La Medicina del Lavoro

Med Lav 2017; 108, 3: 167-173
DOI: 10.23749/mdl.v108i3.6083

A reliability and readability analysis of silicosis-related Italian websites: implications for occupational health

Guglielmo Dini[1,2], Nicola Luigi Bragazzi[3], Beatrice D'Amico[1,2], Alfredo Montecucco[1,2], Stanley C. Igwe[4], Francesco Brigo[5,6], Alessandra Toletone[1], Paolo Durando[1,2]

[1] Department of Health Sciences, Postgraduate School in Occupational Medicine, University of Genoa, Italy
[2] Occupational Medicine Unit, Ospedale Policlinico San Martino, Genoa, Italy
[3] Department of Health Sciences, Postgraduate School of Public Health, University of Genoa, Italy
[4] Department of Neuro-Psychiatry, Federal Teaching Hospital, Abakaliki-Ebonyi State, Nigeria
[5] Department of Neurology, Franz Tappeiner Hospital, Merano, Italy
[6] Department of Neurosciences, Biomedicine, and Movement Sciences, University of Verona, Italy

Key words: Health on the Net Foundation Code of Conduct standards; internet; readability analysis; reliability analysis; silicosis; occupational health promotion

Parole chiave: Criteri del Codice di Condotta della Fondazione Internet; Internet; analisi di leggibilità; analisi di affidabilità; silicosi; promozione della salute in ambito occupazionale

Summary

Background: *Silicosis represents a "classical" occupational disease characterized by a renewed interest. New risk factors are emerging, such as sandblasting in the jeans industry or hydrofracking, leading to clusters of acute or massive cases.* **Objectives:** *Given that the Internet could represent a worker education and empowerment tool, and considering the increase in popularity of silicosis-related information, we aimed at systematically analyzing the reliability and readability of online silicosis-relevant information.* **Methods:** *The search term "silicosi" was used to query 5 top search engines. The first 3 pages of results were screened using two validated readability tools: namely, the Gulpease and the ReadIt DyLanLab grade level scores.* **Results:** *Seventy sites were analyzed. The Gulpease score differed among the types of websites: academic websites differed from institutional websites, as well as encyclopedia/dictionary pages from institutional sites. The Lexical Model – ReadIt DyLanLab grade level differed among the types of websites. Encyclopedia/dictionary pages differed from academic, commercial, health-related, institutional and news sites. Approximately, half of the websites were intended/designed for workers. Only the Global Model – Read-It DyLanLab grade level differed according to the intended/designed target. Only 1.4% of websites adhered to Health on the Net Foundation Code of Conduct.* **Conclusions:** *Our findings may have important practical implications for occupational physicians and health agencies/authorities. They should make efforts in strengthening their online presence, and producing appropriate material. This could lead to positive outcomes in term of occupational health promotion, potentially enabling workers to increase and to improve their work-related health and its determinants.*

Riassunto

«Analisi di affidabilità e leggibilità dei contenuti dei siti web italiani sulla silicosi: possibili implicazioni per la salute in ambito occupazionale». **Introduzione:** *La Silicosi rappresenta una malattia professionale tradizionale caratterizzata recentemente da un rinnovato interesse. Nuovi fattori di rischio come la sabbiatura di jeans e la frat-*

Pervenuto il 27.12.2016 - Revisione pervenuta il 7.4.2017 - Accettato il 20.4.2017
Corrispondenza: Dr. Guglielmo Dini, Department of Health Sciences, Postgraduate School in Occupational Medicine, University of Genoa, Italy - E-mail: guglielmo.dini@unige.it

turazione idraulica, per l'estrazione del petrolio, stanno emergendo come cause di silicosi. **Obiettivi:** *Poiché internet potrebbe rappresentare uno strumento di educazione e responsabilizzazione del lavoratore e considerando l'aumento di popolarità delle informazioni sulla rete relative alla silicosi, abbiamo voluto valutare in modo sistematico l'affidabilità e la leggibilità del materiale disponibile su internet.* **Metodi:** *Il termine "silicosi" è stato utilizzato per cercare sui 5 motori di ricerca più popolari in Italia. Le prime tre pagine sono state selezionate e analizzate usando due strumenti di leggibilità: l'indice Gulpease e l'indice ReadIt DyLanLab.* **Risultati:** *Sono stati analizzati settanta siti internet. L'indice Gulpease differiva tra i vari tipi di siti: quelli accademici differivano da quelli istituzionali così come dizionari/enciclopedie. L'indice ReadIt DyLanLab-modello lessicale differiva tra i vari tipi di sito. I dizionari/enciclopedie differivano da siti accademici, commerciali, di salute, istituzionali e di notizie. Circa metà dei siti erano intesi/realizzati per i lavoratori. Soltanto l'indice ReadIt DyLanLab-modello globale differiva a seconda del target inteso. Solo l'1,4% dei siti internet aderiva al "Codice di Condotta della Fondazione Internet".* **Conclusioni:** *I nostri risultati possono avere importanti implicazioni pratiche per i medici del lavoro e le agenzie/autorità sanitarie. Dovrebbero rafforzare la loro presenza online e produrre materiale appropriato, ai fini della promozione della salute in ambito occupazionale, rendendo i lavoratori in grado di migliorare la propria salute lavorativa ed i fattori che la determinano.*

Introduction

Nowadays, the Internet serves as a major tool for seeking health-related information, within the framework of the so-called "digital health" or "e-health" (7). A multi-centric study has found that approximately 57% of all Italian adults aged 18-65 years use the Internet looking for health or medical information online (20).

However, online information suffers from a number of limitations that should be properly recognized. It can be incorrect or misleading, biased by industrial and commercial interests, omitting preventive measures, avoidable factor risks or complications, or not based on scientific evidence (13, 16). From a patient/worker perspective, the Internet may be regarded as a tool for self-education and self-empowerment (7). Due to asymmetry in information and education, patients/workers may represent a group of users particularly vulnerable to false or biased information (3). Further, because of the gap in health literacy, even if highly accurate and reliable, websites may be difficult to access and to read for lay people (10, 11).

In the current manuscript, we aimed at reviewing and analyzing the online available information on silicosis because it represents a "classical" occupational disease, being an untreatable but preventable fibrotic lung condition that can progressively evolve to lung impairment, respiratory failure and death, even after exposure to crystalline silica has ceased. Further, silicosis represents a disease characterized by a renewed interest and a recent growing popularity. New risk factors, such as hydrofracking or construction and installation of engineered stone kitchens and bathrooms or sandblasting in the jeans industry, are emerging as new causes of silicosis, leading to epidemic clusters of acute, massive cases (1, 2, 4, 12). Campaigns, such as the "Clean Clothes Campaign", and advocacy have raised awareness. Silicosis-related web activities have, indeed, increased in the USA and worldwide (8, 9).

Given that the Internet represents an important worker education tool, and considering the increase in interest and popularity of silicosis-related information, we aimed at systematically describing and analyzing the reliability and readability of online silicosis-relevant information, in the field of "occupational e-health".

Methods

Website pages selection

The search term "*silicosi*" (Italian for silicosis) was used as a keyword to separately query Google®, Yahoo®, Bing®, Ask® and Libero Arianna®. We limited our search to these search engines, in that they currently represent the top 5 search engines used by Italian people to surf the Internet.

Further, we limited our screening to the first 3 pages of results (for a total of 30 entries per search engine) because researchers have found that approximately 90% of users do not go beyond the first 3 pages (18).

The Website pages from the 5 search engines were downloaded in August 2016 exactly in the order they appeared, turning off the location option in order to avoid any bias in the search results. In other words, we did our best efforts in order to avoid that the search engine(s) would make a guess about our location (that is to say, Genoa, Italy) on the basis of our Internet Protocol (IP) address, in such a way that the findings would be representative of all the Italian territory and not only of our local reality. The data were downloaded on the same day to ensure that identical results were independently screened by the two reviewers (NLB and GD).

Inclusion/Exclusion criteria

Website pages were excluded from the current study if they were duplicates, were unrelated or not pertinent to the topic of silicosis, or did not contain enough information for a full review and analysis. YouTube videos were excluded as they should have been transcribed for the readability analysis. Also, presentations for university lectures or congresses/conferences were not retained in the analysis, in that they should have been substantially modified and edited for the readability analysis. Scientific articles or academic materials were not excluded from the current analysis (differently from (3)), since they may represent a source of information for workers.

For further details, the reader is referred to table 1.

Data assessment

The Websites were independently characterized by type (namely, academic, commercial, encyclopedia/dictionary, health, institutional, news, and others) and by target (namely, physician, worker, or both). For further details, the reader is referred to table 1.

Reliability assessment

The two independent reviewers also assessed whether the included website pages adhered to the Health on the Net Foundation Code of Conduct (HONCode) standards, in other words, whether the websites exhibited the logo of the HONCode certification, which is issued only when a given website meets with the required standards of health information reliability (6).

Table 1 - Eligibility criteria for silicosis-related website selection.

Search strategy	Details
Used keyword(s)	Silicosi
Mined search engines	Google®, Bing®, Yahoo®, Ask®, Libero Arianna®
Type of website	Academic (content realized by scientific societies, scholars or researchers), commercial (profit organizations, industries, firms, insurance societies), encyclopedia or dictionary, health, institutional (content realized by local health unities, occupational agencies or authorities, Health Minister), news, other (website pages not falling into the previous categories)
Intended/designed target of website	Physician, worker, both
Language filter	Italian
Location filter	None applied
Time filter	None applied
Exclusion criteria	Website pages containing presentations, videos or other multi-medial contents whose readability index is difficult to assess; website pages unrelated or non pertinent with silicosis; duplicated website pages

Readability assessment

In order to assess the level of ease/difficulty with which Website content could be accessed, read and understood, two validated readability tools for the Italian language were used: namely, the Gulpease and the ReadIt DyLanLab grade level scores.

The Gulpease index (5, 21) was initially defined in 1982 and developed in 1988 at the University Roma "La Sapienza", Rome, Italy, by the Linguistic Pedagogical University Group (in Italian, "Gruppo Universitario Linguistico Pedagogico" or GULP), in collaboration with the Italian section of the International Business Machines Corporation (IBM). It computes the score of ease to read and understand according to the following formula:

$$\text{Gulpease Index} = 89 + \frac{300 \cdot (\text{number of sentences}) - 10 \cdot (\text{number of letters})}{\text{number of words}}$$

In particular, the Gulpease grade level can be interpreted in the following way:

- a Gulpease score <40 suggests that the content is difficult to read for a subject with a higher school diploma;
- a Gulpease score <60 suggests that the content is difficult to read for a subject with a middle/secondary school diploma;
- a Gulpease score <80 suggests that the content is difficult to read for a subject with an elementary diploma.

In other words, lower figures indicate harder texts to read.

The Read-It - DyLanLab score (current version v2.1.9), developed by the Computational Linguistics Institute (in Italian, "Istituto di Linguistica Computazionale"), Pisa, Italy (14), differently from the Gulpease Index, assesses different content components, namely the lexical, syntactic and morpho-syntactic features of the text, among others. As such, the Read-It - DyLanLab tool returns the users with different scores: namely, the Base Model (measuring "classical" readability features, like average length of words and sentences), the Lexical Model (measuring vocabulary size and use, and lexical richness, for example counting the number of difficult words), the Syntax Model (measuring syntactic and morpho-syntactic characteristics, including the use of difficult grammatical rules) and the composite Global Model (combining all the previous models) scores.

These two grade levels are inversely related in that higher Read-It - DylanLab scores indicate harder texts to read; therefore, a higher Gulpease score should correlate to a lower Read-It DyLanLab grade level, and *vice-versa.* However, the two indices are not perfectly complementary but explore two different domains of readability. Read-It DyLanLab grade level explores parameters linked to structural complexity, grammatical correctness and meaning, whilst Gulpease is more focused on distributional parameters, like average number of words per sentence, average length of sentences, and number of syllables per word, among others (21).

Statistical analysis

Continuous data were presented as mean±standard deviation, together with the minimum-maximum ranges, the 25th and the 75th percentiles. Average ranks and medians were also reported, where appropriate. Categorical data were reported as percentages.

Before commencing statistical analyses, data were checked for normal distribution using the Shapiro-Wilk's test. The assumption of equality in variances was controlled performing the Levene's test. In case of significant deviation from normality and from homoscedasticity, non parametric tests (such as the Kruskal-Wallis one-way analysis of variance test) were carried out.

Non-parametric correlational analysis was performed between the different computed readability indices.

All analyses were performed using the commercial software MedCalc Statistical Software version 16.4.3 (MedCalc Software bvba, Ostend, Belgium; https://www.medcalc.org; 2016).

Figures with p-values <0.05 were considered statistically significant.

Results

150 sites were initially screened. Finally, 70 unique and silicosis relevant website pages were

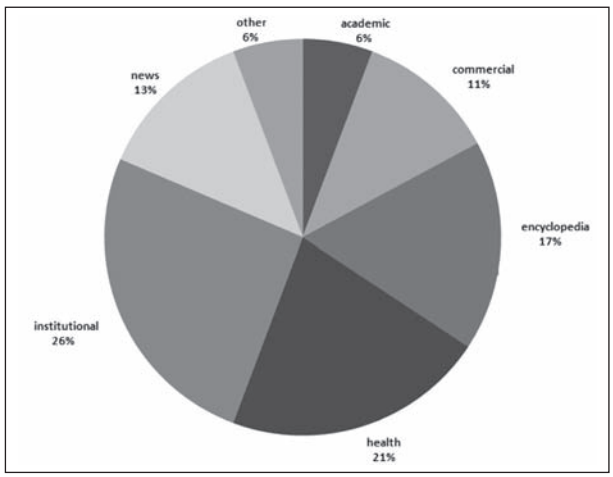

Figure 1 - Included silicosis-related website pages broken down for the different types of websites

analyzed, leftovers were excluded according to exclusion criteria listed in table 1. Approximately half of the websites are institutional and health-related pages (figure 1).

Concerning reliability, only 1.4% of websites adhere to HONCode standards.

Readability of the silicosis-related content is highly variable (table 2). As hypothesized, we observed an inverse relation between the Gulpease score and all the Read-It - DyLanLab grade levels, except for the Lexical model score (table S1).

The Gulpease score (table S2) differed among the different types of websites (Kruskal-Wallis ANOVA=12.93, p-value=0.044, p-value <0.05; figure 2). In particular, at the post-hoc analysis, academic websites significantly differed from institutional websites, as well as encyclopedia/dictionary pages from institutional sites, in a statistically significant way (p-value <0.05).

The Lexical Model - Read-It DyLanLab grade level (table S2) differed among the different types of websites (Kruskal-Wallis ANOVA=23.54, p-value=0.0006, p-value <0.001; figure S1). In particular, encyclopedia/dictionary pages differed from aca-

Table 2 - Readability indices for included websites

Readability Index	Mean	Standard Deviation	Minimum	Maximum	25th - 75th percentiles
Base Model - ReadIt DyLanLab	75.38	31.61	0.7	99.7	71.4 to 97.5
Global Model - ReadtIt DyLanLab	95.87	14.03	7.8	100.0	99.5 to 100.0
Gulpease Index	46.74	7.02	34.7	72.5	42.2 to 48.0
Lexical - ReadIt DyLanLab	73.56	29.20	0.1	100.0	54.0 to 98.0
Syntax Model - ReadIt DyLanLab	81.71	28.46	6.1	100.0	70.0 to 100.0

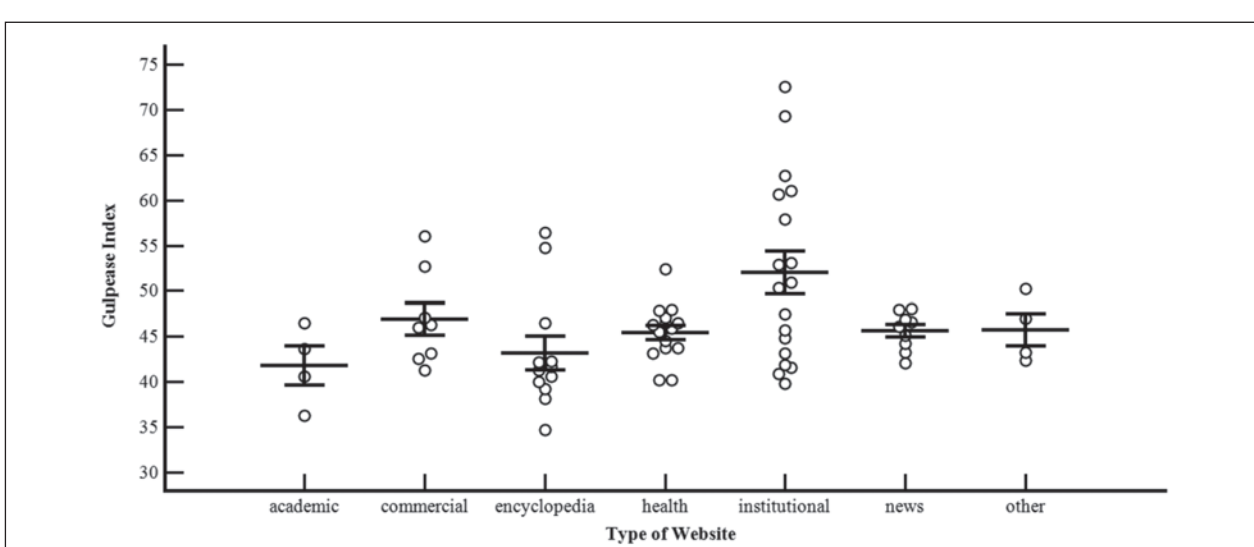

Figure 2 - A box-and-whiskers plot showing the Gulpease Index broken down for the different types of websites

demic, commercial, health-related, institutional and news sites in a statistically significant way (p-value <0.05).

On the contrary, the Base Model, the Syntax Model and the Global Model Read-It - DyLanLab grade levels did not statistically differ among the different types of websites. For further details, the reader is referred to table S2.

Approximately, half of the websites are intended or designed for workers whilst remaining 40% for physicians.Only the Global Model - Read-It DyLanLab grade level differed according to the intended/designed target (Kruskal-Wallis ANOVA=4.56, p-value= 0.031; p-value <0.05; figure S2).

Differently, the other readability indices did not statistically differ among the targets of website pages. For further details, the reader is referred to table S3.

Discussion

This is the first Italian study with respect to a classical occupational disease, investigating the quality of online material in terms of reliability and readability.

We found that very few sites adhere to HON-Code standards. This is in line with the literature, which reports a low reliability level for health-related or medical websites (see for example (3)).

Interestingly, about half of the websites are intended or designed for workers, indicating that currently the Internet is overlooked as a potential tool in the field of occupational health. It could be better exploited for improving work-relate health and wellbeing.

There is a certain degree of variability in readability score in online silicosis-related material, and this is especially true for institutional websites, even though all sites appear equally poor. Of particular concern is the fact that particularly popular pages such as the Wikipedia website, are difficult to read and understand for lay people. This confirms the findings from the extant literature (10, 11). Also in other medical fields, such as nephrology (17), psychiatry (15), oncology and preventive medicine (19), findings indicate that patient information materials - either online and non-Internet materials - are pitched above the average patient's literacy level, whilst current guidelines recommend that patient education materials should target a 5th grade literacy level (around 10-11 years) (17, 22). We could not make a direct comparison with the readability of occupational health-related websites in other countries and in other languages, since so far studies of this kind have not been conducted yet.

Our findings may have important practical implications for health-care workers and, in particular, for occupational physicians and health agencies/authorities. They should make efforts in strengthening their online presence and in further simplifying the online material, for health promotion purpose. This could have positive outcomes in term of occupational health, such as in behavior in the workplace and adoption of proper preventive measures enabling workers to increase and to improve their work-related health and its determinants.

However, this remains a speculation due to unavailability of randomized controlled trials (RCTs) assessing and quantifying the health effect of unbiased, scientific evidence-based occupational health online information, which remains a major limitation of the present study.

On the other hand, our study has a number of strengths that are given by the systematic approach used for selecting the website pages for inclusion, the use of two independent reviewers and of validated tools.

Conclusion

Generally speaking, medical institutions and, more specifically, occupational health agencies and authorities, are well positioned, accounting approximately for a quarter of silicosis-related website pages. However, more efforts should be geared towards further increasing the online presence of institutions on the Internet, in terms of medical and scientific leadership, and improving the quality of the content in terms of reliability, accuracy, and above all readability of information. Particular emphasis should be laid on simplifying online occupational-related materials in order to enhance user comprehension. Readability tools similar to those adopted in the present study could be exploited in this effort.

In conclusion, higher-quality information about silicosis using plain standard language is highly and urgently needed on the Internet to promote occupational health and to provide accurate medical information for affected workers.

No potential conflict of interest relevant to this article was reported by the authors

References

1. Akgün M: Denim production and silicosis. Curr Opin Pulm Med 2016; 22: 165-169
2. Akgun M, Mirici A, Ucar EY, et al: Silicosis in Turkish denim sandblasters. Occup Med (Lond) 2006; 56: 554-558
3. Al-Taha MT, Al Youha SA, Bull CE, et al: What Are Your Patients Reading Online About Soft-tissue Fillers? An Analysis of Internet Information. Plast Reconstr Surg Glob Open 2016; 4: e824
4. Barmania S: Deadly denim: sandblasting-induced silicosis in the jeans industry. Lancet Respir Med 2016; 4: 543
5. Berto D, Peroni M, Milleri S, et al: Evaluation of the readability of information sheets for healthy volunteers in phase-I trials. Eur J Clin Pharmacol 2000; 56: 371-374
6. Boyer C, Baujard V, Geissbuhler A: Evolution of Health Web certification through the HONcode experience. Stud Health Technol Inform 2011; 169: 53-57
7. Bragazzi NL: From P0 to P6 medicine, a model of highly participatory, narrative, interactive, and «augmented» medicine: some considerations on Salvatore Iaconesi's clinical story. Patient Prefer Adherence 2013; 7: 353-359
8. Bragazzi NL, Dini G, Toletone A, et al: Leveraging Big Data for Exploring Occupational Diseases-Related Interest at the Level of Scientific Community, Media Coverage and Novel Data Streams: The Example of Silicosis as a Pilot Study. PLoS One 2016; 11: e0166051
9. Bragazzi NL, Dini G, Toletone A, et al: Infodemiological data concerning silicosis in the USA in the period 2004-2010 correlating with real-world statistical data. Data Brief 2017; 10:457-464
10. Brigo F, Erro R: The readability of the English Wikipedia article on Parkinson's disease. Neurol Sci 2015; 36: 1045-1046
11. Brigo F, Otte WM, Igwe SC, et al: Clearly written, easily comprehended? The readability of websites providing information on epilepsy. Epilepsy Behav 2015; 44: 35-39
12. Cimrin A, Sigsgaard T, Nemery B: Sandblasting jeans kills young people. Eur Respir J 2006; 28: 885-886
13. Copello F, Garbarino S, Messineo A, et al: Occupational Medicine and Hygiene: applied research in Italy. J Prev Med Hyg 2015; 56: E102-110
14. Dell'Orletta F, Montemagni S, Venturi G: READ-IT: Assessing Readability of Italian Texts with a View to Text Simplification. Proceedings of the 2nd Workshop on Speech and Language Processing for Assistive Technologies. Edinburgh, Scotland, UK, July 30, 2011: 73-83.
15. Ipser JC, Dewing S, Stein DJ: A systematic review of the quality of information on the treatment of anxiety disorders on the internet. Curr Psychiatry Rep 2007; 9: 303-309
16. Manzoli L, Sotgiu G, Magnavita N, et al: Evidence-based approach for continuous improvement of occupational health. Epidemiol Prev 2015; 39 (4 Suppl 1): 81-85
17. Morony S, Flynn M, McCaffery KJ, et al: Readability of Written Materials for CKD Patients: A Systematic Review. Am J Kidney Dis 2015; 65: 842-850
18. Peterson G, Aslani P, Williams KA: How do consumers search for and appraise information on medicines on the Internet? A qualitative study using focus groups. J Med Internet Res 2003; 5: e33
19. Shahar S, Shirley N, Noah SA: Quality and accuracy assessment of nutrition information on the Web for cancer prevention. Inform Health Soc Care 2013; 38: 15-26
20. Siliquini R, Ceruti M, Lovato E, et al: Surfing the internet for health information: an Italian survey on use and population choices. BMC Med Inform Decis Mak 2011; 11: 21.
21. Terranova G, Ferro M, Carpeggiani C, et al: Low quality and lack of clarity of current informed consent forms in cardiology: how to improve them. JACC Cardiovasc Imaging 2012; 5: 649-655
22. Tieman J, Bradley SL: Systematic review of the types of methods and approaches used to assess the effectiveness of healthcare information websites. Aust J Prim Health 2013; 19: 319-324