1. Rate readability index (RIX)

The Rate Readability Index (RIX) is given by a formula which is reliable and easy to interpret [50]. It is derived from the Läsbarhet index ('readability index' in Swedish) [51]. This first formula, later referred to as LIX (Läsbarhet index), was later adapted to English by Anderson and renamed RIX [49]. RIX estimates are easily available by counting the number of long words per sentence in a corpus of text, with 'long' generally defined by 7 characters or more [50]. It is almost equivalent to LIX and can be scaled to reflect the grade level needed for the comprehension of content (grade levels as defined by the US education system; for instance, grade 8 for 13–14 years old readers, last year of middle school). See the equation for computing RIX and Table 1 for further information on the interpretation of single scores.

$$RIX = \frac{longwords(\geq 7characters)}{sentences}$$

2.2.2. Flesch Reading ease (FRE)

The Flesch Readability Ease (FRE) is a common index that aims to evaluate the difficulty of understanding a written text in the English language. The Flesch reading ease test is given by the difference between two proportions, the first between the number of words over the number of sentences, and the second

Table 1. Readability scores, interpretation.

	English			Italian
RIX index	FRE	Grade Level	FRE	Readability
/	10–0	Professional	<0	Extremely difficult
/	30–10	College graduate	0–30	Difficult
≥7.2	50–30	College	30–40	Fairly difficult
≥6.2	60–50	12	40–50	Standard
≥5.3	60–50	11	50–60	Fairly easy
≥4.5	60–50	10	60–70	Easy
≥3.7	70–60	9	80–90	Extremely easy
≥3.0	70–60	8	/	/
≥2.4	80–70	7	/	/
≥1.8	90–80	6	/	/
≥1.3	100–90	5	/	/
≥0.8	/	4	/	/
≥0.5	/	3	/	/
≥0.2	/	2	/	/
<0.2	/	1	/	/

Note: No study at present time offers a direct interpretation of RIX scores for Italian.

by the number of syllables over words.

$$FRE^{english} = 206.835 - 1.015 * \left(\frac{words}{sentences}\right) - 84.6 * \left(\frac{syllables}{words}\right)$$

As the original equation pertains only to English, an Italian adaptation was developed by Franchina and Vacca in 1986. The Flesch-Vacca formula, as it is now known, was later modified in 1986 to be more comparable to English [52]. For the current work, this later adaptation was chosen over the original one described in 1972, in order to improve the comparability between sources. See Table 1 for further information on conversion to US Grade levels and for the interpretation of each score.

$$FRE^{italian} = 217 - 0.6 * \left(\frac{words}{sentences}\right) - 1.3 * \left(\frac{syllables}{words}\right)$$

2.3. Scientific support, authoritativeness, and risk of sponsorship bias

The scientific support of cited sources was assessed by manually inspecting all references on each Wikipedia landing page. Each reference was assigned to one of the following categories, by order of authority: peer-reviewed academic publications; scientific or clinical information (e.g. in the form of manuals, textbooks, doctoral theses); information promoted by health agencies, governmental sources, or non-governmental organizations; non-scientific publications (e.g. magazines, anecdotal, interviews).

References were checked for their accessibility, and content which was archived, or which presented non-accessible information at the time of assessment was annotated. Wikipedia was chosen in order to reflect page traffic not directly coming from search engines, but specific and targeted by users exploring health information on the internet [42,43]. The authoritativeness of each website was annotated for what pertains to the following information: type of entity (for-profit or not); area of focus (clinical/medical, generalist, psychotherapy). The potential risk for sponsorship bias was then evaluated by senior psychiatrists, experts in the field of eating disorders (G.C., V.R.), and the results of this assessment were presented narratively in the Discussion section of the present work.

3. Results

3.1. Sample

A total of 120 landing pages were extracted. For both English and Italian, 60 unique sites were identified, as per Methods previously illustrated. Among the most frequent websites (occurring more than once), eight websites out of 11 in English were either institutional, representing a healthcare center of excellence, or a non-governmental agency (total occurrences: 24; 40.00% of the total). For Italian, the same was true only for 3 websites and 10 total occurrences (16.66% of the total). In particular, two for-profit news websites were among the frequent landing pages for English (9 total occurrences, 15%), while for Italian 10 unique sites were found as pertaining to for-profit institutions (31 occurrences, 51.66%). Similarly, no private psychotherapy institute was among the most frequent landing pages in English, while it represented six unique websites for Italian (19 total occurrences, 31.66%). See Table 2 for a description of the most frequent sites across terms.

3.2. Readability

On average, both English and Italian sources had a moderately long text on the landing page for the top Google® search results (range 1,048–2,796 for English and 1,001–2,720 for Italian). The longest article was

Table 2. Website frequency.

English			Italian		
Rank	Website	Number of occurrences	Rank	Website	Number of occurrences
1	https://www.healthline.com	6*	1	https://www.my-personaltrainer.it	6*
2	https://en.wikipedia.org	6	2	https://www.stateofmind.it	5*§
3	https://www.nationaleatingdisorders.org	5	3	https://www.auxologico.it	4
4	https://www.mayoclinic.org	4	4	https://www.ipsico.it	4*§
5	https://www.beateatingdisorders.org.uk	3	5	https://www.ospedalemarialuigia.it	4
6	https://www.nhs.uk	3	6	https://www.apc.it	3*§
7	https://www.webmd.com	3*	7	https://www.eist.it	3*§
8	https://www.eatingdisorders.org.au	2	8	https://it.wikipedia.org	3
9	https://www.hopkinsmedicine.org	2	9	https://www.intherapy.it	2*§
10	https://nedc.com.au	2	10	https://www.issalute.it	2
11	https://www.verywellhealth.com	2	11	https://www.melarossa.it	2*
12	/	/	12	https://www.msdmanuals.com/it-it	2*
13	/	/	13	http://www.psicoterapia-bologna.org	2*§
14	/	/	14	https://www.treccani.it	2*

Note: Only websites occurring more than once reported.
Overall occurrences for frequent websites over total: 38/60–63.33% for English, 44/60–73.33% for Italian.
* For-profit
§ Private psychotherapy institute

about Muscle Dysmorphia in English and about Anorexia Nervosa in Italian. In English, the most readable content was on Bulimia Nervosa and Binge Eating Disorder (RIX 3.99/4.01 – interpretable as readable at an education level comparable to US Grade 9; FRE 66.27/68.1 respectively – similarly indicating a readability level of US Grade 8-9), while for Italian it was on Anorexia Nervosa (FRE 41.69 – interpretable as of standard difficulty for its readability). The least readable content was on Muscle Dysmorphia (RIX 7.00, FRE 49.16 – readable at US Grade Level 12/College) and Orthorexia Nervosa (RIX 6.13, FRE 53.08 – readable at US Grade Level 11/12) for English, while in Italian the least readable content was on EDs in general (FRE 32.98 – fairly difficult to read), Muscle Dysmorphia (FRE 33.80 – fairly difficult to read) and Orthorexia Nervosa (FRE 33.28 – fairly difficult to read). In Italian, this content all had comparably lower readability in comparison to more commonly known diagnoses, namely Anorexia Nervosa (FRE 41.69 – standard difficulty), Bulimia Nervosa (FRE 39.66 – fairly difficult to read) and Binge Eating Disorder (FRE 38.62 – fairly difficult to read). Results were averaged by term, and results were reported in Table 3. See Supplementary Materials eTable 1a and 1b for results pertaining to each landing page.

For English content on Wikipedia, the longest article was on EDs in general, with a total of 12,472 words (words per sentence 20.86; syllables per word 1.62) and the shortest on Orthorexia Nervosa (1,924 words; words per sentence 27.88; syllables per word 1.66). The shortest article – Orthorexia Nervosa – was also the most difficult to read (RIX 9.94; FRE 38.1; both indicating a College education level). The easiest article to read was on Bulimia Nervosa (RIX 5.96; FRE 58; both indicating a US grade level 11). For Italian, the longest article was on Anorexia Nervosa (6,368 words; 298 sentences; words per sentence 21.37; syllables per word 2.26). The shortest, similarly to English, was on Orthorexia Nervosa (203 words). As the content was too short, no readability information

was derived from the specific article. The most readable article in Italian was on Bulimia Nervosa (FRE 53.6 – fairly easy to read). Results for each Wikipedia page can be found in the Supplementary Materials as eTable 2.

3.3. Cited sources

To reflect page traffic not directly coming from search engines, but specific and targeted by users exploring health information on the internet [42,43], the references of each Wikipedia landing page were analyzed and described. Overall, English sources consistently had more references than Italian pages (range 35–182, vs 1–163 respectively; Chi-square 166.524, $p < 0.001$). and had a higher percentage of citations available in the target language (Chi-square 31.258, $p < 0.001$). In fact, at most one reference for each landing page was not in the target language for English landing pages, with the sole exception of EDs (eight references out of 332 not in English, from scientific peer-reviewed or governmental sources). For Italian, three diagnoses out of six had fewer than ten references each (Binge Eating, Muscle Dysmorphia and Orthorexia Nervosa). The content of these references was mainly reflective of peer-reviewed or clinical manuals. For both Italian and English content, a relatively high number of references cited archived content that was not available at the time of assessment. Further results can be found in Table 4.

4. Discussion

The present work revises and expands previous results on the quality of online information for EDs [8,10,13,30,31,53,54]. In fact, while most of the previous evidence concerned content available in English, Italian sources were found to be similarly readable, citing similar references but shorter on average. Italian content seems to offer a synthetic translation of English sources, rather than novel information.

Table 3. Readability, average across 10 top Google® search results.

Diagnosis	Words	Sentences	Words per sentence	Syllable per sentence	RIX	FRE	Readability
English Sources							
Anorexia Nervosa	2120	143	13.36	1.58	4.18	59.6	Grade 9–12
Binge Eating Disorder	1283	75	16.16	1.45	4.01	68.10	Grade 9
Bulimia Nervosa	1567	116	12.86	1.51	3.99	66.27	Grade 9
Eating Disorders	1578	102	15.26	1.56	4.74	59.48	Grade 10–12
Muscle Dysmorphia	2796	128	21.05	1.61	7.00	49.16	Grade 12/College
Orthorexia Nervosa	1048	53	19.29	1.59	6.13	53.08	Grade 11–12
Italian Sources							
Anorexia Nervosa	2720	112	25.64	2.30	10.04	41.69	Standard
Binge Eating Disorder	1001	41	27.06	2.32	10.62	38.62	Fairly Difficult
Bulimia Nervosa	1741	71	25.08	2.34	10.02	39.66	Fairly Difficult
Eating Disorders	1796	66	31.41	2.32	12.23	32.98	Fairly Difficult
Muscle Dysmorphia	1200	43	30.60	2.32	11.67	33.80	Fairly Difficult
Orthorexia Nervosa	1060	46	32.23	2.30	12.02	33.28	Fairly Difficult

RIX = rate readability index.
FRE = Flesch readability ease.

Table 4. Reliability.

Diagnosis	Number of references	Scientific, peer-reviewed sources	Clinical manuals	Governmental or NGO	Others	Archived sources	Available in target language
English Sources							
Anorexia Nervosa	182	125 (68.68%)	39 (21.43%)	7 (3.85%)	11 (6.04%)	17 (9.34%)	182 (100%)
Binge Eating Disorder	70	47 (67.14%)	9 (12.86%)	5 (7.14%)	9 (12.86%)	0 (0%)	69 (98.57%)
Bulimia Nervosa	107	77 (71.96%)	18 (16.82%)	0 (0%)	12 (11.21%)	11 (10.28%)	107 (100%)
Eating Disorders	332	293 (88.25%)	18 (5.42%)	2 (0.60%)	19 (5.72%)	11 (3.31%)	324 (97.59%)
Muscle Dysmorphia	41	35 (85.37%)	5 (12.20%)	0 (0%)	1 (2.44%)	0 (0%)	40 (97.56%)
Orthorexia Nervosa	35	13 (37.14%)	5 (14.29%)	1 (2.86%)	16 (45.71%)	6 (17.14%)	34 (97.14%)
Italian Sources							
Anorexia Nervosa	163	125 (76.69%)	34 (20.86%)	1 (0.61%)	3 (1.84%)	6 (3.68%)	19 (11.66%)
Binge Eating Disorder	5	2 (40.00%)	1 (20.00%)	0 (0%)	2 (40.00%)	0 (0%)	3 (60.00%)
Bulimia Nervosa	63	26 (41.27%)	21 (33.33%)	3 (4.76%)	13 (20.63%)	12 (19.05%)	6 (9.52%)
Eating Disorders	39	35 (89.75%)	4 (10.25%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Muscle Dysmorphia	1	0 (0%)	1 (100%)	0 (0%)	0 (0%)	0 (0%)	1 (100%)
Orthorexia Nervosa	5	2 (40.00%)	1 (20.00%)	0 (0%)	2 (40.00%)	1 (20.00%)	1 (20.00%)

Note: Archived sources were available through 3rd party archiving websites.

Although previous academic work has suggested that translating English sources might represent a viable method to prevent the exacerbation of health disparities between global communities [32], current authors also suggest adapting the content to the society and culture of destination. Adapting the content to the society and culture of destination, health communication strategies could better address divergent risk factors in light of the target audience [55].

In current results, online content about EDs was found to mainly focus on Anorexia Nervosa, Bulimia Nervosa or Binge Eating Disorder. The readability of online content on EDs ranged from difficult to standard (from College level to Grade 9). These results confirm what has been previously described for online content on mental health in general [32]. In fact, online content on mental health has been routinely found to be written well above the 6th to 8th grade level suggested by the National Institutes of Health in the United States [56,57]. Additionally, less commonly known, but increasingly more prevalent disorders, such as Muscle Dysmorphia and Orthorexia Nervosa were found to be less readable across both languages, and less content was offered for both diagnoses compared to Anorexia Nervosa, Bulimia Nervosa or Binge Eating Disorder. In the following paragraphs, implications for public policy and health communication are reviewed, as informed by the comparison between the retrieved content across English and Italian websites.

4.1. Authoritative sources

Institutional websites were more prevalent within English sources. In other words, online content about EDs in English was mostly reflective of institutional or governmental publications. On the other hand, a higher number of unique websites was found for Italian, and they were mainly reflective of for-profit entities, as well as private psychotherapy institutions. Only two total occurrences were found for governmental sources

among the most frequently searched websites in Italian, both represented by the institutional website of the National Institute of Health (for Anorexia Nervosa and Bulimia Nervosa, https://www.issalute.it). A special case was noted for the Italian Society of Eating Psychopathology (SIPA, https://psicopatologiaalimentazione.it), which was among the top 10 results only for EDs in general. This particular result reflects an academic effort to provide high-quality online information, but the website currently does not seem to be represented among the top searches for other particular diagnoses.

For what concerns unique websites among the top Google® rankings, only two instances were managed by for-profit entities in English results, and both were represented by health-oriented news agencies (https://www.healthline.com and https://www.webmd.com). Interestingly, these news agencies currently offer a platform to connect directly with mental health professionals or to schedule an appointment with a specific medical doctor (whether it is a general practitioner or a specialist), signaling a potential conflict of interest over the presented evidence in favor of offered services.

For Italian, health-oriented news agencies were also represented by two websites (https://www.my-personaltrainer.it and https://www.melarossa.it). Both targeted weight – and shape-conscious people at the time of consultation. The first started as a website promoting fitness trainers, and later expanded to publish health-oriented content for specific pathologies [58]. At the time of writing, the website was currently ranked as the 952nd most visited website on the web, worldwide, with an organic traffic of 13.56M visits per month [59]. The second was still currently devoted to sponsoring diet-advice, while also offering a podcast on nutrition and weight-loss [60]. This finding strengthens the need to assess online content for its role in promoting weight-loss, moderating access to healthcare for EDs, especially for what concerns content other than in the English language.

Governmental sources have previously been found to be more readable and reliable for what concerns

online mental health information [23]. The lack of trusted sources in Italian, that is, governmental or institutional sources, thus warrants the attention of both clinicians and researchers on devoting their efforts to public divulgation. Poor readability of online information on mental health can negatively impact access to care, promoting misinformation. For these reasons, previous work suggested further efforts to improve the readability and authoritativeness of online mental health information [21]. However, especially for content other than English, a scarce allocation of resources by public or private funds could reflect the lower possibility or incentive to pursue such endeavors [61].

4.2. Conflicts of interest and sponsorship bias

Although the problem of significant conflicts of interests in scholarly work favoring psychotherapy interventions was previously noted [62–64], this risk seems to be more prevalent than previously reported for online information on EDs [31]. Current results suggest that conflicts of interest in online content mirror what has already been known for clinical trials and society recommendations. In particular, allegiance effects were never disclosed in the retrieved online content [62,63], and potential sponsorship bias was rarely addressed [64]. Allegiance effects characterize the positive confirmation bias of those psychotherapists who promote an intervention in line with their school of thought or training, for example, by promoting a psychodynamic or cognitive–behavioral perspective and increasing their claims of efficacy [62]. On the contrary, sponsorship bias reflects the issue of commercial interests in the promotion of particular interventions [64]. Online information on EDs was found to be at a high risk for both biases, in the Italian language, given the high prevalence of for-profit and private psychotherapy institution among the sponsors of the retrieved content.

Online content can inform the choice of a particular therapist, or dictate the preference over a specific technique. Online content can also guide patients' choice, and it can signal high-risk situations where a preliminary evaluation would be warranted in light of personal symptom severity [65,66]. Therefore, online content is of primary importance for public health policy and practice. Decades-long calls to fund high-quality online content from institutional sources seem to have influenced the quality of mental health information for English sources [15,24]. However, the same has not been found for Italian. Therefore, the authors highlight the need for specific grants reserved to promote authoritative content online, both as a means to promote prevention and as a method to sponsor the activity of a neutral agent capable of offering authoritative information to patients.

4.3. Accessibility and novelty

While an issue of accessibility was noted across both languages, due to articles citing archived content (up to 19.05% of citations), in Italian the cited content was also rarely available in the target language (e.g. EDs with 0 references available in Italian). This result may further indicate how Italian content is often a synthetic translation of English sources, which does not present novel information, nor provides socio-cultural adjustments to their discourse. As Wikipedia is a primary source for online healthcare information, and as it captures a high proportion of online traffic, a higher availability of Italian references would be vital for informing the general public about the scientific consensus on the topic.

For what concerns health-behaviors in general, Wikipedia currently grants first access to authoritative content in languages other than English, allowing for the scientific literature on a single topic to be relatively readable and authoritative for the general public, with a lower probability of conflicts of interests in promoting diagnostic procedures or treatments. Moreover, Wikipedia currently amplifies the impact of peer-reviewed articles, diffusing publicly funded content, but available only by subscription to scientific journals [67]. For these reasons, community-driven content might be supported by the active participation of experts in the field, and online information on eating behaviors revisited and expanded according to current practices. The authors also suggest promoting content produced by national clinical or scientific associations, as well as redirecting to institutional websites on the respective landing pages, to improve the readability and authoritativeness of Italian sources on EDs.

4.4. The gender-bias of online content on EDs

Less commonly known and more recently established diagnoses – such as Muscle Dysmorphia and Orthorexia Nervosa – had less readable content compared to Anorexia Nervosa, Bulimia Nervosa or Binge Eating Disorder across both languages. Along with a low accessibility, a marked difference was related to the authoritativeness of the content on these two diagnoses. Orthorexia Nervosa and Muscle Dysmorphia had a markedly lower number of references compared to other diagnoses (35 and 41, respectively, for English content – 5 and 1 for Italian). Orthorexia Nervosa, specifically, also had the highest prevalence of citation to non-scientific or anecdotal references (45.71% of references for English, 40.00% for Italian). This finding is of primary importance considering how under-diagnosis has been a concern for both entities [68], while actually impacting approximately 7% of the general population, and up to 57.8% of individuals in high-

risk groups [69,70]. Additionally, while Anorexia Nervosa, Bulimia Nervosa and Binge Eating Disorders show a notable gender-prevalence for women, both Orthorexia Nervosa and Muscle Dysmorphia may be either equivalent or more represented among men [69,71]. Therefore, the low accessibility of online content for these diagnoses could reinforce the stigma against diagnosing and treating eating disorders in men [72,73], reinforcing the lower likelihood of men to seek treatment for this category of mental disorders [73].

5. Conclusions

Online content on EDs was found to primarily focus on Anorexia Nervosa, Bulimia Nervosa or Binge Eating Disorder. Its readability ranged from difficult to standard (from College to grade 9 level). For content other than English, Italian sources were evaluated and the same trend was confirmed. The readability of articles was reflective of the same bias, with more commonly known diagnoses being written in a more readable fashion. Shorter texts and fewer references were more likely to appear on the landing page of these diagnoses. Most references pertained to sources in English and scientific articles, regardless of the target language of the article. In agreement with what was previously found within scholarly works, sponsorship bias and allegiance effects were seldom addressed when content offered advice on psychotherapy interventions. Divergent trends were noted across languages. For Italian, online content was more likely published by private psychotherapy institutes or entities promoting weight-loss/diet advice. As online content informs the time to seek treatment, or the choice of a particular center of care, creating, updating, and maintaining online content by third-party, neutral, or institutional sources should be promoted by public and private funding agencies. Finally, more efforts should be devoted to divulging information on Muscle Dysmorphia and Orthorexia Nervosa, especially for languages other than English.

Ethics

This study evaluates publicly available content. In particular, it assesses the readability of online content (written text) using natural language processing algorithms. As the current study did not involve human or animal subjects, ethical approval by the local ethics committee was not required.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Authors' contribution

Study conception and design: L.T.; Data collection: L.T., G.S.; Analysis and interpretation of results: L.T., T.M.B.; Draft manuscript preparation: L.T.; supervision: V.R., G.C.; Manuscript editing and revision: V.R., G.C. All authors reviewed the results and approved the final version of the manuscript.

Data availability

The data and code used to analyze the data are freely available at the following site: https://github.com/detsutut/infoquality-dca. The authors guarantee a minimum of 5 years of maintenance to the repository.

Notes on contributors

Livio Tarchi is a medical resident at the Psychiatry Unit at the University of Florence. His clinical interests focus on gender identities and eating behaviors while using neuroimaging techniques to account for a distributed and embodied perspective of cognition in the brain.

Tommaso Mario Buonocore is a biomedical engineer, currently researching solutions based on natural-language processing to improve healthcare-based prediction tasks leveraging Italian clinical notes. He was awarded his PhD at the University of Pavia.

Giulia Selvi is a medical resident at the University of Florence. Her clinical interests focus on gender identities and eating behaviors.

Prof. Valdo Ricca is a full professor of Psychiatry at the University of Florence, Italy. Prof. Ricca has an established curriculum in the treatment of eating disorders, and he is a leading expert in Psychiatry Genetics.

Prof. Giovanni Castellini is an associate professor of Psychiatry at the University of Florence, Italy. He is specialized in typical and atypical sexuality, as well as in the psychopathology of eating disorders.

ORCID

Livio Tarchi http://orcid.org/0000-0002-9931-5621
Tommaso Mario Buonocore http://orcid.org/0000-0002-2887-088X
Giulia Selvi http://orcid.org/0000-0002-9613-0904
Valdo Ricca http://orcid.org/0000-0002-9291-2124
Giovanni Castellini http://orcid.org/0000-0003-1265-491X

References

[1] American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 5th ed.; 2013. https://doi.org/10.1176/appi.books.9780890425596.

[2] Castelpietra G, Knudsen AKS, Agardh EE, et al. The burden of mental disorders, substance use disorders and self-harm among young people in Europe, 1990-2019: findings from the global burden of disease study 2019. Lancet Reg Health Eur. 2022;16:100341.

[3] Musolino CM, Warin M, Gilchrist P. Embodiment as a paradigm for understanding and treating SE-AN: locating the self in culture. Front Psychiatry. 2020: 11. [cited 2022 February 17]. https://www.frontiersin.org/article/10.3389fpsyt.2020.00534.

[4] Rymarczyk K. The role of personality traits, sociocultural factors, and body dissatisfaction in anorexia readiness syndrome in women. J Eat Disord. 2021;9(1):51.

[5] Andersen TH. Speaking about the unspeakable: sexually abused men striving toward language. Am J Mens Health. 2008;2(1):25–36.

[6] Becker AE. Television, disordered eating, and young women in Fiji: negotiating body image and identity during rapid social change. Cult Med Psychiatry. 2004;28(4):533–59.

[7] Chen A, Couturier J. Triggers for children and adolescents with anorexia nervosa: a retrospective chart review. J Can Acad Child Adolesc Psychiatry. 2019;28(3):134–40.

[8] Mento C, Silvestri MC, Muscatello MRA, et al. Psychological impact of pro-anorexia and pro-eating disorder websites on adolescent females: a systematic review. Int J Environ Res Public Health. 2021;18(4):2186.

[9] Bullivant B, Rhydderch S, Griffiths S, Mitchison D, Mond JM. Eating disorders "mental health literacy": a scoping review. J Ment Health. 2020;29(3):336–49. doi:10.1080/09638237.2020.1713996.

[10] Arts H, Lemetyinen H, Edge D. Readability and quality of online eating disorder information—are they sufficient? A systematic review evaluating websites on anorexia nervosa using DISCERN and Flesch readability. Int J Eat Disord. 2020;53(1):128–32.

[11] Cotten SR, Gupta SS. Characteristics of online and offline health information seekers and factors that discriminate between them. Soc Sci Med. 2004;59(9):1795–806. doi:10.1016/j.socscimed.2004.02.020.

[12] Diviani N, van den Putte B, Giani S, van Weert JC. Low health literacy and evaluation of online health information: a systematic review of the literature. J Med Internet Res. 2015;17(5):e4018.

[13] Dutta-Bergman M. Trusted online sources of health information: differences in demographics, health beliefs, and health-information orientation. J Med Internet Res. 2003;5(3):e893.

[14] Zhao Y, Zhang J. Consumer health information seeking in social media: a literature review. Health Inf Libr J. 2017;34(4):268–83.

[15] Morahan-Martin JM. How internet users find, evaluate, and use online health information: a cross-cultural review. Cyberpsychol Behav. 2004;7(5):497–510. doi:10.1089/cpb.2004.7.497.

[16] Wathen CN, Harris RM. I try to take care of It myself." How rural women search for health information. Qual Health Res. 2007;17(5):639–51.

[17] Ramos E, Weissman SM. The Dawn of consumer-directed testing. Am J Med Genet C Semin Med Genet. 2018;178(1):89–97.

[18] Fox S, Duggan M. Health online 2013. Washington, D.C.: Pew Research Center; 2013.

[19] Laidlaw A, Napier C, Neville F, Collinson A, Cecil JE. Talking about weight talk: primary care practitioner knowledge, attitudes and practice. J Commun Healthc. 2019;12(3-4):145–53. doi:10.1080/17538068.2019.1646061.

[20] Chan T, Drake T, Vollmer RL. A qualitative research study comparing nutrition advice communicated by registered dietitian and non-registered dietitian bloggers. J Commun Healthc. 2020;13(1):55–63. doi:10.1080/17538068.2020.1749351.

[21] Daraz L, Morrow AS, Ponce OJ, et al. Readability of online health information: a meta-narrative systematic review. Am J Med Qual. 2018;33(5):487–92.

[22] Cheng H, Wang Y, Sheng J, et al. Characteristics and variability of structural networks derived from diffusion tensor imaging. NeuroImage. 2012;61(4):1153–64. doi:10.1016/j.neuroimage.2012.03.036.

[23] Mcinnes N, Haglund BJA. Readability of online health information: implications for health literacy. Inform Health Soc Care. 2011;36(4):173–89.

[24] Swire-Thompson B, Lazer D. Public health and online misinformation: challenges and recommendations. Annu Rev Public Health. 2020;41:433–51.

[25] Albano G, Hodsoll J, Kan C, Lo Coco G, Cardi V. Task-sharing interventions for patients with anorexia nervosa or their carers: a systematic evaluation of the literature and meta-analysis of outcomes. Int Rev Psychiatry. 2019;31(4):367–81. doi:10.1080/09540261.2019.1588711.

[26] Silverstein LS, Haggerty C, Sams L, Phillips C, Roberts MW. Impact of an oral health education intervention among a group of patients with eating disorders (anorexia nervosa and bulimia nervosa). J Eat Disord. 2019;7(1):29.

[27] Palumbo R, Annarumma C, Adinolfi P, Musella M, Piscopo G. The Italian health literacy project: insights from the assessment of health literacy skills in Italy. Health Policy. 2016;120(9):1087–94. doi:10.1016/j.healthpol.2016.08.007.

[28] Kutner M, Greenburg E, Jin Y, Paulsen C. The health literacy of America's adults: results from the 2003 national assessment of adult literacy. NCES 2006-483. Natl Cent Educ Stat. Published online September 2006. [cited 2023 March 3]. https://eric.ed.gov/?id = ed493284.

[29] Klare GR. Measurement of readability. Iowa State University Press; 1963. [cited 2022 July 29]. https://scholar.google.com/scholar_lookup?title = measurement+of+readability&author = Klare%2C+George+Roger&publication_year = 1963.

[30] Bragazzi NL, Prasso G, Re TS, Zerbetto R, Del Puente G. A reliability and content analysis of Italian language anorexia nervosa-related websites. Risk Manag Healthc Policy. 2019;12:145–51.

[31] Guardiola-Wanden-Berghe R, Gil-Pérez JD, Sanz-Valero J, Wanden-Berghe C. Evaluating the quality of websites relating to diet and eating disorders. Health Inf Libr J. 2011;28(4):294–301. doi:10.1111/j.1471-1842.2011.00961.x.

[32] Skierkowski DD, Florin P, Harlow LL, Machan J, Ye Y. A readability analysis of online mental health resources. Am Psychol. 2019;74(4):474–83.

[33] Schaefer LM, Burke NL, Anderson LM, et al. Comparing internalization of appearance ideals and appearance-related pressures among women from the United States, Italy, England, and Australia. Eat Weight Disord - Stud Anorex Bulim Obes. 2019;24(5):947–51.

[34] Higgs S, Ruddock H. Social influences on eating. In: Meiselman HL, editor. Handbook of eating and drinking. Cham: Springer International Publishing; 2020. p. 277–91.

[35] Fitzsimmons-Craft EE, Krauss MJ, Costello SJ, Floyd GM, Wilfley DE, Cavazos-Rehg PA. Adolescents and young adults engaged with pro-eating disorder social media: eating disorder and comorbid psychopathology, health care utilization, treatment barriers, and opinions on harnessing technology for treatment.

Eat Weight Disord - Stud Anorex Bulim Obes. 2020;25 (6):1681–92.

[36] Kästner D, Weigel A, Buchholz I, Voderholzer U, Löwe B, Gumz A. Facilitators and barriers in anorexia nervosa treatment initiation: a qualitative study on the perspectives of patients, carers and professionals. J Eat Disord. 2021;9(1):28.

[37] Mitchison D, Mond J, Griffiths S, et al. Prevalence of muscle dysmorphia in adolescents: findings from the EveryBODY study. Psychol Med. Published Online March. 2021;16:1–8.

[38] Niedzielski A, Kaźmierczak-Wojtaś N. Prevalence of orthorexia nervosa and its diagnostic tools—a literature review. Int J Environ Res Public Health. 2021;18(10):5488.

[39] Eysenbach G, Köhler C. 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. Br Med J. 2002;324(7337):573–7.

[40] Hariri N. Relevance ranking on google: Are top ranked results really considered more relevant by the users? Online Inf Rev. 2011;35(4):598–610.

[41] Laurent MR, Vickers TJ. Seeking health information online: does Wikipedia matter? J Am Med Inform Assoc. 2009;16(4):471–9.

[42] Smith DA. Situating wikipedia as a health information resource in various contexts: a scoping review. PLoS One. 2020;15(2):e0228786.

[43] Filter J. clean-text. Functions to preprocess and normalize text. Published online February. 2022;2.

[44] Bird S, Klein E, Loper E. Natural language processing with python. 1st ed. Sebastopol (CA): O'Reilly Media; 2009.

[45] Reback J, McKinney W, jbrockmendel, et al. pandas-dev/pandas: Pandas 1.0.3. Published online March 18, 2020.

[46] Harris CR, Millman KJ, van der Walt SJ, et al. Array programming with NumPy. Nature. 2020;585 (7825):357–62.

[47] Virtanen P, Gommers R, Oliphant TE, et al. Scipy 1.0: fundamental algorithms for scientific computing in python. Nat Methods. 2020;17(3):261–72.

[48] Ayoub G. pyphen: Pure Python module to hyphenate text. Published online 2022. [cited 2022 September 19]. https://www.courtbouillon.org/pyphen.

[49] Anderson J. Lix and Rix: variations on a little-known readability index. J Read. 1983;26(6):490–6.

[50] Björnsson CH. Läsbarhet. Liber; [Solna. Seelig]; 1968.

[51] Franchina V, Vacca R. Adaptation of flesh readability index on a bilingual text written by the same author both in Italian and English languages. Linguaggi. 1986;3:47–49.

[52] Lai C, Pellicano GR, Iuliano S, et al. Why people join pro-Ana online communities? A psychological textual analysis of eating disorder blog posts. Comput Hum Behav. 2021;124:106922.

[53] Bardone-Cone AM, Cass KM. What does viewing a pro-anorexia website do? an experimental examination of website exposure and moderating effects. Int J Eat Disord. 2007;40(6):537–48.

[54] Castellini G, Pellegrino A, Tarchi L, et al. Body-Size perception among first-generation Chinese migrants in Italy. Int J Environ Res Public Health. 2022;19(10):6063. https://doi.org/10.3390/ijerph19106063.

[55] National Institutes of Health. Making Health communication programs work.; 2001. https://www.cancer.gov/publications/health-communication/pink-book.pdf.

[56] National Institutes of Health. How to write easy-to-read health materials internet.; 2017. https://www.nlm.nih.gov/medlineplus/etr.html.

[57] My Personal Trainer. [cited 2023 November 4]. https://www.my-personaltrainer.it/chi-siamo-my-personaltrainer.html.

[58] Top Websites Ranking - most visited websites in September 2022. Similarweb. [cited 2022 September 4]. https://www.similarweb.com/top-websites/.

[59] Chi Siamo. Melarossa. [cited 2023 November 4]. https://www.melarossa.it/chi-siamo/.

[60] Tetroe JM, Graham ID, Foy R, et al. Health research funding agencies' support and promotion of knowledge translation: An international study. Milbank Q. 2008;86(1):125–55.

[61] Dragioti E, Dimoliatis I, Fountoulakis KN, Evangelou E. A systematic appraisal of allegiance effect in randomized controlled trials of psychotherapy. Ann Gen Psychiatry. 2015;14(1):25.

[62] Dragioti E, Dimoliatis I, Evangelou E. Disclosure of researcher allegiance in meta-analyses and randomised controlled trials of psychotherapy: a systematic appraisal. BMJ Open. 2015;5(6):e007206.

[63] Cristea IA, Gentili C, Pietrini P, Cuijpers P. Sponsorship bias in the comparative efficacy of psychotherapy and pharmacotherapy for adult depression: meta-analysis. Br J Psychiatry. 2017;210(1):16–23.

[64] Kilbride MK, Joffe S. The New Age of patient autonomy: implications for the patient-physician relationship. JAMA. 2018;320(19):1973–4. doi:10.1001/jama.2018.14382.

[65] Yang H, Du HS, Shang W. Understanding the influence of professional status and service feedback on patients' doctor choice in online healthcare markets. Internet Res. 2020;31(4):1236–61.

[66] Teplitskiy M, Lu G, Duede E. Amplifying the impact of open access: Wikipedia and the diffusion of science. J Assoc Inf Sci Technol. 2017;68(9):2116–27.

[67] Kanayama G, Pope Jr. HG. Gods, men, and muscle dysmorphia. Harv Rev Psychiatry. 2011;19(2):95–98. doi:10.3109/10673229.2011.565250.

[68] Applewhite B, Olivola M, Tweed C, et al. Body dysmorphic disorder, muscle dysmorphia, weight and shape dissatisfaction and the use of appearance-enhancing drugs in the military: a systematic review. BMJ Mil Health. Published online June 8, 2022;170(3):255–66. doi:10.1136/bmjmilitary-2022-002135.

[69] Varga M, Dukay-Szabó S, Túry F, van Furth Eric F. Evidence and gaps in the literature on orthorexia nervosa. Eat Weight Disord - Stud Anorex Bulim Obes. 2013;18(2):103–11.

[70] Nagata JM, Ganson KT, Murray SB. Eating disorders in adolescent boys and young men: an update. Curr Opin Pediatr. 2020;32(4):476–81.

[71] Jones W, Morgan J. Eating disorders in men: a review of the literature. J Public Ment Health. 2010;9(2):23–31. doi:10.5042/jpmh.2010.0326.

[72] Strother E, Lemberg R, Stanford SC, et al. Eating disorders in men: underdiagnosed, undertreated, and misunderstood. In: Leigh Cohn, Lemberg Raymond, editors. Current findings on males with eating disorders. New York: Routledge; 2013. p. 13–22.

[73] Weltzin TE, Weisensel N, Franczyk D, et al. Eating disorders in men: update. J Mens Health Gend. 2005;2 (2):186–93. doi:10.1016/j.jmhg.2005.04.008.