Bring on the Machines: Could Machine Learning Improve the Quality of Patient Education Materials? A Systematic Search and Rapid Review

abstract

Purpose Clear and trustworthy information is essential for people who are ill. People with cancer, in particular, are targeted with vast quantities of patient education material, but of variable quality. Machine learning technologies are popular across industries for automated tasks, like analyzing language and spotting readability issues. With the experience of patients with cancer in mind, we reviewed whether anyone has proposed, modeled, or applied machine learning technologies for the assessment of patient education materials and explored the utility of this application.

Methods We systematically searched the literature to identify English-language articles published in peer-reviewed journals or as conference abstracts that proposed, used, or modeled the use of machine learning technology to assess patient education materials. Specifically, we searched MEDLINE, Web of Science, CINAHL, and Compendex. Two reviewers assessed study eligibility and performed study screening.

Results We identified 1,570 publications in our search after duplicate removal. After screening, we included five projects (detailed in nine articles) that proposed, modeled, or used machine learning technology to assess the quality of patient education materials. We evaluated the utility of each application across four domains: multidimensionality (2 of 5 applications), patient centeredness (1 of 5 applications), customizability (0 of 5 applications), and development stage (theoretical, 1 of 5 applications; in development, 3 of 5 applications; complete and available, 1 of 5 applications). Combining points across each domain, the mean utility score across included projects was 1.8 of 5 possible points.

Conclusion Given its potential, machine learning has not yet been leveraged substantially in the assessment of patient education materials. We propose machine learning systems that can dynamically identify problematic language and content by assessing the quality of patient education materials across a range of flexible, customizable criteria. Assessment may help patients and families decide which materials to use and encourage developers to improve materials overall.

Clin Cancer Inform. © 2018 by American Society of Clinical Oncology

Catherine H. Saunders
Curtis L. Petersen
Marie-Anne Durand
Pamela J. Bagley
Glyn Elwyn

Author affiliations and support information (if applicable) appear at the end of this article.

Corresponding author:
Glyn Elwyn, MD, PhD,
The Preference Laboratory,
The Dartmouth Institute
for Health Policy &
Clinical Practice, Level 5
Williamson Translational
Research Bldg, 1 Medi-
cal Center Dr, Lebanon,
NH 03756; e-mail:
glynelwyn@gmail.com.

INTRODUCTION

Clear, balanced, and evidence-based patient education materials are an essential component of high-quality medical care. Across many disease states, high-quality information exchange can improve patient engagement, reduce distress, improve adherence to treatment plans, and foster understanding among people who are ill and their families about what to expect.^{1,2}

Clear communication is especially important for people with cancer, a disease highly correlated with psychological distress.^{3,4} In addition to confronting suffering associated with diagnosis,

prognosis, treatment effects, and the possibility of death, people with cancer and their families also have to make complex decisions about how to proceed with management of their disease.¹ These decisions necessitate a firm grasp of the illness itself, realistic treatment options, and risks and benefits of various treatments.¹

Written education materials constitute a significant element of patient communication. Since the meaningful use requirements went into effect, health care systems routinely administer after-visit summaries to patients, which contain written information about patients' health condition,

treatment, and care plan.⁵ Patients also receive information in the form of brochures, handouts, and flyers from their clinicians. A reported 77% of people seek health information on the Internet, which has resources of variable quality.⁶

People with cancer also expect and prefer honest communication about their disease and as much information as possible,^{7,8} but often, they find the information provided by the health care system disappointing.^{9,10} Common flaws in cancer materials include poor risk communication, biased presentation of information, insufficient information about medical facts, and lack of transparency about who developed the resources.^{11,12} They also receive so much information that they often cannot recall needed details or recall them accurately.^{9,10}

Given the ubiquity and patient-reported importance of written health information, assessment of the quality of these materials is essential, especially in cancer care. Assessments of patient education quality can help patients make decisions about which sources to trust. The science in this area, however, is handicapped by limited methods. Analyses span from simple arithmetic or algorithmic readability assessments to in-depth content analyses. One example is The Patient Education Materials Assessment Tool (PEMAT), which guides the assessment of the understandability and actionability of education materials.¹³ The Health on the Net Foundation Code of Conduct (HONcode) offers certifications of content credibility, for a fee, that are based on manually evaluated criteria.¹⁴ The DISCERN Instrument is a free health content quality tool for the assessment of a variety of factors, including bias and information sources.¹⁵ Patient decision aids are another type of education material often evaluated with a set of quality criteria, such as the International Patient Decision Aid Standards (IPDAS) or the newer National Standards for the Certification of Patient Decision Aids from the National Quality Forum (NQF).^{16,17} Cancer decision aids often score poorly on these quality assessments.¹⁸ In a systematic Web review of decision aids for women who face genetic testing for familial breast cancer, for instance, Williams et al¹⁸ found that none of the tools scored > 50% on the IPDAS criteria. Like the PEMAT, raters must assess IPDAS and NQF criteria manually. In addition, interrater reliability, even among trained raters, can be variable.^{16,19} Perhaps

because of their manual nature, these tools do not provide mechanisms to carefully assess written content in depth, including accuracy, phrasing, and potential bias. This more rigorous assessment of written materials would require more time and human resources.

Few techniques exist for rapidly reviewing the quality of large bodies of written health information across domains that matter, such as evidence-based information, tone, and bias. Quick quality scoring and problem identification could increase the quantity of total reviewed patient education material and improve the quality and trustworthiness of materials overall.

Machine learning methods may be able to help. Since the emergence of computers in the 1950s and 1960s, methods have emerged to analyze data. Much of this work was and still is focused on classification problems.²⁰ Natural language processing uses a wide range of techniques for analyzing and interpreting human language.^{21,22} Almost from the beginning, the potential application of machine learning algorithms in medicine was evident, and health care professionals have used these algorithms in everything from diagnosis to treatment.²⁰ In other industries, machine learning has been used for tasks as varied as increasing efficiency and productivity to handling questions and response systems.^{23,24} In health care, machine learning applications and algorithms mainly concern the delivery of medicine and the distillation of medical knowledge. Less attention has been paid to communication directly with patients.

We conducted a preliminary review of the literature and could not find any instances of researchers who applied machine learning technology to written patient education materials. Therefore, we conducted a systematic search and rapid review to determine whether any applications of machine learning technology to the assessment of patient education materials exist and report on their utility. The purpose of this exercise is to assess the state of the field and recommend future applications.

METHODS

Aim, Objectives, Research Question

The aim of this study was to determine whether machine learning technology (intervention) has been applied to written patient education materials

(problem) to assess and improve their quality (outcome of interest). We did not have a comparison intervention for this study. Our objectives were to identify all peer-reviewed publications and academic medical conference abstracts about the application of machine learning techniques to the assessment of patient education material quality and to assess the utility of these applications of machine learning technology to patient education materials. Our research questions were: (1) has anyone has proposed, modeled, or applied machine learning technology to patient education materials? (2) if so, what is the utility of these applications? This was a rapid review completed in a short timeframe, therefore, we did not publish a protocol ahead of time.

Eligibility Criteria

We included studies about the application of machine learning technology (intervention) to written patient education materials (problem) and assessed their quality (the outcome of interest). We did not have a comparator group. We selected English-language studies published in peer-reviewed journals as well as abstracts from medical conferences.

We did not limit by study design so that we could identify any evidence that machine learning has been used to assess written patient education materials. If we were evaluating efficacy of machine learning technology for the assessment of written patient education materials, we would have limited by study design. We also did not limit by country.

We excluded readability assessments of patient education material in this review, although some readability assessments use simple machine learning algorithms. Our rationale was that readability assessment mechanisms are well understood, especially in cancer care.²⁵⁻²⁷ In addition, machine learning readability efforts have drawn mainly from the work of existing readability assessment algorithms, such as the Flesch-Kincaid readability test.²⁸

Machine learning criteria (level 1). We defined machine learning as a set of processes wherein a computer learns without being programmed by a human being.²⁹ We classified publications as concerning machine learning if they used any of the established machine learning technologies, including but not limited to decision tree, support

vector machines, naive Bayes classifiers, nearest neighbors, k-nearest neighbors algorithms, k-means, random forests, dimensionality reduction algorithms, gradient boosting algorithms, feature-based analysis, vision-based detection, expert systems, knowledge-based systems, voice recognition, neural networks, fuzzy modeling, pattern matching algorithms, and natural language processing.^{20,21,23} We excluded linear and logistic regression because researchers typically use these technologies for more straightforward analyses.

Patient education material criteria (level 2). We defined patient education material as an artifact intended to educate someone about his or her health condition. We excluded face-to-face interactions with health care practitioners, but we did include online resources, booklets, pamphlets, videos, and other similar items.

Assessment criteria (level 3). To be considered an assessment, projects must have included some rating, ranking, or analysis of patient education material. The generation, recall, and organization of patient education materials would not meet these criteria. Similarly, the identification of patients' education needs would not meet these criteria. Furthermore, the authors must have proposed, modeled, or applied one of these methods to patient education materials.

Information Sources

We used a search strategy developed in partnership with a research librarian (P.J.B.) to identify all peer-reviewed publications and conference abstracts on natural language processing and patient education materials published in English-language journals between 1966 and 2017. We included the following databases in the search: MEDLINE (Ovid), Web of Science, CINAHL, and Compendex. We reviewed the references of relevant articles to identify studies we may have missed with the digital search strategy. We performed the search on October 20, 2017. The complete search strategy is presented in Appendix [Table A1](#).

Study Selection

We imported the results of each database search into the reference management software Mendeley. Two reviewers (C.H.S. and C.L.P.)

independently reviewed titles and abstracts of each article to determine eligibility for inclusion. When the reviewers disagreed, they reached consensus through discussion. Both researchers examined the full texts of articles that seemed to meet inclusion criteria or that we could not rule out by reviewing the titles and abstracts. Again, disagreements were resolved through conversation and consensus.

Data Collection Process

Both reviewers independently extracted data of interest for each included study using a data collection form developed collaboratively (Appendix [Table A2](#)). These data items were date, title, authors, data source, study design or article type, machine learning method, data source, and patient education material type. The data collection form also included assessments of utility. If we could not find the information of interest, we planned to reach out to the study authors to identify missing information.

Risk of Bias

Given that the objective was to identify any instances of the application of machine learning technology to assess written patient education materials and that we were not interested in examining the effectiveness of these techniques, we did not include a risk of bias assessment for individual studies in the analysis. In addition, model development and design articles typically do not follow experimental procedures; however we noted the study type with the understanding that randomized studies inherently have the least bias, nonrandomized studies have a higher risk of bias, and pieces such as editorials have the highest risk of bias. Furthermore, we did not anticipate and did not assess risk of bias across studies.

Summary Measures

The intended summary measure was a quantification of the number of publications that mentioned machine learning technology and its application or potential application to written patient education materials. We also assessed the utility of each model.

Analysis

We developed and piloted a measure of machine learning application utility for the assessment of patient education materials on the basis of the reviewers' (C.H.S. and C.L.P.) assessment of which utilities were important to include. Our utility scorecard included four items: dimensionality, customizability, patient centeredness, and development stage. We assessed each item on a 1-point scale, with 1 point for meeting the criteria and 0 points for not meeting the criteria. The exception was the development stage, which we assessed on a 2-point scale.

We defined dimensionality as whether the program addressed more than one specific identifiable category of components. For example, an assessment of readability would be unidimensional even if readability was assessed through a number of means. A program that assessed both readability and tone of language would be multidimensional. We defined customizability as the ability of the end user to customize assessments on the basis of their needs. For instance, a person with cancer may be particularly interested in the accuracy and reliability of survival statistics, whereas a person with a chronic condition such as diabetes may be more interested in the overall tone of the content, like its patient centeredness. Patient centeredness was defined as the inclusion of patients and other end users in the model development.

We evaluated development stage on a 2-point scale. Publications or abstracts received 0 points for a purely theoretical model, 1 point for the in-development process for either the model or the corpus, and 2 points for a completed and available application. For the development stage, we gave 0 points to projects that had not implemented anything beyond a conceptual or theoretical examination. We gave 1 point to projects that had implemented some degree of machine learning algorithm to assessment or corpus development. Publications that had implemented a developed tool on a new corpus received 2 points.

We calculated a summary utility score out of 5 points for each included project (see Appendix [Table A3](#) for the utility scorecard). Both reviewers contributed to the development of the utility scorecard, which we piloted on included studies. We revised the utility scorecard twice.

RESULTS

Study Selection

The search returned 1,570 publications after the removal of duplicates (Fig 1). We excluded 197 articles that did not concern machine learning technology (level 1) and 127 that did not concern patient education (level 2). In addition, we excluded 96 articles that did not involve the assessment of patient education materials (level 3) but, rather, the development, classification, or anticipation of patients' need for them.³⁰ After full-text review, we excluded an additional four articles that did not concern the assessment of patient-education materials (level 3). We added four articles after reference review.

Although we required that publications concern the assessment of patient education materials, we screened out many articles about similar subjects. For example, the unscreened results contained many cases of machine learning technologies supporting clinicians in decision making, which we screened out. This phenomenon is well documented in the existing literature.³¹ In addition, there were many instances of the use of machine learning technology to

classify, customize, or generate patient education material. Like the assessment of patient education material, this is not well documented in the literature but did not qualify for inclusion in our review. Furthermore, we did not systematically assess the number of articles that fell into each excluded category beyond the three levels because this was not related to our a priori research question and search strategy (Appendix Table A4).

Study Characteristics

We ultimately included nine publications concerning five discrete projects in our assessment. All nine articles concerned machine learning model development for patient education material assessment. The lead authors for three articles were from the United States, one was from a team of Spanish researchers, and the authors of five articles were from Switzerland. Three projects leveraged natural language processing, one leveraged feature-based scoring, and another used a fuzzy linguistic approach. One project, which included five articles, used a combination of various machine learning approaches (Table 1).

Allam et al³² outlined the early stages of development for a machine learning algorithm, specifically natural language processing, that can automatically assess the quality of online materials using the DISCERN instrument, a free health information quality assessment tool.¹⁵ Similarly, the HONcode project by Boyer et al³³⁻³⁵ and the studies by Gaudinat et al^{36,37} assess an automated mechanism for identifying an article's adherence to the HONcode principles using a variety of machine learning technologies. The publicly available European Union medical search engine Khresmoi for Everyone uses similar machine learning capabilities to assess HONcode compliance of patient education material.

The other three articles' authors used machine learning technologies to assess not necessarily the overall quality of written health information but one-dimensional characteristics such as text simplification, information continuity, and online usability.³⁸⁻⁴⁰ Kauchak and Leroy³⁸ proposed the use of natural language processing to assess patient education materials and replace confusing text with more explicit language. The work is too preliminary to determine feasibility. Ballerio³⁹

Fig 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses flow diagram.

Table 1. Characteristics of Included Projects

First Author	Country	Design	Method	Source	Material
Quality assessment					
Allam ³²	United States	Model development	Probabilistic feature-based scoring	Online information	272 Web pages that discussed treatment options in breast cancer, arthritis, and depression
Boyer ³³⁻³⁵ ; Gaudinat ^{36,37}	Switzerland	Model development and assessment	Various	Online information	5,000 Web pages, 27 Web sites, Web page fragments,
Text simplification					
Kauchak ³⁸	United States	Model development	Natural language processing	Online information	Consumer and professional health information
Information continuity assessment					
Ballerio ³⁹	United States	Model development	Natural language processing	Discharge summaries	413 discharge letters
Usability assessment					
Moreno ⁴⁰	Spain	Model development	Fuzzy linguistic approach	Online health information	Health Web sites

NOTE. The Boyer and Gaudinat articles were part of the same project, so they are represented together.

modeled a natural language processing algorithm to detect follow-up instructions and other indications of care continuity in hospital discharge letters and concluded that more sophisticated natural language processing technologies are needed before the model can be implemented in routine care. Moreno et al⁴⁰ used a fuzzy linguistic approach to assess the usability and quality of health-related Web sites from a Web-based design perspective. They found that their method was feasible.

Analysis

With the use of our utility checklist, we assessed the average summary utility of the included studies as 1.8 of 5 possible points (Table 2). We evaluated the utility of five articles together because they were by the same authors about the same project. The article by Allam et al³² described a multidimensional quality assessment approach to automating the DISCERN criteria. All the included HONcode articles concerned a multidimensional quality assessment using HONcode.³⁵ Similarly, the Moreno et al project operationalized multidimensional criteria derived from qualitative research using machine learning technology.⁴⁰ The other included studies concerned unidimensional quality assessments. Kauchak and Leroy³⁸ involved end users in the development process, whereas the other authors did not. Kauchak and Leroy noted that

their feature-based approach improves the ability of their model to be flexible and adaptable to new features, although their existing features are all readability based, so it seems unlikely that the addition of another construct would be simple. Similarly, we did not determine that any of the included studies were customizable by the end user. The authors of the included publications possibly will adapt their models for customizability, but that has not happened to date. With regard to development stage, only the HONcode project scored 2 of 2 possible points because the platform was developed and tested and is now publicly available. The article by Moreno et al,⁴⁰ was mostly theoretical and thus scored 1 point. The remaining articles scored 1 point each because their models were in development but not yet publicly available for use on the new corpora.

DISCUSSION

The potential of machine learning to transform the way patient education materials are used and developed for cancer is unrealized. We found that machine learning technology has been harnessed to assess the quality of patient education materials but to a limited extent, context, and scope. Many machine learning applications for the assessment of patient education materials concern discrete concepts, such as readability and usability. Two instances exist of researchers

Table 2. Utility Scores

First Author	Dimensionality (0-1 points)	Patient Involvement (0-1 points)	Customizability (0-1 points)	Development (0-2 points)	Utility Summary Score (0-5 points)
Quality assessment					
Allam ³²	1	0	0	1	2
Boyer ^{35*}	1	0	0	2	3
Text simplification					
Kauchak ³⁸	0	1	0	1	2
Information continuity assessment					
Ballerio ³⁹ ; Gaudinat ^{36,37}	0	0	0	1	1
Usability assessment					
Moreno ⁴⁰	1	0	0	0	1

NOTE. Mean summary score = 1.8.

*The Health on the Net Foundation Code of Conduct studies were combined for assessment because they were about the same project and published by the same authors.

applying this technology to assess the quality of written patient education materials across multiple quality dimensions, but these applications merely automate existing analog quality assessment instruments.

The work of Allam et al³² on the DISCERN criteria and Boyer et al³³⁻³⁵ and Gaudinat et al^{36,37} on the HONcode project suggests that machine learning technology can be used to assess patient education materials. We also know that machine learning technologies, including natural language processing, have been successfully used in other health care capacities. For instance, in a systematic review of natural language processing applications in radiology, Pons et al⁴¹ found that 30% of models were currently integrated into routine care beyond test environments, which further supports the feasibility of machine learning technologies in health care. Like clinical notes, patient education materials are unstructured, which can make analysis challenging. For these types of analyses, typical natural language processing analyses or topic mapping techniques include term frequency-inverse document frequency or methods that use word frequency and probabilistic models for word presence.^{42,43} Another approach is sentiment analysis, which is typically a sum of sentiment scores over a sentence, paragraph, or sequence of items (*n*-gram) and requires precise rules.^{27,44}

Because of the inherent ambiguity and subjectivity associated with determining the quality of

patient education materials, use of these typical machine learning and natural language processing methods might not be feasible. Even highly trained researchers have variable agreement.¹⁹

We could, however, explore alternative machine learning methods for patient education material assessment. Researchers have not yet investigated these approaches in detail. Possibilities include latent semantic analysis (and by extension, latent dirichlet allocation), autoencoders, or deep artificial neural networks. Socher et al,⁴⁵ for example, developed a new model called the recursive neural tensor network and assessed its performance using excerpts of movie reviews. Their model obtained > 80% accuracy in fine-grained sentiment prediction, which surpasses previous attempts. Researchers could reasonably appropriate this approach for the assessment of patient education materials. Unsupervised machine learning methods like these have been used to automatically assess patient medical records to identify trends beyond the ability of human readers to grasp, such as co-occurrence of certain free-text medical language and subsequent medical events.⁴⁶ In this case, the researchers would develop a corpus of sentiments specific to the language and format of patient education materials.⁴⁵ Latent semantic analysis and latent dirichlet allocation methods might be the most appropriate because they could examine the underlying language

similarities of similarly categorized patient education materials.

Our review was limited to the published academic literature. Hospitals, health systems, or other private entities might have already operationalized machine learning concepts to improve patient education through assessment beyond what we found. We could not confirm or deny this because we did not systematically query stakeholders or organizations. In addition, determining whether publications concerned machine learning was challenging. The broad definition of machine learning has changed over time as methods and terms have developed and evolved. The elusiveness of a consensus machine learning definition is well documented.⁴⁷ For example, many articles used expert system methods, and although these usually are considered machine learning, they are rigid algorithms that are not based on statistical learning methods. The wide range of machine learning definitions could lead to misclassification or unidentified works but highlights the importance of the human review of each publication to mitigate such risk.

The rise of machine learning technologies in health care delivery and research has been meteoric in the past few years. Most of these efforts concern the improvement of clinician workflow, diagnostics, and other issues, whereas the potential of machine learning technology for improving patient concerns remains underused. Simultaneously, researchers have made painstaking, iterative efforts to develop reliable, valid, and user-centered measures of patient education materials over the past 15 years. The DISCERN instrument, the IPDAS collaboration, and the NQF criteria are just a few examples.^{15,17,48}

Given these two independently evolving efforts, it is surprising that machine learning application to written patient education materials has been so limited.

Machine learning methods may enable real-time assessment of patient education materials, but the technology allows for evaluations that go beyond simple manual standards. Quantitative assessment of characteristics that represent the trustworthiness of content across many domains, such as directedness (Do they tell patients what to do?), emotive language (Do they use descriptive words that may sway behavior?), and omission (Do they exclude viable options?), may be possible. Previously, these types of granular assessment were only possible with laborious, human analyses. Machine learning, specifically natural language processing and deep artificial neural networks, may democratize this knowledge, putting information about the ability of education materials to communicate critical concepts, decision points, and information in patients' hands.

Moving forward, we recommend the use of machine learning technology to assess patient education materials by using existing analog measures such as DISCERN and IPDAS. We also recommend the use of new technology to expand the scope of quality assessment to include more-detailed assessments in line with what matters to patients. Machine learning technology may facilitate faster testing of items, customizable quality reports, and more.

DOI: <https://doi.org/10.1200/CCI.18.00010>

Published online on ascopubs.org/journal/cci on November 9, 2018.

AUTHOR CONTRIBUTIONS

Conception and design: All authors

Collection and assembly of data: Catherine H. Saunders, Curtis L. Petersen, Marie-Anne Durand, Pamela J. Bagley

Data analysis and interpretation: Catherine H. Saunders, Curtis L. Petersen, Marie-Anne Durand

Manuscript writing: All authors

Final approval of manuscript: All authors

Accountable for all aspects of the work: All authors

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

The following represents disclosure information provided by authors of this manuscript. All relationships are considered compensated. Relationships are self-held unless noted. I = Immediate Family Member, Inst = My

Institution. Relationships may not relate to the subject matter of this manuscript. For more information about ASCO's conflict of interest policy, please refer to www.asco.org/rwc or ascopubs.org/jco/site/ifc.

Catherine H. Saunders

No relationship to disclose

Curtis L. Petersen

No relationship to disclose

Marie-Anne Durand

Consulting or Advisory Role: ACCESS Community Health Network, EBSCO Health

Patents, Royalties, Other Intellectual Property: Co-developer of the OptionGrid patient decision aids, which are licensed to EBSCO Health

Travel, Accommodations, Expenses: EBSCO Health, ACCESS Community Health Network

Pamela J. Bagley

No relationship to disclose

Glyn Elwyn

Consulting or Advisory Role: ACCESS Community Health Network, Chicago (Federally Qualified Medical Centers), EBSCO Health

Patents, Royalties, Other Intellectual Property: Director of &think LLC, which owns the registered trademark for Option Grids™ patient decision aids. Owns copyright in measures of shared decision making and care integration, namely, collaboRATE, integRATE, and Observer OPTION-5 and Observer OPTION-12, which are freely available for use.

Travel, Accommodations, Expenses: ACCESS Community Health Network, Chicago (Federally Qualified Medical Centers), EBSCO Health

Affiliations

Catherine H. Saunders, Curtis L. Petersen, Marie-Anne Durand, and Glyn Elwyn, The Dartmouth Institute for Health Policy & Clinical Practice; **Curtis L. Petersen,** Geisel School of Medicine at Dartmouth; and **Pamela J. Bagley,** Dartmouth College, Lebanon, NH.

REFERENCES

1. McPherson CJ, Higginson IJ, Hearn J: Effective methods of giving information in cancer: A systematic literature review of randomized controlled trials. *J Public Health Med* 23:227-234, 2001
2. Mills ME, Sullivan K: The importance of information giving for patients newly diagnosed with cancer: A review of the literature. *J Clin Nurs* 8:631-642, 1999
3. Krebber AMH, Buffart LM, Kleijn G, et al: Prevalence of depression in cancer patients: A meta-analysis of diagnostic interviews and self-report instruments. *Psycho-oncology* 23:121-130, 2014
4. Bultz BD, Carlson LE: Emotional distress: The sixth vital sign in cancer care. *J Clin Oncol* 23:6440-6441, 2005
5. Hummel J: Providing Clinical Summaries to Patients After Each Office Visit: A Technical Guide, 2012. <https://www.healthit.gov/sites/default/files/measure-tools/avs-tech-guide.pdf>
6. Fox S: 2012 Health Survey Data, 2013. <http://www.pewinternet.org/2013/02/11/2012-health-survey-data>
7. Butow PN, Maclean M, Dunn SM, et al: The dynamics of change: Cancer patients' preferences for information, involvement and support. *Ann Oncol* 8:857-863, 1997
8. Cassileth BR, Zupkis RV, Sutton-Smith K, et al: Information and participation preferences among cancer patients. *Ann Intern Med* 92:832-836, 1980
9. Ley P: Psychology and Medicine Series: Communicating With Patients: Improving Communication, Satisfaction and Compliance. New York, NY, Croom Helm, 1988
10. Hogbin B, Fallowfield L: Getting it taped: The "bad news" consultation with cancer patients. *Br J Hosp Med* 41:330-333, 1989
11. Slaytor EK, Ward JE: How risks of breast cancer and benefits of screening are communicated to women: Analysis of 58 pamphlets. *BMJ* 317:263-264, 1998
12. Fagerlin A, Rovner D, Stableford S, et al: Patient education materials about the treatment of early-stage prostate cancer: A critical review. *Ann Intern Med* 140:721-728, 2004
13. Shoemaker SJ, Wolf MS, Brach C: Development of the Patient Education Materials Assessment Tool (PEMAT): A new measure of understandability and actionability for print and audiovisual patient information. *Patient Educ Couns* 96:395-403, 2014
14. Health On the Net: Health On the Net Code of Conduct (HONcode). <https://www.hon.ch/HONcode/Patients>
15. Charnock D, Shepperd S, Needham G, et al: DISCERN: An instrument for judging the quality of written consumer health information on treatment choices. *J Epidemiol Community Health* 53:105-111, 1999

16. Elwyn G, O'Connor A, Stacey D, et al: Developing a quality criteria framework for patient decision aids: Online international Delphi consensus process. *BMJ* 333:417, 2006
17. National Quality Forum: National Standards for the Certification of Patient Decision Aids. Washington, DC, National Quality Forum, 2016
18. Williams L, Jones W, Elwyn G, et al: Interactive patient decision aids for women facing genetic testing for familial breast cancer: A systematic Web and literature review. *J Eval Clin Pract* 14:70-74, 2008
19. Elwyn G, O'Connor AM, Bennett C, et al: Assessing the quality of decision support technologies using the International Patient Decision Aid Standards instrument (IPDASI). *PLoS One* 4:e4705, 2009
20. Kononenko I: Machine learning for medical diagnosis: History, state of the art and perspective. *Artif Intell Med* 23:89-109, 2001
21. Chowdhury GG: Natural language processing. *Annu Rev Inf Sci Technol* 37:51-89, 2005
22. Liddy ED: Natural language processing, in Drake MA (ed): *Encyclopedia of Library and Information Science* (ed 2). New York, NY, Marcel Dekker, 2003, pp 2126-2136
23. Bose I, Mahapatra RK: Business data mining—a machine learning perspective. *Inf Manage* 39:211-225, 2001
24. Riezler S, Vasserman A, Tsochantaridis I, et al: Statistical machine translation for query expansion in answer retrieval. *Proc Assoc Computational Linguistics* 1:464-471, 2007
25. Friedman DB, Hoffman-Goetz L: A systematic review of readability and comprehension instruments used for print and Web-based cancer information. *Heal Educ Behav* 33:352-373, 2006
26. Schwarm SE, Ostendorf M: Reading level assessment using support vector machines and statistical language models. *Proc Assoc Computational Linguistics*, 523-530, 2005
27. Petersen SE, Ostendorf M: A machine learning approach to reading level assessment. *Comput Speech Lang* 23:89-106, 2009
28. Flesch R: A new readability yardstick. *J Appl Psychol* 32:221-233, 1948
29. Samuel AL: Some studies in machine learning using the game of checkers. *IBM J Res Dev* 3:210-229, 1959
30. Lisboa PJ, Taktak AFG: The use of artificial neural networks in decision support in cancer: A systematic review. *Neural Networks* 19:408-415, 2006
31. Kerr WT, Anderson A, Lau EP, et al: Automated diagnosis of epilepsy using EEG power spectrum. *Epilepsia* 53:e189-e192, 2012
32. Allam A, Schulz PJ, Krauthammer M: Toward automated assessment of health Web page quality using the DISCERN instrument. *J Am Med Informatics Assoc* 24:481-487, 2017
33. Boyer C, Dolamic L, Grabar N: Automated detection of health websites' HONcode conformity: Can N-gram tokenization replace stemming? *Stud Health Technol Inform* 216:1064, 2015
34. Boyer C, Dolamic L: Feasibility of automated detection of HONcode conformity for health-related websites. *Int J Adv Comput Sci Appl* 5:69-74, 2014
35. Boyer C, Dolamic L: Automated detection of HONcode Website conformity compared to manual detection: An evaluation. *J Med Internet Res* 17:e135, 2015
36. Gaudinat A, Grabar N, Boyer C: Machine learning approach for automatic quality criteria detection of health web pages. *Stud Health Technol Inform* 129:705-709, 2007
37. Gaudinat A, Grabar N, Boyer C: Combination of heterogeneous criteria for the automatic detection of ethical principles on health web sites. *AMIA. Annu Symp proceedings AMIA Symp*, 264-268, 2007
38. Kauchak D, Leroy G: Moving beyond readability metrics for health-related text simplification. *IT Prof* 18:45-51, 2016

39. Ballerio S: Automatic analysis of electronic discharge letters as a means to evaluate the continuity of information and of patient care. *Proc 13th Int Conf Electron Publ*, 607-612, 2009
40. Moreno JM, Morales del Castillo JM, Porcel C, et al: A quality evaluation methodology for health-related websites based on a 2-tuple fuzzy linguistic approach. *Soft Comput* 14:887-897, 2010
41. Pons E, Braun LMM, Hunink MGM, et al: Natural language processing in radiology: A systematic review. *Radiol* 279:329-343, 2016
42. Liu B: *Sentiment Analysis and Opinion Mining*. San Rafael, CA, Morgan & Claypool, 2012
43. Medhat W, Hassan A, Korashy H: Sentiment analysis algorithms and applications: A survey. *Ain Shams Eng J* 5:1093-1113, 2014
44. Liu B: Sentiment analysis and subjectivity, 2010. <http://www.cs.uic.edu/~liub/FBS/NLP-handbook-sentiment-analysis.pdf>
45. Socher R, Perelygin A, Wu J, et al: Recursive deep models for semantic compositionality over a sentiment treebank, 2013. http://nlp.stanford.edu/~socherr/EMNLP2013_RNTN.pdf
46. Titus A, Faill R, Das A: Automatic Identification of Co-Occuring Patient Events. *Proc 7th ACM Int Conf Bioinformatics Computational Biol Health Inform*, 579-586, 2016
47. Fundamental Issues in Machine Learning. <http://ttic.uchicago.edu/~pengjian/MLCourse/intro.pdf>
48. International Patient Decision Aid Standards (IPDAS) Collaboration: IPDAS 2005: Criteria for judging the quality of patient decision aids, 2005. http://ipdas.ohri.ca/ipdas_checklist.pdf

Appendix

Table A1. Search Strategy

Concept		
Database	Machine Learning	Patient Education Materials
Ovid MEDLINE (epub ahead of print, in process, other nonindexed citations, daily) 1946 to present		
Thesaurus terms	Machine Learning/ or exp Supervised Machine Learning/ or Unsupervised machine learning/ or Fuzzy logic/ or exp natural language processing/ or exp "Neural Networks (Computer)"/ or Pattern Recognition, Automated/ or exp Artificial Intelligence/	exp Patient Education as Topic/ or Consumer health information/
Text words	(machine learning or fuzzy logic or automated pattern recognition or natural language processing or k clustering or support vector machines or decision tree* or neural network* or deep learning or artificial intelligence).ti,ab,kw.	(patient education or patient information or consumer education or consumer health).ti,ab,kw. ADJ5 (brochure* or handout* or leaflet* or material* or pamphlet* or publication* or resource* or information or written or decision aid* or Decision support tool* or communication* or source*).ti,ab,kw. OR (patient or patients or consumer*).ti,ab,kw. ADJ5 Health information.ti,ab,kw. Or Educational material*.ti,ab,kw. Or Education material*.ti,ab,kw. Or Prompt sheet*.ti,ab,kw. Or patient discharge instruction*.ti,ab,kw. or after visit summar*.ti,ab,kw. or AVS.ti,ab,kw. or patient portal.ti,ab,kw. or Video.ti,ab,kw. or Web.ti,ab,kw. Or patient discharge instructions.ti,ab,kw.
CINAHL (EBSCOhost; EBSCO Information Services, Ipswich, MA)		
Thesaurus terms	(MH "Artificial Intelligence+")	(MH "Patient Education+")
Text words	"machine learning" OR "fuzzy logic" OR "automated pattern recognition" OR "natural language processing" OR "k clustering" OR "support vector machine" OR "decision tree*" OR "neural network*" OR "deep learning" OR "artificial intelligence"	((("patient education" OR "patient information" OR "consumer education" OR "consumer health") N5 (brochure* OR handout* OR leaflet* OR material* OR pamphlet* OR publication* OR resource* OR information OR written OR "decision aid*" OR "Decision support tool*" OR communication* OR source*)) OR ((patient OR patients OR consumer*) N5 "Health information" OR "Educational material*" OR "Education material*" OR "Prompt sheet*" or "patient discharge instruction*" OR "after visit summar*" OR AVS OR "patient portal" OR Video or Web OR "patient discharge instructions"))
Web of Science (Clarivate Analytics, Philadelphia, PA)		

(Continued on following page)

Table A1. Search Strategy (Continued)

Concept		
Database	Machine Learning	Patient Education Materials
Text words	"machine learning" OR "fuzzy logic" OR "automated pattern recognition" OR "natural language processing" OR "k clustering" OR "support vector machine" OR "decision tree*" OR "neural network*" OR "deep learning" OR "artificial intelligence"	((("patient education" OR "patient information" OR "consumer education" OR "consumer health") NEAR/5 (brochure* OR handout* OR leaflet* OR material* OR pamphlet* OR publication* OR resource* OR information OR written OR "decision aid*" OR "Decision support tool*" OR communication* OR source*)) OR ((("patient education" OR "patient information" OR "consumer education" OR "consumer health") NEAR/5 ("Health information" OR "Educational material*" OR "Education material*" OR "Prompt sheet*" OR "patient discharge instruction*" OR "after visit summar*" OR AVS OR "patient portal" OR Video or Web OR "patient discharge instructions"))
Compendex (Engineering Village; Elsevier, New York, NY)		
Thesaurus terms	(((((WN CV)) OR ((AUTOMATED PATTERN RECOGNITION SYSTEM)) WN CV)) OR ((FUZZY LOGIC)) WN CV) OR ((NEURAL NETWORKS)) WN CV) OR ((SUPPORT VECTOR MACHINES)) WN CV) OR ((DECISION TREES)) WN CV) OR ((ARTIFICIAL INTELLIGENCE - MEDICAL APPLICATIONS)) WN CV))	
Text words (searched in title, abstract, and subject)	"machine learning" OR "fuzzy logic" OR "automated pattern recognition" OR "natural language processing" OR "k clustering" OR "support vector machine" OR "decision tree*" OR "neural network*" OR "deep learning" OR "artificial intelligence"	((({machine learning} OR {fuzzy logic} OR {automated pattern recognition} OR {natural language processing} OR {k clustering} OR {support vector machine} OR {decision tree*} OR {neural network*} OR {deep learning} OR {artificial intelligence}) WN KY AND (1884-2018 WN YR)))) AND (((({patient education} OR {patient information} OR {consumer education} OR {consumer health}) AND (brochure* OR handout* OR leaflet* OR material* OR pamphlet* OR publication* OR resource* OR \$information OR \$written OR {decision aid*} OR {Decision support tool*} OR communication* OR source*)) OR ((patient OR patients) AND ({Health information} OR {Educational material*} OR {Education material*} OR {Prompt sheet*} OR {patient discharge instruction*} OR {after visit summar*} OR \$AVS OR {patient portal} OR \$Video OR \$Web OR {patient discharge instructions}))))

Table A2. Data Collection Instrument

Year	Title	Authors	Source	Design	Method	Source	Country	Material	Notes
Year of publication	Title of publication	Authors of publication	Source of publication	Research design	Machine learning method used	Data source, test set, or corpus	Country of researchers	Type of patient education material assessed	

Table A3. Utility Checklist

Dimensionality	Customizability	Patient Centeredness	Development Stage	Utility Summary
Whether the program addresses more than one specific, identifiable category of components (1 point for yes, 0 points for no)	The ability of the end user to customize assessments on the basis of their needs (1 point for yes, 0 points for no)	Inclusion of patients and other end users in the model development (1 point for yes, 0 points for no)	Publications or abstracts received 0 points for a purely theoretical model, 1 point for the in-development process for either the model or the corpus, and 2 points for a completed and available assessment (2 points for completed assessment, 1 point for in development, 0 points theoretical)	Summary score that comprised the sum of four categories

Table A4. PRISMA Checklist

Section/Topic	Item No.	Checklist Item	Reported on Page No.
Title			
Title	1	Identify the report as a systematic review, meta-analysis, or both	1
Abstract			
Structured summary	2	Provide a structured summary, including, as applicable, background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number	2
Introduction			
Rationale	3	Describe the rationale for the review in the context of what is already known	3-4
Objectives	4	Provide an explicit statement of questions being addressed with reference to PICOS	4
Methods			
Protocol and registration	5	Indicate whether a review protocol exists and whether and where it can be accessed (eg, Web address) and, if available, provide registration information, including registration number	4
Eligibility criteria	6	Specify study characteristics (eg, PICOS, length of follow-up) and report characteristics (eg, years considered, language, publication status) used as criteria for eligibility, giving rationale	4
Information sources	7	Describe all information sources (eg, databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	5
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated	5; Appendix Table A1
Study selection	9	State the process for selecting studies (ie, screening and eligibility included in systematic review and, if applicable, included in the meta-analysis)	5
Data collection process	10	Describe method of data extraction from reports (eg, piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators	5
Data items	11	List and define all variables for which data were sought (eg, PICOS, funding sources) and any assumptions and simplifications made	5
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level) and how this information is to be used in any data synthesis	5
Summary measures	13	State the principal summary measures (eg, risk ratio, difference in means)	6
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (eg, I^2) for each meta-analysis	NA
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (eg, publication bias, selective reporting within studies)	5
Additional analyses	16	Describe methods of additional analyses (eg, sensitivity or subgroup analyses, metaregression), if done, indicating which were prespecified	6
Results			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram	6; Figure 1
Study characteristics	18	For each study, present characteristics for which data were extracted (eg, study size, PICOS, follow-up period) and provide the citations	7; Table 1

(Continued on following page)

Table A4. PRISMA Checklist (Continued)

Section/Topic	Item No.	Checklist Item	Reported on Page No.
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12)	NA
Results of individual studies	20	For all outcomes considered (benefits or harms), present for each study: (1) simple summary data for each intervention group, and (2) effect estimates and CIs, ideally with a forest plot	NA
Synthesis of results	21	Present results of each meta-analysis done, including CIs and measures of consistency	NA
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see item 15)	NA
Additional analysis	23	Give results of additional analyses, if done (eg, sensitivity or subgroup analyses, metaregression [see item 16])	NA
Discussion			
Summary of evidence	24	Summarize the main findings, including the strength of evidence for each main outcome; consider their relevance to key groups (eg, health care providers, users, and policymakers)	7
Limitations	25	Discuss limitations at the study and outcome level (eg, risk of bias) and at the review level (eg, incomplete retrieval of identified research, reporting bias)	7
Conclusions	26	Provide a general interpretation of the results in the context of other evidence and implications for future research	8
Funding			
Funding	27	Describe sources of funding for the systematic review and other support (eg, supply of data) and the role of funders for the systematic review	NA

NOTE. Adapted with permission from Moher et al: PLoS Med 6: e1000097, 2009.

Abbreviations: NA, not applicable; PICOS, participants, interventions, comparisons, outcomes, and study design; PRISMA, Preferred Reporting Items for Systematic Reviews and Meta-Analyses.