Validation of a Semiautomated Natural Language Processing–Based Procedure for Meta-Analysis of Cancer Susceptibility Gene Penetrance

Zhengyi Deng, MS¹; Kanhua Yin, MD, MPH¹; Yujia Bao, MA²; Victor Diego Armengol¹; Cathy Wang, MS^{3,4}; Ankur Tiwari, MBBS¹; Regina Barzilay, PhD²; Giovanni Parmigiani, PhD^{3,4}; Danielle Braun, PhD^{3,4}; and Kevin S. Hughes, MD^{1,5}

PURPOSE Quantifying the risk of cancer associated with pathogenic mutations in germline cancer susceptibility genes—that is, penetrance—enables the personalization of preventive management strategies. Conducting a meta-analysis is the best way to obtain robust risk estimates. We have previously developed a natural language processing (NLP)–based abstract classifier which classifies abstracts as relevant to penetrance, prevalence of mutations, both, or neither. In this work, we evaluate the performance of this NLP-based procedure.

MATERIALS AND METHODS We compared the semiautomated NLP-based procedure, which involves automated abstract classification and text mining, followed by human review of identified studies, with the traditional procedure that requires human review of all studies. Ten high-quality gene–cancer penetrance meta-analyses spanning 16 gene–cancer associations were used as the gold standard by which to evaluate the performance of our procedure. For each meta-analysis, we evaluated the number of abstracts that required human review (workload) and the ability to identify the studies that were included by the authors in their quantitative analysis (coverage).

RESULTS Compared with the traditional procedure, the semiautomated NLP-based procedure led to a lower workload across all 10 meta-analyses, with an overall 84% reduction (2,774 abstracts v 16,941 abstracts) in the amount of human review required. Overall coverage was 93%—we are able to identify 132 of 142 studies—before reviewing references of identified studies. Reasons for the 10 missed studies included blank and poorly written abstracts. After reviewing references, nine of the previously missed studies were identified and coverage improved to 99% (141 of 142 studies).

CONCLUSION We demonstrated that an NLP-based procedure can significantly reduce the review workload without compromising the ability to identify relevant studies. NLP algorithms have promising potential for reducing human efforts in the literature review process.

JCO Clin Cancer Inform. © 2019 by American Society of Clinical Oncology

INTRODUCTION

The magnitude of cancer risk—that is, penetrance—of pathogenic variants in hereditary cancer susceptibility genes is important for determining appropriate management strategies.¹ Quantifying risk allows clinicians to personalize preventive management strategies for patients. Relevant scientific literature, which clinicians need to review, is not only large, but also growing exponentially. In addition, most publications regarding cancer risk for specific gene–cancer associations vary in sample size and study quality, which results in large differences in risk estimates for the same gene–cancer association. Whereas most clinicians try to intuit which of the studies provide the most accurate estimates, a more robust and evidence-based approach to leveraging these data is to conduct a meta-analysis to estimate the risk for each gene–cancer association.

Unfortunately, performing a high-quality, methodologically sound meta-analysis requires a significant amount of effort. An analysis of 195 systematic reviews shows that the mean estimated time to complete and publish a project is approximately 17 months, whereas the mean ratio of studies that are ultimately included in the review to the number of studies found in the database searches—that is, yield rate—is only approximately 2.94%.² In cataloging the work process, study selection, including abstract screening, is the most labor-intensive and time-consuming step, accounting for an estimated 26% of project work.³ As the literature on cancer genetics grows, it is imperative to develop an effective procedure that can reduce the literature review burden without compromising comprehensiveness.

Natural language processing (NLP) is a computer science technology that uses algorithms to extract

ASSOCIATED CONTENT

Appendix

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

Accepted on July 16, 2019 and published at ascopubs.org/journal/cci on August 16, 2019; DOI <https://doi.org/10.1200/CCI.19.00043>

CONTEXT

Key Objective

We developed and evaluated a semiautomated natural language processing–based procedure for abstract screening that can substantially decrease the review workload for cancer gene penetrance meta-analyses without compromising the ability to identify relevant studies.

Knowledge Generated

Ten published high-quality meta-analyses were used as the gold standard by which to evaluate performance. This procedure decreases the abstract review workload by 84% and achieves 93% and 99% coverage before and after additional reference review, respectively.

Relevance

This procedure has significant potential to reduce the review time necessary for conducting meta-analyses on gene–cancer penetrance studies and lead to a more rapid translation of the accumulating body of knowledge of prevention and clinical activities.

useful information from unstructured narrative text.^{4,5} NLP algorithms have been widely applied to medical research in recent years for various tasks, including extracting protein–protein interactions,⁶ predicting gene–disease associations from biomedical literature databases,⁷ improving the sensitivity of screening for suicide behaviors among pregnant women from electronic health record systems,⁸ and correlating mammographic imaging features with pathologic findings.⁹ Our group previously used an NLP algorithm to automatically parse breast pathologic reports in both English and Chinese.^{10–12}

Although prior studies have demonstrated that using an NLP-based procedure is a promising approach for reducing the human labor required for an extensive literature review,^{13,14} no NLP-based procedure has been developed and validated in the context of conducting a meta-analysis of penetrance for gene–cancer associations. Here, we introduce and evaluate the performance of a semiautomated NLP-based procedure to partially automate the literature review process required for a meta-analysis of penetrance studies for gene–cancer associations.

MATERIALS AND METHODS

Institutional review board approval was not needed as this study did not involve human participants.

Semiautomated NLP-Based Procedure Design

We compared a traditional procedure of performing a meta-analysis with the semiautomated NLP-based procedure. The traditional procedure is typically based on the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement¹⁵ (Fig 1A). In brief, multiple medical literature databases are searched with predesigned search terms. Human reviewers screen titles and abstracts of these records independently and decide whether a paper is relevant to the research question. Full texts of papers selected at this stage—we refer to these as filtered abstracts (traditional; Fig 1A)—are retrieved and reviewed by human

reviewers. After this review, papers that meet the specified inclusion and exclusion criteria are included in the final analysis—we refer to these as initially included papers (traditional; Fig 1A). In addition, references of the initially included papers (traditional) are reviewed, and additional relevant papers from references are added, resulting in the final set of papers—we refer to these as final included papers (traditional; Fig 1A).

The semiautomated NLP-based procedure developed by our group automates a portion of the traditional procedure (Fig 1B). We use the National Center for Biotechnology Information Entrez Programming Utilities application programming interface (API)¹⁶ to automatically retrieve abstracts on the basis of a PubMed query. A standardized PubMed search query is used with the following structure: “gene name” [TIAB] OR “medical subject headings (MeSH) for that gene” OR “related syndrome name” [TIAB] OR “MeSH for that syndrome”. For example, for a meta-analysis studying *ATM* and breast cancer, the search term is: “*ATM*” [TIAB] OR “Ataxia Telangiectasia Mutated Proteins” [Mesh] OR “Ataxia Telangiectasia” [TIAB] OR “Ataxia Telangiectasia” [Mesh]. Search terms are purposefully designed to be broad, which limits the gene but not the specific cancer. After abstract retrieval, we apply a previously developed NLP-based abstract classifier (code is available at <https://github.com/YujiaBao/PubmedClassifier>), which uses a Support Vector Machine algorithm, to classify abstracts as relevant to penetrance, prevalence, both, or neither (Bao et al, manuscript submitted for publication). We apply this classifier to the extracted abstracts and include only abstracts that are classified as relevant to prevalence, penetrance, or both for additional review. We then apply text mining to the resulting abstracts to identify the presence of the cancer of interest in the abstract. Only abstracts that contain the cancer of interest are included for additional review—we refer to these as filtered abstracts (semiautomated; Fig 1B). Additional steps in the review process are identical to those of the traditional procedure,

FIG 1. Workflows. (A) Traditional versus (B) semiautomated natural language processing (NLP) –based procedure.

including full-text review of the filtered abstracts (semi-automated). Similarly, references of the initially included papers (semiautomated) are reviewed, which results in a final set of papers—the final included papers (semi-automated; Fig 1B).

Published Meta-Analysis Selection

To evaluate the performance of the semiautomated NLP-based procedure, we used 10 high-quality published cancer gene penetrance meta-analyses. These 10 meta-analyses were identified from a systematic literature review conducted by our group as part of the creation of a knowledgebase composed of penetrance studies on gene–cancer associations, which, in turn, is used as the foundation for the clinical decision support tool, ASK2ME.¹⁷ We searched our knowledge database and identified

38 meta-analyses that were relevant to the penetrance of cancer susceptibility genes. Several of the identified studies were on differing gene–cancer associations, but some gene–cancer associations had multiple studies. For those, we chose the study of highest quality and evaluated considering study design, ascertainment criteria, appropriateness of ascertainment adjustment if the study is not population based, and sample size of individual studies included in the meta-analysis. If multiple meta-analyses for the same gene–cancer association were of similar quality, we chose the most recent. Meta-analyses that focused on polymorphisms, grouped syndrome–cancer associations without individual gene–cancer associations, or with multiple gene–multiple cancer combinations were excluded. On the basis of these selection criteria, 10 of 38 identified meta-analyses were included.¹⁸⁻²⁷

As all 10 meta-analyses were performed using the traditional approach—with the exception of the reference review step, as the authors of only one paper stated that they performed the reference review step—we therefore assumed the final included papers (traditional) for each published meta-analysis was the set of studies included by the authors in their quantitative analysis. Basic information for each of these meta-analyses, including author, published year, gene mutation, corresponding cancer, and the number of final included papers (traditional) is listed in Table 1.

Performance Evaluation

As our procedure is based on an abstract classification algorithm, the goal of the procedure is to minimize reviewers' workload by minimizing the number of abstracts a human would need to review but still maximizing the identification of relevant papers. In terms of scientific value, the cost of missing one important paper can be much higher than that of reviewing additional abstracts. Thus, we used both the abstract review workload and the coverage as our two primary outcome measures.

We defined the abstract review workload for the semiautomated NLP-based procedure as the number of abstracts that are classified by the NLP algorithm and text mining as being relevant to penetrance, prevalence, or both on the gene–cancer association of interest. This is also the number of abstracts that required human review using the semiautomated NLP-based procedure (abstracts identified from NLP and text mining; Fig 1B). The abstract review workload of the traditional procedure used by the original meta-analysis papers (abstracts identified from literature database search; Fig 1A) is estimated as follows: If the authors used the PubMed database only and the number of abstracts reviewed was reported in the paper, we used the reported numbers directly. If the authors used the PubMed database only but did not report the exact number of abstracts reviewed, or if they used multiple databases, we created a PubMed search query according to the authors' search strategy description and used the resulting number. Detailed search terms for each of the 10 meta-analyses are

listed in the Appendix. In addition, by estimating the time spent for each abstract review as 3 minutes, we calculated the total time saved for abstract review for each meta-analysis and for all meta-analyses combined for the semiautomated NLP-based procedure compared with the traditional procedure.

We defined coverage as the ability of the semiautomated procedure to identify final included papers (traditional) for each published meta-analysis. We calculated it as the ratio of the number of final included papers (semiautomated) to the number of final included papers (traditional). A coverage of 100% indicates that all final included papers (traditional) in the published meta-analyses were identified by the semiautomated NLP-based procedure. We measured and compared the coverage before and after the reference check. We assumed that the final included papers (traditional) are the gold standard—that is, the positive labels; therefore, the coverage of the semiautomated NLP-based procedure after the reference check can be thought of as the sensitivity (recall)—that is, the number of predicted positives out of the number of positive labels, where the numerator and denominator correspond to the number of final included papers (semiautomated) and the number of final included papers (traditional), respectively.

RESULTS

Summary of the 10 Meta-Analyses

The 10 meta-analyses were published from 2007 to 2016. Among them, four were on breast cancer, three on prostate cancer, two on colorectal cancer, and one on both breast and ovarian cancer (Table 1). Genes studied included *ATM*, *BRCA1*, *BRCA2*, *BRIP1*, *HOXB13*, *MUTYH*, *NBN*, *PALB2*, *RAD50*, and Lynch syndrome genes (*MLH1*, *MSH2*, *MSH6*, *PMS2*, and *EPCAM*). All 10 meta-analyses used PubMed as one of the literature sources. Five of these used PubMed as the sole literature database, whereas the rest used EMBASE, Web of Science, or the Cochrane database in addition to PubMed. A total of 147 studies were included by the authors in their final quantitative analysis. We excluded five (3%) of the 147 studies as they either

TABLE 1. Basic Information Summarizing the 10 Published Meta-Analyses

Meta-Analysis	Gene	Cancer	Date Search Ended	No. of Final Included Papers
Chen et al ¹⁹	<i>BRCA1</i> , <i>BRCA2</i>	Breast, ovarian	02/20/2006	11
Weischer et al ²⁶	<i>CHEK2</i>	Breast	03/01/2007	14
Theodoratou et al ²⁴	<i>MUTYH</i>	Colorectal	07/20/2009	14
Zhang et al ²⁷	<i>NBN</i>	Breast	03/26/2012	11
Shang et al ²³	<i>HOXB13</i>	Prostate	01/03/2013	7
Ryan et al ²²	Lynch syndrome genes	Prostate	05/09/2013	12
Jenkins et al ²⁰	<i>MLH1</i> , <i>MSH2</i>	Colorectal	03/31/2013	4
Aloraifi et al ¹⁸	<i>ATM</i> , <i>CHEK2</i> , <i>PALB2</i> , <i>BRIP1</i> , <i>NBN</i> , <i>RAD50</i>	Breast	06/01/2014	42
Wang et al ²⁵	<i>CHEK2</i>	Prostate	01/27/2015	8
Marabelli et al ²¹	<i>ATM</i>	Breast	02/10/2015	19

could not be identified from PubMed ($n = 2$) or were unpublished ($n = 3$), which left 142 final included papers (traditional) (mean, 14 ± 10 per meta-analysis; range four to 42 studies), with an overall yield rate of 0.8% (range, 0.09% to 58%).

Abstract Review Workload

Overall, the semiautomated NLP-based procedure was associated with a more than six-fold decrease in the abstract review workload (2,774 v 16,941 abstracts; 84% reduction; Table 2). Assuming an average review time of 3 minutes per abstract, this is equivalent to saving 42,501 minutes (approximately 30 full days) of human effort. The semiautomated NLP-based procedure yielded a lower abstract review workload for all 10 meta-analyses (range, 8% to 96% reduction). Three papers were associated with greater than 70% workload reduction, four with 30% to 70% workload reduction, and the remaining three with less than 30% workload reduction. Table 2 lists the comparison of abstract review workload between original meta-analyses and the semiautomated NLP-based procedure.

Coverage

Before the reference review process, overall coverage for the semiautomated NLP-based procedure was 93% (132 of 142; Table 3). Coverage of the 10 meta-analyses ranged from 88% to 100%, with full coverage (100%) for four individual meta-analyses. After adding the reference review, the semiautomated NLP-based procedure identified almost all final included papers (traditional) in these 10 meta-analyses (141 of 142; overall coverage, 99%), only missing one paper. Table 3 lists the coverage for the semiautomated NLP-based procedure before and after reference check.

DISCUSSION

To our knowledge, our study is the first to evaluate a semiautomated NLP-based procedure for abstract

screening that can substantially decrease the review workload for cancer gene penetrance meta-analyses without compromising the ability to identify relevant studies. Using 10 published meta-analyses, we demonstrate that this procedure decreases the abstract review workload by 84% and achieves 93% and 99% coverage before and after reference review, respectively.

With the dramatic reduction in genetic testing costs, it is becoming increasingly important to have up-to-date estimates of gene–cancer mutation penetrance to assist clinical decision making. As penetrance estimates can vary widely by study, even among similar study populations, in many settings integrating different estimates by meta-analysis is an important step in estimating risk. However, as the existing literature on genetic cancer risk is vast and growing daily, it is becoming increasingly difficult to perform such meta-analyses manually.

Since Cohen et al²⁸ first applied an NLP algorithm for a systematic review by developing a voting perceptron–based citation classification system, NLP methods have been used to automate the extraction of information from biomedical literature and automate the screening of titles and abstracts. Gates et al²⁹ developed an online machine learning tool, Abstrackr, that can semiautomate title and abstract screening. Whereas our procedure does not require any input from the users except for the PubMed search query, Abstrackr requires human annotators to provide key positive and negative features for each initial paper. By validating using four published studies, Abstrackr achieved more than 75% sensitivity and a median precision of 26.6%.^{14,29-31} Lerner et al³² reported an automated screening algorithm for updating network meta-analysis on the basis of machine learning and word embeddings. It achieved more than 97% sensitivity on four network meta-analyses, decreasing the abstract review burden by 53% while only missing three of 124 eligible

TABLE 2. Abstract Review Workload Comparison Between Original Meta-Analyses and the NLP Approach

Meta-Analysis	Original (papers)*	NLP Procedures (papers)	Work Saved, %	Time Saved, Minutes†
Chen et al ¹⁹	1,631	1,007	38	1,872
Weischer et al ²⁶	129	69	47	180
Theodoratou et al ²⁴	134	96	28	114
Zhang et al ²⁷	114	46	60	204
Shang et al ²³	12	11	8	3
Ryan et al ²²	381	91	76	870
Jenkins et al ²⁰	4,014	793	80	9,663
Aloraifi et al ¹⁸	10,209	438	96	29,313
Wang et al ²⁵	47	30	36	51
Marabelli et al ²¹	270	193	29	231
Total	16,941	2,774	84	42,501 (≈ 30 days)

Abbreviation: NLP, natural language processing.

*Estimation methods are detailed in Materials and Methods.

†On the basis of an estimated 3 minutes per abstract.

TABLE 3. Coverage of the Semiautomated Natural Language Processing–Based Procedure for Each Meta-Analysis

Meta-Analysis (No. of final papers) and Reference Check	Identified, No.	Missed, No.	Coverage, %
Chen et al ¹⁹ (n = 11)			
Before	11	0	100
After	11	0	100
Weischer et al ²⁶ (n = 14)			
Before	13	1	93
After	14	0	100
Theodoratou et al ²⁴ (n = 14)			
Before	13	1	93
After	14	0	100
Zhang et al ²⁷ (n = 11)			
Before	11	0	100
After	11	0	100
Shang et al ²³ (n = 7)			
Before	7	0	100
After	7	0	100
Ryan et al ²² (n = 12)			
Before	11	1	92
After	12	0	100
Jenkins et al ²⁰ (n = 4)			
Before	4	0	100
After	4	0	100
Aloraifi et al ¹⁸ (n = 42)			
Before	38	4	90
After	42	0	100
Wang et al ²⁵ (n = 8)			
Before	7	1	88
After	8	0	100
Marabelli et al ²¹ (n = 19)			
Before	17	2	89
After	18	1	95
Overall (N = 142)			
Before	132	10	93
After	141	1	99

citations. Tsafnate et al³³ extracted four characteristics from the titles and abstracts—population, exposure, confounders, and outcomes—and trained an algorithm for screening. The best model they tested achieved a recall of 98%. The coverage of our semiautomated NLP-based procedure was 99% and can be thought of as the sensitivity (recall) described in Materials and Methods. Thus, our proposed procedure has performance that is comparable to the algorithms proposed by both Lerner et al and Tsafnate et al. In addition, using keyword matching and a bag-of-words

representation model, Birgmeier et al and Deisseroth et al applied NLP techniques in Mendelian disorders and have developed such tools as Automatic Mendelian Literature Evaluation (AMELIE) and ClinPhen, which can extract and prioritize candidate genes and patient phenotypes from literature and clinical notes.^{34,35} However, none of these is like our procedure, which is designed specifically for gene–cancer penetrance studies and, as shown in this study, leads to a significant workload reduction in this context.

The semiautomated NLP-based procedure that we propose can largely reduce literature review workload for meta-analyses. As our NLP abstract classifier performs well but is not perfect (Bao et al, manuscript submitted for publication), we chose to manually review all abstracts that were classified as penetrance or prevalence, rather than only reviewing abstracts classified as penetrance, to minimize the possibility of missing important studies. Although this additional review increases workload, we estimated the overall workload reduction compared with traditional procedure to be as high as 84%, which suggests that this procedure significantly improves the efficiency of performing penetrance meta-analysis in this context.

Moreover, our procedure achieved 93% coverage of final included papers before adding an additional reference check. In all cases, we missed important studies (n = 10) because of abstract-related issues, including blank abstracts from PubMed (seven of 10), unclearly written abstracts (one of 10), and abstracts missing critical findings (one of 10). Only one paper (one of 10) was missed because our procedure did not catch it. Therefore, it would be useful for academic journals to require an abstract for all papers and for authors to report their findings more completely and precisely in abstracts. Adding a reference review mitigated this limitation and improved coverage to 99%, which suggests that reviewing references remains a crucial step when applying this procedure to other cancer penetrance meta-analysis projects.

This study has several potential limitations. First, our procedure was designed on the basis of searching the PubMed database. In the current study, five final included papers among all 10 meta-analyses were not indexed in PubMed, including three unpublished papers, which suggests that PubMed itself has indexed the majority of cancer penetrance results. Nevertheless, it may be worthwhile to extend the NLP-based procedure to other databases, such as EMBASE³⁶ or Web of Science,³⁷ to have a more comprehensive scanning of the literature. Second, we estimated the time it takes to review one abstract as 3 minutes, which was based on our own experience. In addition to constantly adding new annotations and trying to improve the accuracy of our NLP algorithm, another future research direction is the development of NLP algorithms to classify papers on the basis of the full text, which may improve accuracy and further decrease workload.

In this study, we integrated an NLP algorithm into the literature review workflow for conducting meta-analysis in the context of cancer genetics (Fig 1B). One important lesson we learned is that the strong performance of the NLP algorithm itself does not ensure high efficiency of the entire multistep workflow. In fact, assessing the performance of the entire workflow independently is necessary to identify potential limitations—for example, adding a reference review step is required to achieve good coverage—and to understand the application potential of this informatics tool. The ultimate goal is to develop tools that can automate additional components of the literature review workflow, not just abstract review. There are several challenges that must be addressed to achieve this: automating the reference review process, extracting useful information from tables and figures when applying NLP to full-text papers, and

correctly identifying penetrance information from the multiple quantities reported in the paper. Addressing these challenges is the subject of ongoing and future work. Nonetheless, the current study provides a valuable framework for the development, evaluation, and translation of such multistep informatics tools to practice.

In conclusion, we have evaluated the performance of a semiautomated NLP-based procedure that partially automates the abstract screening process for gene–cancer penetrance studies. We have demonstrated that this procedure can substantially decrease the abstract review workload and achieve comprehensive coverage in identifying relevant studies. It has significant potential to reduce the review time necessary for conducting meta-analyses on gene–cancer penetrance studies.

AFFILIATIONS

¹Massachusetts General Hospital, Boston, MA

²Massachusetts Institute of Technology, Boston, MA

³Harvard TH Chan School of Public Health, Boston, MA

⁴Dana-Farber Cancer Institute, Boston, MA

⁵Harvard Medical School, Boston, MA

CORRESPONDING AUTHOR

Danielle Braun, PhD, Department of Biostatistics, Harvard TH Chan School of Public Health, Department of Data Sciences, Dana-Farber Cancer Institute, 677 Huntington Ave, SPH 4th Floor, Boston, MA 02115; e-mail: dbraun@mail.harvard.edu.

EQUAL CONTRIBUTION

Z.D. and K.Y. contributed equally to this work.

PRIOR PRESENTATION

Presented at the 71st Annual Meeting of the Massachusetts General Hospital Scientific Advisory Committee, Boston, MA, April 3–4, 2019, and the 2019 Dana-Farber/Harvard Cancer Center Cancer Genetics Retreat, Boston, MA, June 13, 2019.

SUPPORT

Supported by National Cancer Institute Grants No. 5T32-CA009337-32 and 4P30-CA006516-51 and the Koch Institute/Dana-Farber Cancer Institute/Harvard Cancer Center Bridge Project (Footbridge).

AUTHOR CONTRIBUTIONS

Conception and design: Zhengyi Deng, Kanhua Yin, Victor Diego Armengol, Regina Barzilay, Giovanni Parmigiani, Danielle Braun, Kevin S. Hughes

Financial support: Giovanni Parmigiani, Kevin S. Hughes

Administrative support: Kevin S. Hughes

Provision of study materials or patients: Kevin S. Hughes

Collection and assembly of data: Zhengyi Deng, Kanhua Yin, Yujia Bao, Victor Diego Armengol, Cathy Wang, Ankur Tiwari, Kevin S. Hughes

Data analysis and interpretation: Zhengyi Deng, Kanhua Yin, Victor Diego Armengol, Ankur Tiwari, Regina Barzilay, Danielle Braun, Kevin S. Hughes

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/ifo.

Regina Barzilay

Honoraria: Merck

Consulting or Advisory Role: Janssen

Research Funding: Bayer, Amgen

Giovanni Parmigiani

Leadership: Phaeno Biotech

Stock and Other Ownership Interests: HRA Health

Consulting or Advisory Role: Biogen, Konica Minolta

Patents, Royalties, Other Intellectual Property: Patent: Genetic Alterations in Malignant Gliomas, copyright: BayesMendel Software (Inst)

Expert Testimony: Natera

Travel, Accommodations, Expenses: Konica Minolta

Kevin S. Hughes

Stock and Other Ownership Interests: Hughes riskApps

Honoraria: Focal Therapeutics, 23andMe, Hologic

Consulting or Advisory Role: Health Beacons

No other potential conflicts of interest were reported.

REFERENCES

- Turnbull C, Sud A, Houlston RS: Cancer genetics, precision prevention and a call to action. *Nat Genet* 50:1212–1218, 2018 [Erratum: *Nat Genet* 51:196, 2019]
- Borah R, Brown AW, Capers PL, et al: Analysis of the time and workers needed to conduct systematic reviews of medical interventions using data from the PROSPERO registry. *BMJ Open* 7:e012545, 2017

3. Pham B, Bagheri E, Rios P, et al: Improving the conduct of systematic reviews: A process mining perspective. *J Clin Epidemiol* 103:101-111, 2018
4. Selby LV, Narain WR, Russo A, et al: Autonomous detection, grading, and reporting of postoperative complications using natural language processing. *Surgery* 164:1300-1305, 2018
5. Zhong Q-Y, Mittal LP, Nathan MD, et al: Use of natural language processing in electronic medical records to identify pregnant women with suicidal behavior: Towards a solution to the complex classification problem. *Eur J Epidemiol* 34:153-162, 2019
6. Yu K, Lung P-Y, Zhao T, et al: Automatic extraction of protein-protein interactions using grammatical relationship graph. *BMC Med Inform Decis Mak* 18:42, 2018 (suppl 2)
7. Bouaziz J, Mashiach R, Cohen S, et al: How artificial intelligence can improve our understanding of the genes associated with endometriosis: Natural language processing of the PubMed database. *BioMed Res Int* 2018:6217812, 2018
8. Zhong Q-Y, Karlson EW, Gelaye B, et al: Screening pregnant women for suicidal behavior in electronic medical records: Diagnostic codes vs. clinical notes processed by natural language processing. *BMC Med Inform Decis Mak* 18:30, 2018
9. Patel TA, Puppala M, Ogunti RO, et al: Correlating mammographic and pathologic findings in clinical decision support using natural language processing and data mining methods. *Cancer* 123:114-121, 2017
10. Tang R, Ouyang L, Li C, et al: Machine learning to parse breast pathology reports in Chinese. *Breast Cancer Res Treat* 169:243-250, 2018
11. Yala A, Barzilay R, Salama L, et al: Using machine learning to parse breast pathology reports. *Breast Cancer Res Treat* 161:203-211, 2017
12. Buckley JM, Coopey SB, Sharko J, et al: The feasibility of using natural language processing to extract clinical information from breast pathology reports. *J Pathol Inform* 3:23, 2012
13. Miwa M, Thomas J, O'Mara-Eves A, et al: Reducing systematic review workload through certainty-based screening. *J Biomed Inform* 51:242-253, 2014
14. Wallace BC, Trikalinos TA, Lau J, et al: Semi-automated screening of biomedical citations for systematic reviews. *BMC Bioinformatics* 11:55, 2010
15. Moher D, Liberati A, Tetzlaff J, et al: Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Med* 6:e1000097, 2009
16. Sayers E: E-utilities quick start, in *Entrez Programming Utilities Help*. Bethesda, MD, National Center for Biotechnology Information, 2010. <https://www.ncbi.nlm.nih.gov/books/NBK25500/>
17. Braun D, Yang J, Griffin M, et al: A clinical decision support tool to predict cancer risk for commonly tested cancer-related germline mutations. *J Genet Couns* 27:1187-1199, 2018
18. Aloraifi F, McCartan D, McDevitt T, et al: Protein-truncating variants in moderate-risk breast cancer susceptibility genes: A meta-analysis of high-risk case-control screening studies. *Cancer Genet* 208:455-463, 2015
19. Chen S, Parmigiani G: Meta-analysis of *BRCA1* and *BRCA2* penetrance. *J Clin Oncol* 25:1329-1333, 2007
20. Jenkins MA, Dowdy JG, Ait Ouakrim D, et al: Short-term risk of colorectal cancer in individuals with lynch syndrome: A meta-analysis. *J Clin Oncol* 33:326-331, 2015
21. Marabelli M, Cheng S-C, Parmigiani G: Penetrance of ATM gene mutations in breast cancer: A meta-analysis of different measures of risk. *Genet Epidemiol* 40:425-431, 2016
22. Ryan S, Jenkins MA, Win AK: Risk of prostate cancer in Lynch syndrome: A systematic review and meta-analysis. *Cancer Epidemiol Biomarkers Prev* 23:437-449, 2014
23. Shang Z, Zhu S, Zhang H, et al: Germline homeobox B13 (HoxB13) G84E mutation and prostate cancer risk in European descendants: A meta-analysis of 24,213 cases and 73,631 controls. *Eur Urol* 64:173-176, 2013
24. Theodoratou E, Campbell H, Tenesa A, et al: A large-scale meta-analysis to refine colorectal cancer risk estimates associated with MUTYH variants. *Br J Cancer* 103:1875-1884, 2010
25. Wang Y, Dai B, Ye D: CHEK2 mutation and risk of prostate cancer: A systematic review and meta-analysis. *Int J Clin Exp Med* 8:15708-15715, 2015
26. Weischer M, Bojesen SE, Ellervik C, et al: *CHEK2**1100delC genotyping for clinical assessment of breast cancer risk: Meta-analyses of 26,000 patient cases and 27,000 controls. *J Clin Oncol* 26:542-548, 2008
27. Zhang Z-H, Yang L-S, Huang F, et al: Current evidence on the relationship between two polymorphisms in the NBS1 gene and breast cancer risk: A meta-analysis. *Asian Pac J Cancer Prev* 13:5375-5379, 2012
28. Cohen AM, Hersh WR, Peterson K, et al: Reducing workload in systematic review preparation using automated citation classification. *J Am Med Inform Assoc* 13:206-219, 2006
29. Gates A, Johnson C, Hartling L: Technology-assisted title and abstract screening for systematic reviews: A retrospective evaluation of the Abstrackr machine learning tool. *Syst Rev* 7:45, 2018
30. Frunza O, Inkpen D, Matwin S, et al: Exploiting the systematic review protocol for classification of medical abstracts. *Artif Intell Med* 51:17-25, 2011
31. Rathbone J, Hoffmann T, Glasziou P: Faster title and abstract screening? Evaluating Abstrackr, a semi-automated online screening program for systematic reviewers. *Syst Rev* 4:80, 2015
32. Lerner I, Créquit P, Ravaut P, et al: Automatic screening using word embeddings achieved high sensitivity and workload reduction for updating living network meta-analyses. *J Clin Epidemiol* 108:86-94, 2019
33. Tsafnat G, Glasziou P, Karystianis G, et al: Automated screening of research studies for systematic reviews using study characteristics. *Syst Rev* 7:64, 2018
34. Birgmeier J, Haeussler M, Deisseroth CA, et al: AMELIE accelerates Mendelian patient diagnosis directly from the primary literature. *bioRxiv* doi:[10.1101/171322](https://doi.org/10.1101/171322) [Epub ahead of print on August 2, 2017]
35. Deisseroth CA, Birgmeier J, Bodle EE, et al: ClinPhen extracts and prioritizes patient phenotypes directly from medical records to expedite genetic disease diagnosis. *Genet Med* 21:1585-1593, 2019
36. Embase: Life science solutions. <http://embase.com>
37. Clarivate Analytics: Web of Science databases. <https://clarivate.com/products/web-of-science/databases/>

APPENDIX**Chen et al**

PubMed search query reported in the publication: Yes

Number of PubMed papers reported in the publication: No

We reran the PubMed query by using the reported search query and search end date:

Search query: ("risk" OR "penetrance") AND ("breast cancer" OR "ovarian cancer") AND ("BRCA1" OR "BRCA2")

Search end date: 02/20/2006

Number of papers found: 1,631

Weischer et al

PubMed search query reported in the publication: No

Number of PubMed papers reported in the publication: No

We created a PubMed query according to the authors' search strategy description:

Search term: ("CHEK2" OR "CHEK2*1100delC" OR "CHK2") AND ("breast cancer")

Search end date: 03/01/2007

Number of papers found: 129

Theodoratou et al

PubMed search query reported in the publication: Yes

Number of PubMed papers reported in the publication: No

We reran the PubMed query by using the reported search query and search end date:

Search query: ("MYH" OR "MUTYH") AND "colorectal cancer"

Search end date: 07/20/2009

Number of papers found: 134

Zhang et al

PubMed search query reported in the publication: Yes

Number of PubMed papers reported in the publication: No

We reran the PubMed query by using the reported search query and search end date:

Search query: ("NBS1" or "NBN" or "657del5" or "1171v" or "511A>G" or "rs61754966" or "11e171val") and ("cancer" or "carcinoma" or "tumor" or "neoplasms") and ("breast" or "mammary")

Search end date: 03/26/2012

Number of papers found: 114

Shang et al

PubMed search query reported in the publication: Yes

Number of PubMed papers reported in the publication: No

We reran the PubMed query by using the reported search query and search end date:

Search query: ("Prostatic Neoplasms"[MeSH] OR "Prostate carcinoma" OR "Prostate cancer" OR "Prostatic carcinoma" OR "Prostatic cancer") AND ("Polymorphism, Single Nucleotide"[MeSH] OR "Polymorphisms" OR "SNP" OR "mutation"[MeSH]) AND ("HOXB13 protein, human"[Supplementary Concept] OR "HOXB13" OR "rs138213197" OR "G84E")

Search end date: 01/03/2013

Number of papers found: 12

Ryan et al

PubMed search query reported in the publication: Yes

Number of PubMed papers reported in the publication: Yes

We used the numbers reported in the publication directly:

Number of papers reported in the publication: 381

Jenkins et al

PubMed search query reported in the publication: No

Number of PubMed papers reported in the publication: No

We created the PubMed query according to the authors' search strategy description:

Search query: ("colon" [TIAB] or "colorectal" [TIAB]) and ("cancer" [TIAB] or "neoplasm" [TIAB] or "tumor" [TIAB]) and ("Lynch syndrome" [TIAB] or "HNPCC" [TIAB] or "mismatch repair" [TIAB] or "MLH1" [TIAB] or "MSH2" [TIAB])

Search end date: 03/31/2013

Number of papers found: 4,014

Aloraifi et al

PubMed search query reported in the publication: Yes

Number of PubMed papers reported in the publication: Yes

We used the numbers reported in the publication directly:

Number of papers reported in the publication: 10,209

Wang et al

PubMed search query reported in the publication: No

Number of PubMed papers reported in the publication: No

We created the PubMed query according to the authors' search strategy description:

Search term: ("prostate cancer" OR "prostate carcinoma" OR "prostate neoplasm") AND ("CHEK2")

Search end date: 01/27/2015

Number of papers found: 47

Marabelli et al

PubMed search query reported in the publication: Yes

Number of PubMed papers reported in the publication: No

We reran the PubMed query by using the reported search query and search end date:

Search query: ("ATM") AND ("penetrance" OR "risk") AND ("breast")

Search end date: 02/10/2015

Number of papers found: 270

NOTE. The search end date used was the date reported in the publication or, if not reported, the publication date of the most recent final included paper in the meta-analysis.