

# Evolution of colorectal cancer screening research in the past 25 years: text-mining analysis of publication trends and topics

Shelly Soffer<sup>1</sup>, Eyal Klang, Noam Tau, Roni Zemet, Shomron Ben-Horin, Yiftach Barash and Uri Kopylov<sup>2</sup>

## Abstract

**Background:** There is a growing research effort in the field of colorectal cancer (CRC) screening, with varying topics and shifting research foci over the years. The aim of this study was to apply a text-mining technique to evaluate trends in publications for CRC screening in the last 25 years.

**Methods:** We retrieved MEDLINE/PubMed datasets from 1992–2017. We selected keywords from Medical Subject Headings to include CRC screening related publications. For each article, we extracted the following data: title, journal, publication date, abstract, article type, citation frequency, and country of origin. Articles were categorized into topics using word combination and title match technique.

**Results:** In 1992–2017, 14,119 CRC screening related papers were published. The US had the highest number of papers ( $n=4824$ ) and China had the highest growth rate in publications. Overall, the most researched topic was “screening and surveillance programs” (38%). The topics of “quality assurance” ( $r=0.87$ ) and “racial disparities” ( $r=0.91$ ) have gained increased research attention over the years. In total, 11 of the 20 most cited articles in the field were published in *The New England Journal of Medicine*.

**Conclusion:** The number of publications devoted to CRC screening has grown, with high-quality research reaching top-tier journals. A surge in the number of publications has been increasing in countries previously less involved in research in the field. Screening programs remain the most researched topic, and quality indicators is attracting a growing attention. Text-mining analysis of CRC screening research contributes to an understanding of publication trends and topics and can point to the need for potential future investigations.

**Keywords:** cancer screening, colorectal cancer, text mining

Received: 28 March 2020; revised manuscript accepted: 15 June 2020.

## Introduction

Colorectal cancer (CRC) is a world-wide health-care problem with high morbidity and mortality. It is the second leading cause of cancer-related death throughout the world.<sup>1</sup> The annual expenditures associated with CRC are substantial. They are estimated at 14 billion US dollars per year in the US alone.<sup>2</sup>

Several factors make CRC applicable for population screening. These include high incidence, a

prolonged disease development course, and effective endoscopic treatment options in premalignant stages.<sup>3</sup> CRC screening has therefore been adopted on a wide scale and has been incorporated into international guidelines.<sup>4,5</sup> CRC screening has significantly reduced the incidence and mortality rates of the disease in the last two decades.<sup>6,7</sup>

Extensive research in the field of CRC screening has been published. These studies assessed various

*Ther Adv Gastroenterol*

2020, Vol. 13: 1–12

DOI: 10.1177/  
1756284820941153

© The Author(s), 2020.  
Article reuse guidelines:  
sagepub.com/journals-  
permissions

Correspondence to:

**Shelly Soffer**  
Department of Diagnostic  
Imaging, Sheba Medical  
Center, Tel Hashomer,  
5265601, Israel

DeepVision Lab, Sheba  
Medical Center, Tel  
Hashomer, 5265601, Israel  
[soffer.shelly@gmail.com](mailto:soffer.shelly@gmail.com)

**Shelly Soffer**  
Department of Diagnostic  
Imaging, Sheba Medical  
Center, Tel Hashomer,  
Israel

Sackler Medical School,  
Tel Aviv University, Tel  
Aviv, Israel DeepVision  
Lab, Sheba Medical  
Center, Tel Hashomer,  
Israel

**Eyal Klang**  
Department of Diagnostic  
Imaging, Sheba Medical  
Center, Tel Hashomer,  
Israel

Sackler Medical School,  
Tel Aviv University, Tel  
Aviv, Israel DeepVision  
Lab, Sheba Medical  
Center, Tel Hashomer,  
Israel

**Noam Tau**  
Department of Diagnostic  
Imaging, Sheba Medical  
Center, Tel Hashomer,  
Israel, and Sackler Medical  
School, Tel Aviv University,  
Tel Aviv, Israel

**Roni Zemet**  
Department of Obstetrics  
and Gynecology, Sheba  
Medical Center, Tel  
Hashomer, Israel, and  
Sackler Medical School,  
Tel Aviv University, Tel  
Aviv, Israel

**Shomron Ben-Horin**  
**Uri Kopylov**  
Department of  
Gastroenterology, Sheba  
Medical Center, Tel  
Hashomer, Israel

Sackler Medical School,  
Tel Aviv University, Tel  
Aviv, Israel

**Yiftach Barash**

Department of Diagnostic Imaging, Sheba Medical Center, Tel Hashomer, Israel

Sackler Medical School, Tel Aviv University, Tel Aviv, Israel DeepVision Lab, Sheba Medical Center, Tel Hashomer, Israel

aspects, including available screening tools (endoscopic, fecal, radiologic, and blood tests), screening programs, high-risk populations, and cost-effectiveness. Manual summarization of this large amount of research data is both infeasible and impractical. When examining the literature in other medical fields, we note that endeavors have been made to identify research trends. These studies have been narrowed down to a defined number of cited articles,<sup>8</sup> to specific journals,<sup>9</sup> or to a limited number of years.<sup>10</sup>

Current computational power and machine learning advancements have prompted a technique termed “text-mining.” This technique extracts information from texts using computational statistical methods.<sup>11</sup> Text-mining can be applied to identify trends and to investigate the dynamics in a research field.<sup>12–15</sup>

The aim of our study was to apply a text-mining technique to evaluate published literature for CRC screening in the last 25 years. We performed trend analysis to discover patterns in CRC screening publications.

### Methods

An institutional review board (IRB) approval was granted for this study. Informed consent was waived by the IRB committee.

#### Search strategy

The US National Library of Medicine (NLM) produces an annual version of MEDLINE/PubMed data which is freely available for download.<sup>16–18</sup> We used the 2018 MEDLINE/PubMed baseline dataset in this study. We retrieved all available MEDLINE/PubMed annual datasets from 1992 to the end of 2017 (25 years). Data lock and citation retrieval were performed on 1 August 2019.

#### Data processing

The data processing and results visualization were written in Python (version 3.6.5, 64 bits). We used the open-source Pandas library (version 0.24.2) for data handling, open-source Geopandas (version 0.4.1) for geographical visualization, open-source SCIPY (version 1.3.0), open-source NLTK (version 3.4.4) for text handling, and open-source Matplotlib (version 3.1.0) for results visualization.

For text-mining, each title, abstract, and the first author’s affiliation were tokenized. All punctuation and double spaces were removed, and each word became a single entry in a list.

#### Inclusion and exclusion criteria

For creating a subset of papers which were relevant to our desired topic, we used Medical Subject Headings (MeSH). MeSH terms are used as a supervised glossary for searching in the PubMed database.<sup>13</sup> The following keywords were selected from MeSH to create the subset of articles relating to the colon and rectum: “colorectal,” “colon,” “rectal,” “rectum,” “colonic,” and “CRC.” These terms were matched to the tokenized title list and a subset of records was retrieved.

We then included papers which had one of the terms “cancer,” “carcinoma,” “adenocarcinoma,” “adenoma,” “polyp,” or “mass” in the abstract, and also one of the following terms: “screening,” “surveillance,” or “screen.” Abstracts shorter than 50 words were excluded.

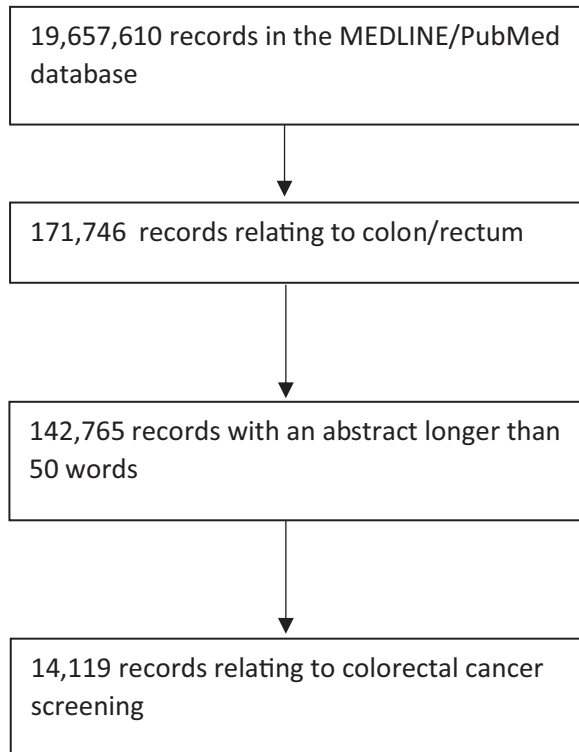
#### Data extraction

The following data was extracted from each of the included articles: PubMed unique article ID (PMID), title, journal, publication date (year and month), abstract text, article type (e.g. review, randomized control trial), article language, and authors (including the first author’s affiliation, if available). We then used a free-for-use application provided by the National Center for Biotechnology Information (NCBI) to retrieve the number of times each article was cited, based on its PMID.<sup>19</sup>

The first author’s country was retrieved from the affiliation data, if available. The first author’s affiliation was compared with a country list extracted from the Geopandas library. We normalized the number of publications and the number of citations for each country according to its population by extracting the yearly population size of each country from the World Bank Catalog.<sup>20</sup>

#### Topic modeling

All included studies were divided into topics using the following methodology: each study’s title was analyzed after omitting stop words such as “the,” “a,” “an,” and “in”, which are detailed



**Figure 1.** Flow diagram of included studies using MEDLINE/PubMed search.

10 topics. Each study record was then matched to one of the 10 topics by comparing the words in the title with the topic list.

#### *Data analysis*

We used Pearson correlation to evaluate normalized trends in topics for intervals of 5 years. We used univariable linear regression to evaluate country growth rate trends. The slope statistical significance is presented through the  $p$ -value which is the regression output (open-source SCIPY).

For article type and country analysis, the citation rate was calculated by dividing the overall number of times articles were cited by the overall number of publications.

#### **Results**

A total of 19,657,610 records were retrieved from the NLM database between 1992 and 2017. Of these, 14,119 publications were related to CRC screening. A flow diagram of the search is provided in Figure 1. Almost all papers were in the English language (93.5%).

in NLTK version 3.4.4 StopWords corpus. The 1000 most frequent two-word combinations in all titles were listed in descending order of frequency. A gastroenterologist specialist physician (KU) defined 10 topics in the field of CRC screening. Topics included: Screening and surveillance programs, risk stratification, non-invasive screening, epidemiology, inflammatory bowel disease (IBD) screening, quality assurance, racial disparities, treatment, quality of life, and cost-effectiveness.

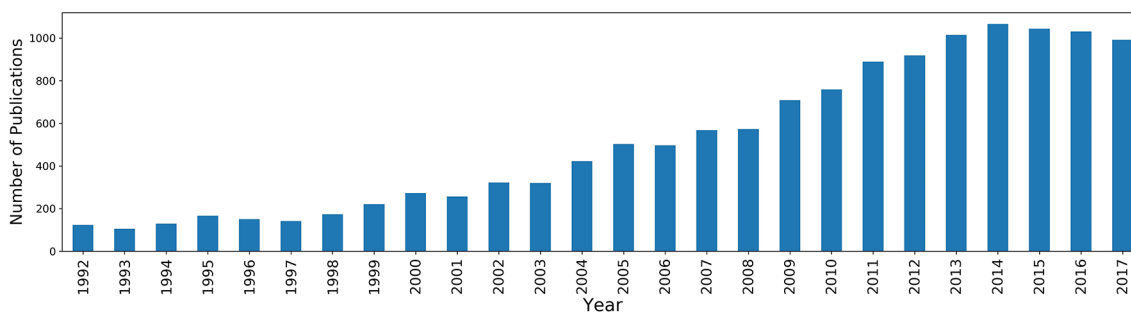
Each word combination in the list was manually labeled as either non-specific or related to 1 of the

#### *Time trend analysis*

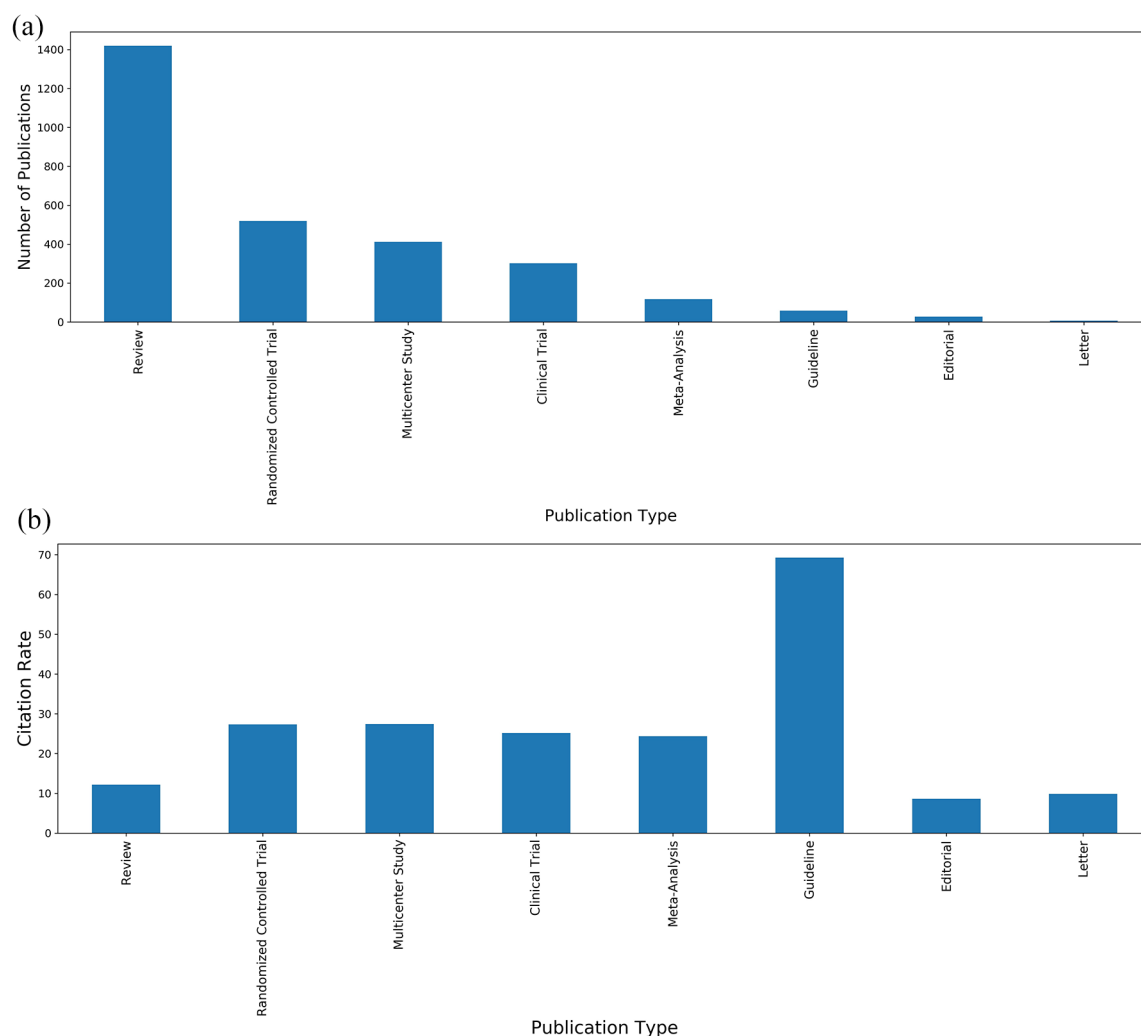
The number of annual publications relating to CRC screening increased between 1992 and 2014 (Figure 2), with a slight decline since 2014. The overall number of annual publications increased from 124 publications in 1992 to 992 publications in 2017.

#### *Article type analysis*

MEDLINE/PubMed article type was specified for 2862/14,119 (20.3%) papers. Among those,



**Figure 2.** Trends in the number of colorectal cancer screening research from 1992 to 2017.



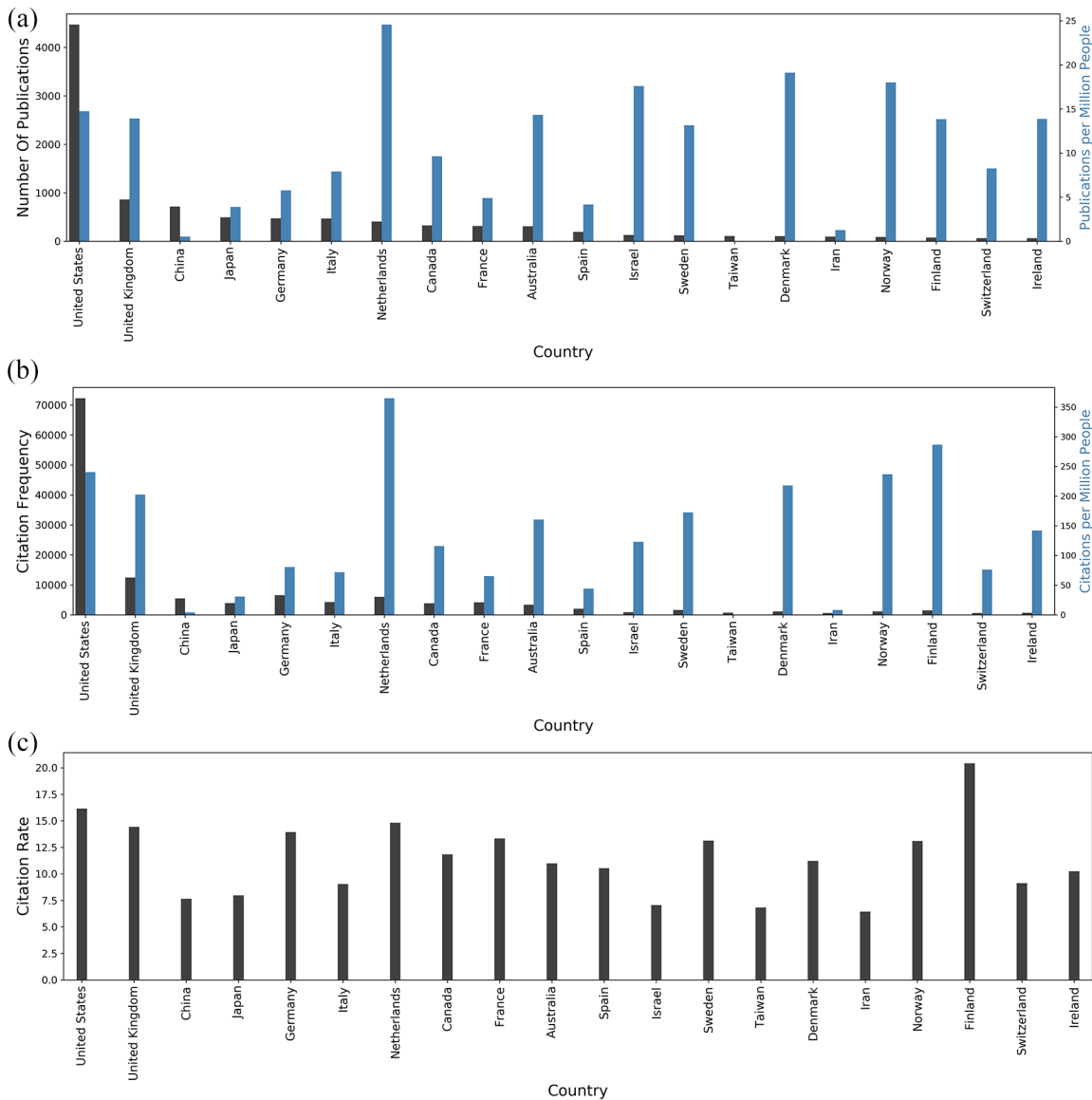
**Figure 3.** Distribution of colorectal cancer screening by article type, indicating (a) publication volume and (b) citation rate (i.e. number of citations per number of publications).

1429/2862 (50.0%) were review articles, 519/2862 (18.1%) were randomized controlled trials, and 412/2862 (14.4%) were multi-center studies. The article type with the highest citation rate (number of citations/number of publications) was guideline papers (69.2) followed by multi-center studies (27.4) and randomized controlled trials (27.3). Figure 3 shows the distribution of article types and their corresponding citation rate.

#### Country analysis

Publications on CRC screening originated from 97 countries, mainly from North America and Europe. The US had the highest number of papers ( $n=4824$ ), followed by the UK ( $n=927$ ), and

China ( $n=848$ ). After normalizing to population size, the Netherlands had the largest number of publications per million people ( $n=27$ ), followed by Denmark ( $n=23$ ), and Israel ( $n=20$ ). The countries with the highest citation number were the US ( $n=73,638$ ), the UK, ( $n=12,678$ ) and Germany ( $n=6642$ ). When normalizing the number of citations to the population size, the Netherlands had the largest number of citations per million people ( $n=373$ ), followed by Finland ( $n=288$ ), and the US ( $n=250$ ). The countries with the highest citation rate (number of citations/number of publications) were Finland (18), US (15), and UK (13). Figure 4 presents the distributions of publications, citations, and citation rates among the 20 countries with the highest number of publications.



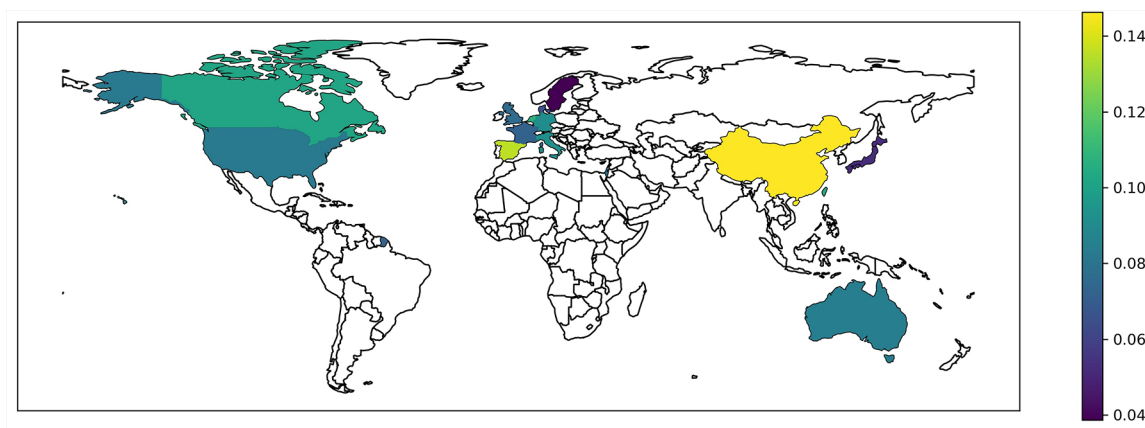
**Figure 4.** Distribution of colorectal cancer screening by country, indicating (a) publication volume, (b) citation frequency, and (c) citation rate (i.e. number of citations per number of publications). The left axis presents absolute numbers and right axis presents normalization by country population size.

Figure 5 shows the growth rate in number of annual publications in respect to the country of origin, China (0.14,  $p < 0.001$ ), Spain (0.13,  $p < 0.001$ ), and Taiwan (0.10,  $p < 0.001$ ) had the highest growth rate over time.

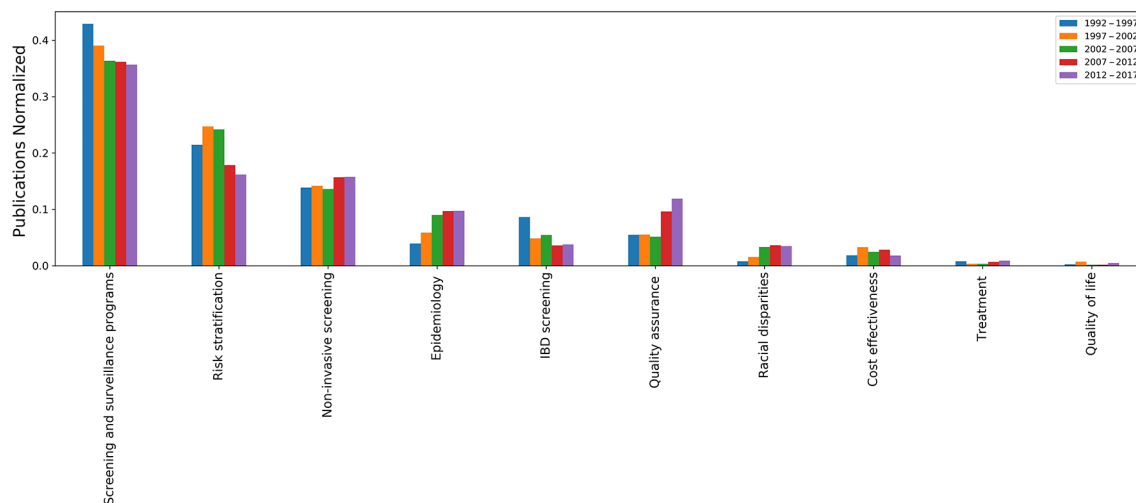
#### Topic analysis

The most researched topic is “screening and surveillance programs” (38%). Yet, a continuous decrease in research attention is shown for this topic over the past 25 years ( $r = -0.91$ ,  $p = 0.035$ ).

Another area with a constant research interest is “non-invasive screening” (14%;  $r = 0.85$ ,  $p = 0.93$ ). The topic of “risk stratification” shows a non-linear correlation in time ( $r = -0.72$ ,  $p = 0.63$ ) with an increase until 1997 (24%) followed by a constant decrease until the end of 2017 (16%). The topic of “quality assurance” was steady until 2007 and has gained increased research attention in the past decade ( $r = 0.87$ ,  $p = 0.052$ ). Research interest in “racial disparities” increased fivefold from 1992–2007 (0.7–3.3%) and has remained stable since then



**Figure 5.** World map indicating colorectal cancer screening publication growth rate by country. The color index represents the calculated growth rate. Countries with less than 100 overall publications were omitted from the growth rate analysis (demonstrated by the white color).



**Figure 6.** Published article topic popularity and trends over time.

( $r=0.91$ ,  $p=0.032$ ). Quality assurance was the most trending topic in the last 5 years (years 2012–2017). The distribution of all assigned topics according to the year of publication is presented in Figure 6.

#### Most frequently cited articles

Table 1 lists the top 20 most cited articles published on CRC screening in the past 25 years. The mean number of citations per article is 98. The top 20 most cited articles were published in five journals, with the greatest number in *The New England Journal of Medicine* ( $n=11$ ) and have originated mainly from the US ( $n=13$ ).

#### Discussion

In our study, we applied a text-mining approach to present an overview of 14,119 CRC screening publications over the past 25 years.

The number of CRC screening publications has increased over the years. In 2017, the number of published papers in CRC screening was eight times greater than in 1992. This increase in the number of published articles coincides with the general trend of increased global publications in the medical field.<sup>41</sup> Several factors can be attributed to this particular trend in CRC screening publications. This growth can be a result of the expansion of CRC screening programs and the

**Table 1.** The top 20 most frequently cited articles published on colorectal cancer screening in the past 25 years.

| Article rank | Study   | Title  | Publication year | Number of times cited | Country        | Journal   |
|--------------|---|--|------------------|-----------------------|----------------|---|
| 1            | U.S. Preventive Services Task Force <sup>21</sup> | Screening for colorectal cancer: U.S. Preventive Services Task Force recommendation statement.   | 2008             | 560                   | United States  | <i>Annals of Internal Medicine</i>              |
| 2            | Mandel <i>et al.</i> <sup>22</sup>                | Reducing mortality from colorectal cancer by screening for fecal occult blood. Minnesota Colon Cancer Control Study.   | 1993             | 554                   | United States  | <i>The New England Journal of Medicine</i>      |
| 3            | Levin <i>et al.</i> <sup>23</sup>                 | Screening and surveillance for the early detection of colorectal cancer and adenomatous polyps, 2008: a joint guideline from the American Cancer Society, the US Multi-Society Task Force on Colorectal Cancer, and the American College of Radiology. | 2008             | 541                   | United States  | <i>Gastroenterology</i>                         |
| 4            | Edwards <i>et al.</i> <sup>24</sup>               | Annual report to the nation on the status of cancer, 1975–2006, featuring colorectal cancer trends and impact of interventions (risk factors, screening, and treatment) to reduce future rates.  | 2009             | 526                   | United States  | <i>Cancer</i>                                   |
| 5            | Hardcastle <i>et al.</i> <sup>25</sup>            | Randomised controlled trial of faecal-occult-blood screening for colorectal cancer.  | 1996             | 497                   | United Kingdom | <i>Lancet (London, England)</i>                 |
| 6            | Zauber <i>et al.</i> <sup>26</sup>                | Colonoscopic polypectomy and long-term prevention of colorectal-cancer deaths.   | 2012             | 470                   | United States  | <i>The New England Journal of Medicine</i>      |
| 7            | Kronborg <i>et al.</i> <sup>27</sup>              | Randomised study of screening for colorectal cancer with faecal-occult-blood test.   | 1996             | 437                   | Denmark        | <i>Lancet (London, England)</i>                 |
| 8            | Winawer <i>et al.</i> <sup>28</sup>               | Colorectal cancer screening and surveillance: clinical guidelines and rationale-update based on new evidence.  | 2003             | 416                   | United States  | <i>Gastroenterology</i>                         |
| 9            | Rex <i>et al.</i> <sup>29</sup>                   | American College of Gastroenterology guidelines for colorectal cancer screening 2009.  | 2009             | 403                   | United States  | <i>The American Journal of Gastroenterology</i> |
| 10           | Atkin <i>et al.</i> <sup>30</sup>                 | Once-only flexible sigmoidoscopy screening in prevention of colorectal cancer: a multicentre randomised controlled trial.  | 2010             | 385                   | United Kingdom | <i>Lancet (London, England)</i>                 |
| 11           | Lieberman <i>et al.</i> <sup>31</sup>             | Use of colonoscopy to screen asymptomatic adults for colorectal cancer. Veterans Affairs Cooperative Study Group 380.  | 2000             | 304                   | United States  | <i>The New England Journal of Medicine</i>      |

(Continued)

Table 1. (Continued)

| Article rank | Study                                 | Title   | Publication year | Number of times cited | Country       | Journal                                    |
|--------------|---------------------------------------|---|------------------|-----------------------|---------------|--|
| 12           | Pickhardt <i>et al.</i> <sup>32</sup> | Computed tomographic virtual colonoscopy to screen for colorectal neoplasia in asymptomatic adults.                     | 2003             | 303                   | United States | <i>The New England Journal of Medicine</i> |
| 13           | Mandel <i>et al.</i> <sup>33</sup>    | The effect of fecal occult-blood screening on the incidence of colorectal cancer.                                       | 2000             | 300                   | United States | <i>The New England Journal of Medicine</i> |
| 14           | Baxter <i>et al.</i> <sup>34</sup>    | Association of colonoscopy and death from colorectal cancer.  | 2008             | 292                   | Canada        | <i>Annals of Internal Medicine</i>         |
| 15           | Hampel <i>et al.</i> <sup>35</sup>    | Screening for the Lynch syndrome (hereditary nonpolyposis colorectal cancer).   | 2005             | 287                   | United States | <i>The New England Journal of Medicine</i> |
| 16           | Kaminski <i>et al.</i> <sup>36</sup>  | Quality indicators for colonoscopy and the risk of interval cancer.   | 2010             | 282                   | Poland        | <i>The New England Journal of Medicine</i> |
| 17           | Nishihara <i>et al.</i> <sup>37</sup> | Long-term colorectal-cancer incidence and mortality after lower endoscopy.  | 2013             | 260                   | United States | <i>The New England Journal of Medicine</i> |
| 18           | Järvinen <i>et al.</i> <sup>38</sup>  | Controlled 15-year trial on screening for colorectal cancer in families with hereditary nonpolyposis colorectal cancer. | 2000             | 248                   | Finland       | <i>Gastroenterology</i>                    |
| 19           | Selby <i>et al.</i> <sup>39</sup>     | A case-control study of screening sigmoidoscopy and mortality from colorectal cancer.                                   | 1992             | 235                   | United States | <i>The New England Journal of Medicine</i> |
| 20           | Whitlock <i>et al.</i> <sup>40</sup>  | Screening for colorectal cancer: a targeted, updated systematic review for the U.S. Preventive Services Task Force.     | 2008             | 229                   | United States | <i>Annals of Internal Medicine</i>         |

implementation of population-based programs.<sup>42,43</sup> The awareness of CRC screening is consistent with the worldwide endeavors that have focused on cancer prevention.<sup>44–46</sup> Furthermore, the increase in CRC screening publications could be linked to the rise in CRC incidence, particularly in countries in Eastern Europe, Asia, and South America.<sup>47</sup> Another possible factor is the emergence of new technologies. For example, in 1994 computed tomographic colonography was introduced by Vining *et al.* and was followed by a large number of related publications.<sup>48</sup>

Research in the field of CRC screening started with several seminal publications 25 years ago.<sup>3,22,25</sup> These papers established the understanding that CRC screening can effectively

reduce CRC mortality rate. They showed that colonoscopic polypectomy resulted in a lower-than-expected incidence of CRC and that annual fecal occult-blood test decreased mortality from CRC. These papers have likely promoted interest in CRC screening research and added momentum to the production of publications.

When analyzing the type of articles, the most frequently cited were guideline articles. Guidelines are usually composed of the accumulation of a large research body that can influence the clinical setting.<sup>49</sup> Over the past 25 years, guidelines for CRC screening have been composed by professional groups and by a panel of expert gastroenterologists. They offer recommendations to assist practitioners and patients in decisions regarding



screening variables such as average-risk persons, high-risk family history, screening tools, and quality indicators.<sup>23,28,50,51</sup> The beneficial effects of guidelines depend on the successful adaptation to clinical settings. The high citation rate of CRC screening guidelines reflects their contribution to the field.

Most of the CRC screening research studies have been performed in North America and Europe. In these countries, greater resources are available and screening is more frequently implemented.<sup>4</sup> CRC incidence and the implementation of CRC screening differs among continents and countries.<sup>4</sup> The US leads in the number of publications and citations in CRC screening, which reflects the prominent role of the US gastroenterology community and its dominant position in international CRC screening research. The advancements of screening programs in the US can also be attributed to the endeavors of various national societies including the U.S. Preventive Services Task Force, American Cancer Society, American Gastroenterological Association, American Society for Gastrointestinal Endoscopy, and National Colorectal Cancer Roundtable. The extensive research in this field as well as the progression of screening programs in the US have resulted in a decrease in CRC incidence and mortality over the past two decades, as reported by the American Cancer Society.<sup>52</sup>

Over the last few decades, there is an increasing trend in CRC incidence and mortality in Asia.<sup>53</sup> We have demonstrated a high growth rate of CRC screening publications in China. The screening programs in this country are still relatively lacking.<sup>54,55</sup> Hopefully, the rising trend in CRC screening publications can promote the understanding of screening significance, which will ultimately influence screening behavior for the wide population.

In our study, we performed a text-mining analysis of two-word combinations. This allowed us to study “hot topics” in the field of CRC screening. Naturally, the most researched topic in the field of CRC screening was “screening and surveillance programs.” This topic has remained relatively stable over 25 years. We found that “quality assurance” was the most commonly trending topic over the last 5 years. This may help predict, to a definite extent, future trends in CRC publications. “Quality assurance” defines optimization

of the benefit to risk ratio of colonoscopy screening.<sup>36,56</sup> Initially, research focused on the implementation of CRC screening programs but, with time, an emphasis has also been placed on the quality of screening.

The topic of “non-invasive tests” is a prominent subject with a slight non-significant increase in the number of studies during the last decade. New laboratory tests include DNA, RNA, and protein biomarker stool and blood tests.<sup>57</sup> Novel imaging tests include colon capsule endoscopy<sup>58</sup> and magnetic resonance colonography.<sup>59</sup> The focus of research on this topic can be attributed to the attempts to develop and implement non-invasive tests, thereby reducing the need for colonoscopy for low-risk populations.

Although a relatively small number of studies have focused on “race disparities,” in 2002 race related research showed an increase in interest and has plateaued since then. Disparities in CRC screening are experienced by minority groups. Screening rates remain low for African Americans, Hispanics, and Asians.<sup>60–62</sup> The research accumulated on “racial disparities” can promote effective intervention designed to decrease gaps in CRC screening.

The research topics of “screening among IBD patients” and “risk stratification” have declined over the years. This may indicate that a foundation for recommendations for screening high-risk groups has already been effectively formulated.

When observing the 20 most cited articles, we can note that these studies have been published in the top-ranking world medical journals. In total, 11 have been published in *The New England Journal of Medicine*, thus reflecting the importance of the subject.

Our research has several limitations. First, this is a comprehensive study that includes 25 years of research conducted in 97 countries. As such, it can only provide a representation of CRC screening research on a global level. Second, the citation frequency was extracted from data provided by NCBI, while other options such as google scholar might have produced different results. Lastly, we used two-word combinations for topic modeling. Other approaches are available, such as latent Dirichlet allocation, but were found to be less effective in our study.

In conclusion, the number of publications devoted to CRC screening is steadily rising, with high-quality research reaching top-tier journals. A surge in the number of publications on the topics has been increasing in countries previously much less involved in academic research in the field. Screening programs remain the most researched topic, and quality indicators in screening colonoscopy has been attracting attention in recent years. A text-mining analysis of CRC screening research contributes to the understanding of current publication trends and topics. This technique has predictive value in illuminating future trends in CRC publications.

### Conflict of interest statement

The authors declare that there is no conflict of interest.

### Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

### ORCID iDs

Shelly Soffer  <https://orcid.org/0000-0002-7853-2029>

Uri Kopylov  <https://orcid.org/0000-0002-7156-0588>

### References

- Jemal A, Siegel R, Ward E, *et al.* Cancer statistics, 2008. *CA Cancer J Clin* 2008; 58: 71–96.
- Simon K. Colorectal cancer development and advances in screening. *Clin Interv Aging* 2016; 11: 967–976.
- Winawer SJ, Zauber AG, Ho MN, *et al.* Prevention of colorectal cancer by colonoscopic polypectomy. *N Engl J Med* 1993; 329: 1977–1981.
- Schreuders EH, Ruco A, Rabeneck L, *et al.* Colorectal cancer screening: a global overview of existing programmes. *Gut* 2015; 64: 1637–1649.
- Chen C, Stock C, Hoffmeister M, *et al.* Optimal age for screening colonoscopy: a modeling study. *Gastrointest Endosc* 2019; 89: 1017–1025.e12.
- Kahi CJ, Imperiale TF, Juliar BE, *et al.* Effect of screening colonoscopy on colorectal cancer incidence and mortality. *Clin Gastroenterol Hepatol* 2009; 7: 770–775.
- Ansa B, Coughlin S, Alema-Mensah E, *et al.* Evaluation of colorectal cancer incidence trends in the United States (2000–2014). *J Clin Med* 2018; 7: 22.
- Ohba N, Nakao K, Isashiki Y, *et al.* The 100 most frequently cited articles in ophthalmology journals. *Arch Ophthalmol* 2007; 125: 952–960.
- Filardo G, da Graca B, Sass DM, *et al.* Trends and comparison of female first authorship in high impact medical journals: observational study (1994–2014). *BMJ* 2016; 352: i847.
- Sepehrvand N, Alemayehu W, Das D, *et al.* Trends in the explanatory or pragmatic nature of cardiovascular clinical trials over 2 decades. *JAMA Cardiol* 2019; 4: 1122–1128.
- Thuraisingham B. *Data mining: technologies, techniques, tools, and trends*. Boca Raton: CRC Press, 2014.
- Barash Y, Klang E, Tau N, *et al.* Evolution of inflammatory bowel disease research from a bird's-eye perspective: a text-mining analysis of publication trends and topics. *Inflamm Bowel Dis*. Epub ahead of print 22 May 2020. DOI: 10.1093/ibd/izaa091.
- Wang S-H, Ding Y, Zhao W, *et al.* Text mining for identifying topics in the literatures about adolescent substance use and depression. *BMC Public Health* 2016; 16: 279.
- Zhang Y, Tao J, Wang J, *et al.* Trends in diatom research since 1991 based on topic modeling. *Microorganisms* 2019; 7: 213.
- Song M and Kim SY. Detecting the knowledge structure of bioinformatics by mining full-text collections. *Scientometrics* 2013; 96: 183–201.
- N.I. of H. U.S. National Library of Medicine, Download MEDLINE/PubMed Data, (n.d.). [www.nlm.nih.gov/databases/download/pubmed\\_medline.html](http://www.nlm.nih.gov/databases/download/pubmed_medline.html) (accessed 1 January 2019)
- Ripple AM, Mork JG, Knecht LS, *et al.* A retrospective cohort study of structured abstracts in MEDLINE, 1992–2006. *J Med Libr Assoc* 2011; 99: 160–163.
- Zou C. Analyzing research trends on drug safety using topic modeling. *Expert Opin Drug Saf* 2018; 17: 629–636.
- Software Tools—Download—NCBI. [online], (n.d.) <https://www.ncbi.nlm.nih.gov/home/tools/> (accessed 1 January 2019)
- The World Bank. World development indicators (WDI), [Datacatalog.worldbank.org](http://datacatalog.worldbank.org) (2019). <https://datacatalog.worldbank.org/dataset/world-development-indicators> (accessed 1 January 2019)

21. US Preventive Services Task Force. Screening for colorectal cancer: US Preventive Services Task Force recommendation statement. *Ann Intern Med* 2008; 149: 627–637.
22. Mandel JS, Bond JH, Church TR, *et al.* Reducing mortality from colorectal cancer by screening for fecal occult blood. *N Engl J Med* 1993; 328: 1365–1371.
23. Levin B, Lieberman DA, McFarland B, *et al.* Screening and surveillance for the early detection of colorectal cancer and adenomatous polyps, 2008: a joint guideline from the American Cancer Society, the US Multi-Society Task Force on Colorectal Cancer, and the American College of Radiology. *CA Cancer J Clin* 2008; 58: 130–160.
24. Edwards BK, Ward E, Kohler BA, *et al.* Annual report to the nation on the status of cancer, 1975–2006, featuring colorectal cancer trends and impact of interventions (risk factors, screening, and treatment) to reduce future rates. *Cancer* 2010; 116: 544–573.
25. Hardcastle JD, Chamberlain JO, Robinson MH, *et al.* Randomised controlled trial of faecal-occult-blood screening for colorectal cancer. *Lancet* 1996; 348: 1472–1477.
26. Zauber AG, Winawer SJ, O'Brien MJ, *et al.* Colonoscopic polypectomy and long-term prevention of colorectal-cancer deaths. *N Engl J Med* 2012; 366: 687–696.
27. Kronborg O, Fenger C, Olsen J, *et al.* Randomised study of screening for colorectal cancer with faecal-occult-blood test. *Lancet* 1996; 348: 1467–1471.
28. Winawer S, Fletcher R, Rex D, *et al.* Colorectal cancer screening and surveillance: clinical guidelines and rationale—update based on new evidence. *Gastroenterology* 2003; 124: 544–560.
29. Rex DK, Johnson DA, Anderson JC, *et al.* American College of Gastroenterology guidelines for colorectal cancer screening 2008. *Am J Gastroenterol* 2009; 104: 739.
30. Atkin WS, Edwards R, Kralj-Hans I, *et al.* Once-only flexible sigmoidoscopy screening in prevention of colorectal cancer: a multicentre randomised controlled trial. *Lancet* 2010; 375: 1624–1633.
31. Lieberman DA, Weiss DG, Bond JH, *et al.* Use of colonoscopy to screen asymptomatic adults for colorectal cancer. *N Engl J Med* 2000; 343: 162–168.
32. Pickhardt PJ, Choi JR, Hwang I, *et al.* Computed tomographic virtual colonoscopy to screen for colorectal neoplasia in asymptomatic adults. *N Engl J Med* 2003; 349: 2191–2200.
33. Mandel JS, Church TR, Bond JH, *et al.* The effect of fecal occult-blood screening on the incidence of colorectal cancer. *N Engl J Med* 2000; 343: 1603–1607.
34. Baxter NN, Goldwasser MA, Paszat LF, *et al.* Association of colonoscopy and death from colorectal cancer. *Ann Intern Med* 2009; 150: 1–8.
35. Hampel H, Frankel WL, Martin E, *et al.* Screening for the Lynch syndrome (hereditary nonpolyposis colorectal cancer). *N Engl J Med* 2005; 352: 1851–1860.
36. Kaminski MF, Regula J, Kraszewska E, *et al.* Quality indicators for colonoscopy and the risk of interval cancer. *N Engl J Med* 2010; 362: 1795–1803.
37. Nishihara R, Wu K, Lochhead P, *et al.* Long-term colorectal-cancer incidence and mortality after lower endoscopy. *N Engl J Med* 2013; 369: 1095–1105.
38. Järvinen HJ, Aarnio M, Mustonen H, *et al.* Controlled 15-year trial on screening for colorectal cancer in families with hereditary nonpolyposis colorectal cancer. *Gastroenterology* 2000; 118: 829–834.
39. Selby JV, Friedman GD, Quesenberry CP Jr, *et al.* A case-control study of screening sigmoidoscopy and mortality from colorectal cancer. *N Engl J Med* 1992; 326: 653–657.
40. Whitlock EP, Lin JS, Liles E, *et al.* Screening for colorectal cancer: a targeted, updated systematic review for the US Preventive Services Task Force. *Ann Intern Med* 2008; 149: 638–658.
41. Chi Y. Global trends in medical journal publishing. *J Korean Med Sci* 2013; 28: 1120–1121.
42. Rim SH, Joseph DA, Steele CB, *et al.* Colorectal cancer screening—United States, 2002, 2004, 2006, and 2008. *MMWR Suppl* 2011; 60: 42–46.
43. Smith RA, Andrews KS, Brooks D, *et al.* Cancer screening in the United States, 2019: a review of current American Cancer Society guidelines and current issues in cancer screening. *CA Cancer J Clin* 2019; 69: 184–210.
44. European Union. Council recommendation on cancer screening. *Off J Eur Union* 2003.
45. World Health Organization. International agency for research of cancer; IARC handbooks of cancer prevention. In: *Carotenoids*. Vol. 2. Lyon, France, 1998.

46. Hampton T. Cancer prevention efforts stalled. *JAMA* 2008; 299: 2264.
47. Arnold M, Sierra MS, Laversanne M, *et al.* Global patterns and trends in colorectal cancer incidence and mortality. *Gut* 2017; 66: 683–691.
48. Vining DJ, Hunt GW, Ahn DK, *et al.* *Automatic analysis in virtual endoscopy*. U.S. Patent No. 5,920,319, 1999.
49. Graham ID and Harrison MB. Evaluation and adaptation of clinical practice guidelines. *Evid Based Nurs* 2005; 8: 68–72.
50. Rex DK, Boland CR, Dominitz JA, *et al.* Colorectal cancer screening: recommendations for physicians and patients from the U.S. Multi-Society Task Force on Colorectal Cancer. *Gastrointest Endosc* 2017; 86: 18–33.
51. European Colorectal Cancer Screening Guidelines Working Group. European guidelines for quality assurance in colorectal cancer screening and diagnosis: overview and introduction to the full supplement publication. *Endoscopy* 2013; 45: 51–59.
52. Siegel R, DeSantis C and Jemal A. Colorectal cancer statistics, 2014. *CA Cancer J Clin* 2014; 64: 104–117.
53. Pourhoseingholi MA. Increased burden of colorectal cancer in Asia. *World J Gastrointest Oncol* 2012; 4: 68–70.
54. Goss PE, Strasser-Weippl K, Lee-Bychkovsky BL, *et al.* Challenges to effective cancer control in China, India, and Russia. *Lancet Oncol* 2014; 15: 489–538.
55. Sung JJ, Choi SY, Chan FK, *et al.* Obstacles to colorectal cancer screening in Chinese: a study based on the health belief model. *Am J Gastroenterol* 2008; 103: 974–981.
56. Rex DK, Schoenfeld PS, Cohen J, *et al.* Quality indicators for colonoscopy. *Gastrointest Endosc* 2015; 81: 31–53.
57. Ouyang DL, Chen JJ, Getzenberg RH, *et al.* Noninvasive testing for colorectal cancer: a review. *Am J Gastroenterol* 2005; 100: 1393–1403.
58. Eliakim R, Fireman Z, Gralnek I, *et al.* Evaluation of the PillCam Colon capsule in the detection of colonic pathology: results of the first multicenter, prospective, comparative study. *Endoscopy* 2006; 38: 963–970.
59. Pappalardo G, Poletini E, Frattaroli FM, *et al.* Magnetic resonance colonography versus conventional colonoscopy for the detection of colonic endoluminal lesions. *Gastroenterology* 2000; 119: 300–304.
60. Jerant AF, Fenton JJ and Franks P. Determinants of racial/ethnic colorectal cancer screening disparities. *Arch Intern Med* 2008; 168: 1317–1324.
61. Shokar NK, Carlson CA and Weller SC. Factors associated with racial/ethnic differences in colorectal cancer screening. *J Am Board Fam Med* 2008; 21: 414–426.
62. Liss DT and Baker DW. Understanding current racial/ethnic disparities in colorectal cancer screening in the United States: the contribution of socioeconomic status and access to care. *Am J Prev Med* 2014; 46: 228–236.