Therapeutic Advances in Gastroenterology Original Research

A bibliometric analysis of 23,492 publications on rectal cancer by machine learning: basic medical research is needed

Ther Adv Gastroenterol

2020, Vol. 13: 1–11

DOI: 10.1177/1756284820934594

Kangtao Wang, Chenzhe Feng, Ming Li, Qian Pei, Yuqiang Li, Hong Zhu, Xiangping Song, Haiping Pei and Fengbo Tan

Abstract

Background and Aims: The aim of this study was to analyse the landscape of publications on rectal cancer (RC) over the past 25 years by machine learning and semantic analysis.

Methods: Publications indexed in PubMed under the Medical Subject Headings (MeSH) term 'Rectal Neoplasms' from 1994 to 2018 were downloaded in September 2019. R and Python were used to extract publication date, MeSH terms and abstract from the metadata of each publication for bibliometric assessment. Latent Dirichlet allocation was applied to analyse the text from the articles' abstracts to identify more specific research topics. Louvain algorithm was used to establish a topic network resulting in identifying the relationship between the topics.

Results: A total of 23,492 papers published were identified and analysed in this study. The changes of research focus were analysed by the changing of MeSH terms. Studied contents extracted from the publications were divided into five areas, including surgical intervention, radiotherapy and chemotherapy intervention, clinical case management, epidemiology and cancer risk as well as prognosis studies.

Conclusions: The number of publications indexed on RC has expanded rapidly over the past 25 years. Studies on RC have mainly focused on five areas. However, studies on basic research, postoperative quality of life and cost-effective research were relatively lacking. It is predicted that basic research, inflammation and some other research fields might become the potential hotspots in the future.

Keywords: bibliometric analysis, LDA analyses, machine learning, rectal cancer

Received: 18 December 2019; revised manuscript accepted: 11 May 2020.

Correspondence to:
Haiping Pei
Department of Gastrointestinal Surgery, The Xiangya Hospital Central South University, Xiangya Road 110#Changsha 410008, China
haipingpei1966@hotmail.com

Fengbo Tan
Department of Gastrointestinal Surgery, The Xiangya Hospital Central South University, Xiangya Road 110#, Changsha, Hunan 410008, China
fengbotan@csu.edu.cn

Kangtao Wang
Qian Pei
Department of Gastrointestinal Surgery, The Xiangya Hospital, Central South University, Changsha, China

Chenzhe Feng
Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing, China

Ming Li
Department of Immunology, College of Basic Medical Science, Central South University, Changsha, China

Yuqiang Li
Department of General, Visceral and Thoracic Surgery, University Medical Centre Hamburg-Eppendorf, Hamburg, Germany

Hong Zhu
Department of Oncology, The Xiangya Hospital, Central South University, Changsha, China

Xiangping Song
UPMC Hillman Cancer Centre, Pittsburgh, PA, USA

Introduction

The understanding and treatment of rectal cancer (RC) have made great progress in the past 25 years. The publications showed that the emergence of new diagnostic and therapeutic technologies has led to a downward trend in the prevalence and mortality of RC worldwide.[1] The mortality rate of RC in the United States decreased from 100–200/100,000 in the 1990s to 8–12/100,000 in 2017. The 5-year survival rate of RC increased from 45–58% in 1983–1989 to 68–76% in 2010–2015.[1–6] However, RC remains a deadly malignant tumour, which causes huge disease burden. Moreover, because of rectum unique physiological position and function, diagnosis and treatment of RC have brought huge physiological and psychological suffering to patients.

Most of the progress of studies on RC treatment is reported in the form of publications. With the continuous development of the publishing industry, the number of papers published has risen every year. The number of academic publications is often presented as a trend that reflects research

interest and public attention within a certain period.[7] Bibliometrics is a method to analyse academic literature quantitatively. It is important and useful to quantitatively analyse a large number of literature data to summarize what has been done and speculate on possible future research. The last bibliometric study on colorectal cancer (CRC) was carried out in 2016 by Stout, identifying the top 100 cited articles.[8] However, as we are lacking effective semantic analysis tools to analyse a large amount of literature, so far, the overall landscape of all publications on RC remains uncharacterized.

Natural language processing (NLP) is a computational technique for analysing human language which was applied to process medical information. There have been substantial research achievements on applications in processing medical information by using NLP methods.[9,10] The latent Dirichlet allocation (LDA) is the most classic topic modelling method used in bibliometrics to characterize large numbers of unstructured texts. It creates a feature glossary of terms based on how often the vocabulary coexists in the document set. Once the vocabulary glossary is created, LDA determines the probability that an article belongs to a particular topic based on the frequency with which the vocabulary appears in each document.[11,12] It was reported that NLP and LDA were used to identify specific themes in a large number of publications by investigating topics and trends in cancer rehabilitation and cardiovascular diseases.[13,14]

In the present study, we analysed the themes of publications on RC in the past 25 years (1994–2018) indexed by PubMed under the Medical Subject Headings (MeSH) term 'Rectal Neoplasm'. Inspired by Stout's research on cancer rehabilitation,[14] and guided by NLP and LDA analysis, we modified the machine learning methodology to conduct a more detailed analysis on publications on RC resulting in finding areas where the evidence is robust, highlighting fields where publication deficits exist, and identifying specific areas where opportunities will exist in the future.

Materials and methods

We searched the public version of PubMed indexed under the MeSH term 'Rectal Neoplasms' from 1994 to 2018, in September 2019. The complete record of the search results was downloaded in XML format. R was used to extract metadata including the year when each article was published, the abstract, the type of research and the MeSH terms. LDA was used to identify more specific research topics in each article. Python was used to model the topics by analysing the abstracts of all the indexed articles in our record and set the number of topics identified to 50. The criteria of topic number selection are based on proper perplexity, redundancy and legibility.[10,15] Based on the subject probability calculated by the algorithm, we finally determined the main topic of each article. We named each term glossary through manual checking based on the abstract and MeSH terms. Finally, we used the Louvain algorithm for cluster analysis in order to establish a topic network resulting in identifying the relationship between the topics. In each article, we identified two topics with the highest probability of attribution, counted the number of simultaneous occurrences of the two topics in each document and established a link between the topics.

XML format analysis was performed with the R programming language in R Studio (https://rstudio.org). The R code is publicly available on GitHub (https://github.com/christopherBelter/pubmedXML). We published all the relevant code on GitHub (https://github.com/yanwen0614/Medicine-Bibliometric-Analysis). The network visualizations in the article were performed with the software package Gephi (https://github.com/gephi/gephi).[16] This article does not require approval by the relevant institutional review board or ethics committee. All the R and Python codes and raw XML data are available in the Supplemental Material online.

Results

Overview data

The search string revealed 23,492 publications indexed under the MeSH term 'Rectal Neoplasms' from 1994 to 2018. An average of 603 articles were published per year from 1994 to 2002, while the body of literature has grown at a rate of 6–13% per year from 2003 to 2016 (Figure 1). However, there was a numerical decline in 2017 and 2018. It may reflect an indexing delay rather than a decline in publications.[14] The 23,492 publications could be roughly classified into eight types: systemic review, meta-analysis, comment, multi-centre study, letter, case report, review and clinical trial (Figure 2).

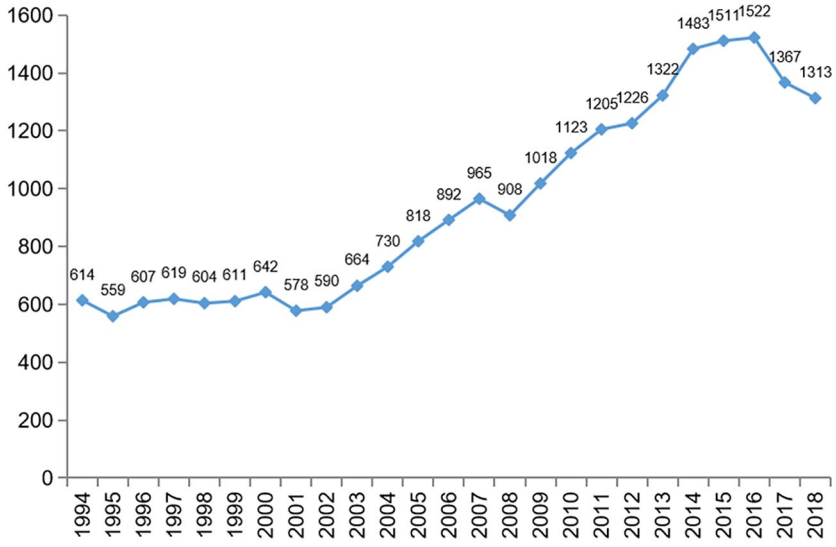

Figure 1. PubMed search results: articles published per year.

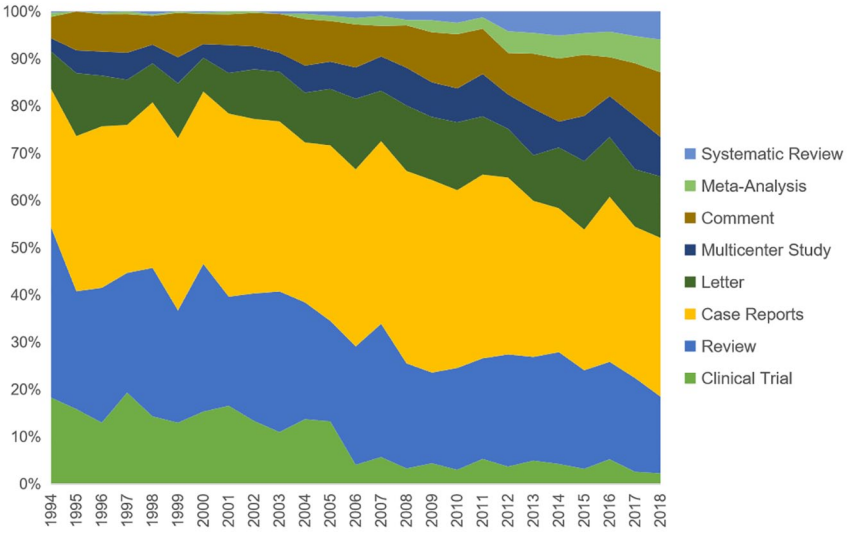

Figure 2. PubMed search results: distribution of each type of publications per year.

It is remarkable that papers of the systematic reviews as well as meta-analysis increased from less than 1% in the 1990s to 13% in 2018, while papers of clinical trials decreased from 18% in 1994 to 4% in 2018. The number of other types of articles published has not changed much over the years.

MeSH analysis

We have found a total 7237 MeSH terms in all the publications. They occurred 296,490 times. In order to simplify our study, 771 MeSH terms whose occurrence was more than 25 were chosen in the study. We found that 106 MeSH terms appeared after 1994, such as chemoradiotherapy, chemoradiotherapy adjuvant, capecitabine and so on (Figure 3). Although these terms appeared 9049 times in all, accounting for only 3% of all the occurrences, they represented the emergence of some new research fields. In fact, many new technologies represented by these newly emerging MeSH terms are currently part of the first line

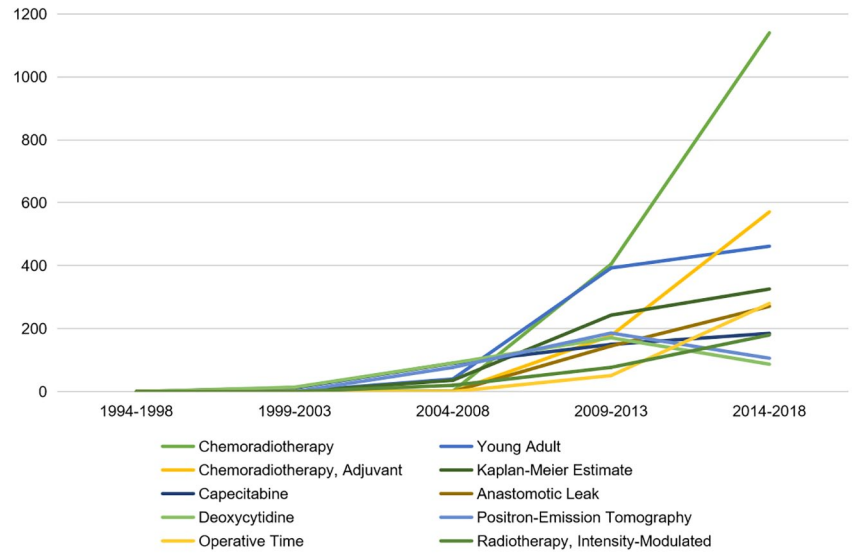

Figure 3. Top 10 MeSH terms appearing after 1994 and changes per five years.

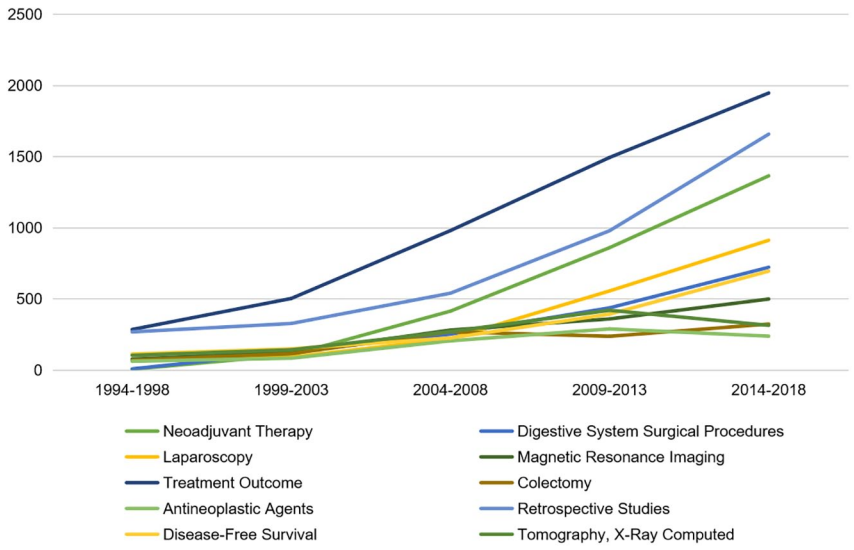

Figure 4. Top 10 MeSH term appearing before 1994 and changes per 5 years.

treatments, such as chemoradiotherapy and neoadjuvant therapy.

Among the remaining 665 MeSH terms, we list the top 10 MeSH terms, such as neoadjuvant therapy, digestive system surgical procedures, laparoscopy and so on, which appeared newly in 1994, but more frequently after 1994 (Figure 4). Among these terms, we are interested in some MeSH terms, including neoadjuvant therapy, laparoscopy, fluorouracil, radiotherapy adjuvant, chemotherapy adjuvant. Figure 5 shows the changing proportion of these terms over the past 25 years. On the one hand, the proportion of neoadjuvant chemoradiotherapy increased gradually after its first appearance in 1998. The proportions of laparoscopy and robotic surgical procedures were also gradually increasing. On the other hand, the proportions of adjuvant radiotherapy and adjuvant chemotherapy have gradually decreased since 2002. The proportions of magnetic resonance imaging (MRI) as well as X-ray computed tomography have been flat over

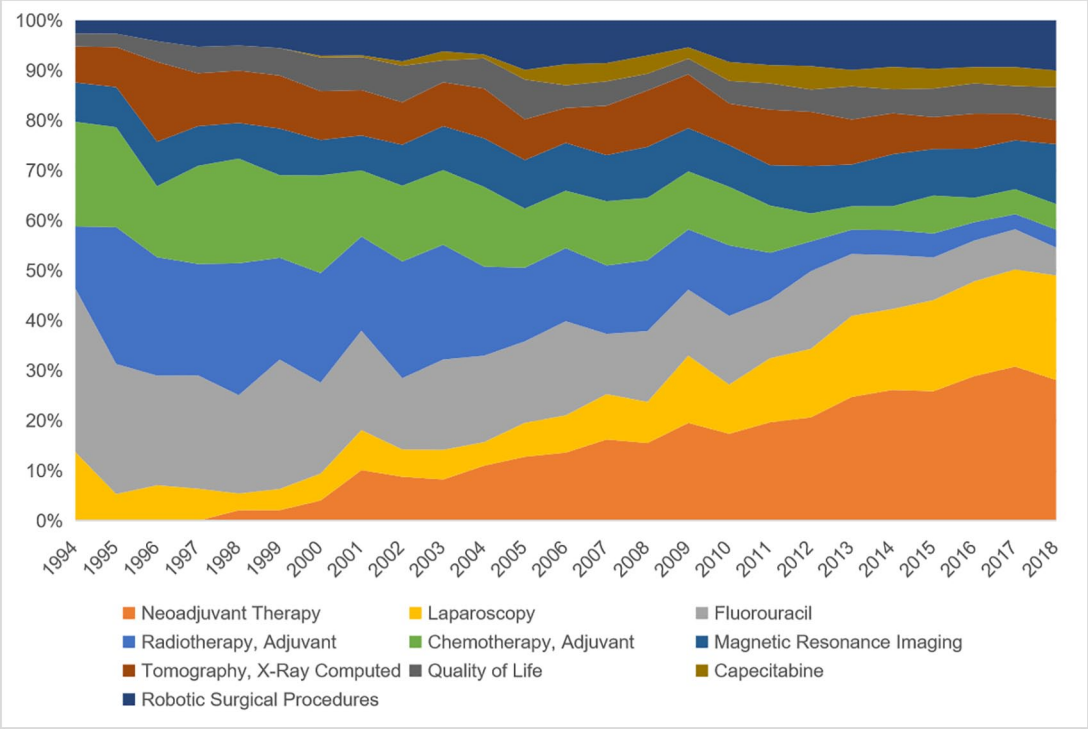

Figure 5. Proportion changes of some representative MeSH terms.

25 years. These findings suggest that the focus on RC treatment strategies has shifted from adjuvant radiotherapy and chemotherapy to neoadjuvant therapy, endoscopy and robotic surgery. In addition, we also found that the occurrences of some MeSH terms, such as positron-emission tomography and deoxycytidine, decreased in recent 25 years. In all, these findings showed the changes of different research fields. (Supplemental Material provides details of MeSH terms.)

LDA analysis

Topic network analysis by LDA and Louvain algorithm highlights areas where clusters of interrelated topics co-occur and provides insight into the relationships between prominent topics of interest. Five topic-network clusters analysed by LDA are shown in five different colours in Figure 6. The five areas are Surgical intervention (marked in green), Radiotherapy and chemotherapy intervention (marked in purple), Clinical case management (marked in blue), Epidemiology and cancer risk study (marked in orange) and Prognosis study (marked in red). The size of the bubble in each area represents the number of the related publications. The magnitude of the relationships among five areas can be seen both within and between topic clusters by the network line connections, with thicker lines representing greater numbers of articles that use certain terms.

In the Surgical intervention cluster (green area), the topics that have been studied more were local recurrence, transanal surgery, laparoscopic and open surgery, as well as surgery complication. Strong relationships are seen among these topics, suggesting that they are closely related and carefully studied in this area. However, there were a small number of publications on the two topics quality of life and anal function protection. Furthermore, there was weak connection between these two topics. On the whole, the cluster of Surgery intervention is strongly related to both of the clusters of Radiotherapy and chemotherapy intervention and Prognostic study.

Case reports accounted for the majority of publications in the cluster of Clinical case management (blue area). Most case reports were relatively isolated and reported descriptions of symptomatology and rare cases. A few cases reported symptoms, endoscopic resection or rare and anal cancer. In view of the five areas, this cluster was

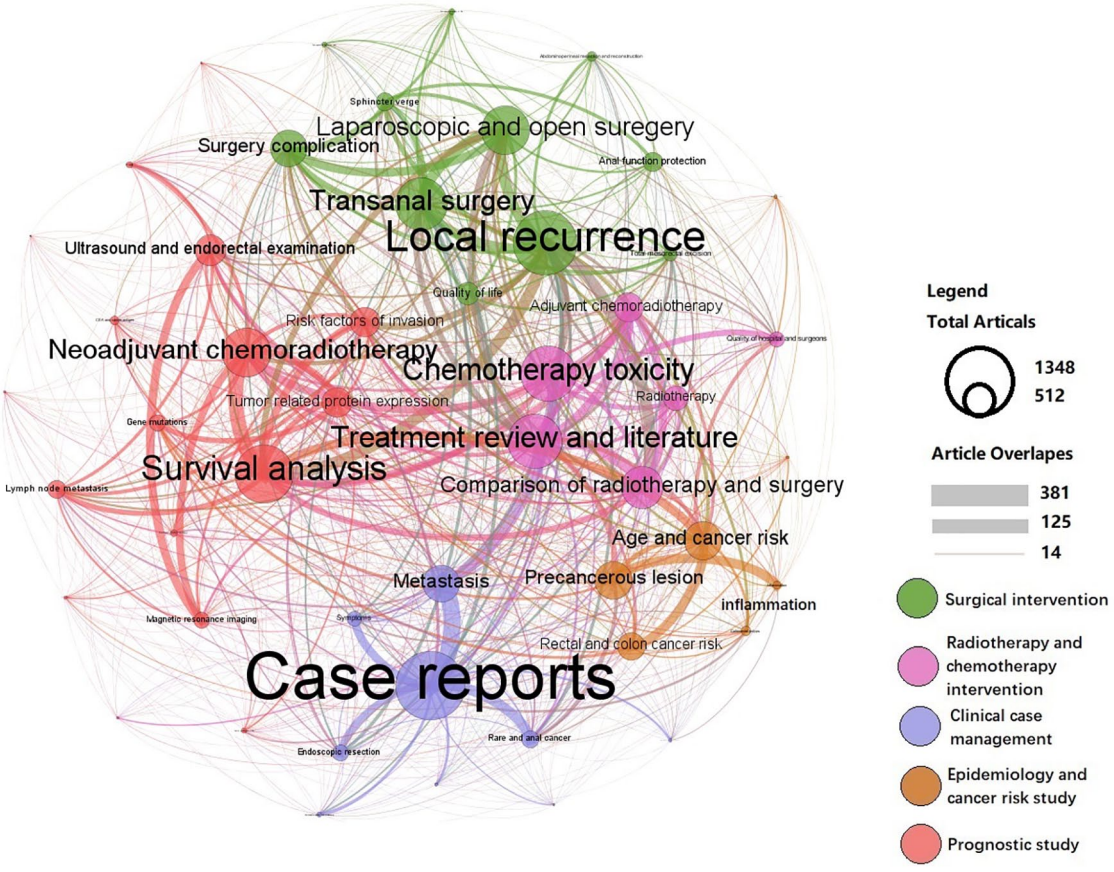

Figure 6. Topic cluster network studied by Latent Dirichlet allocation: inter- and intra-relationships. Green, Surgical intervention; Purple, Radiotherapy and chemotherapy intervention; Blue, Clinical case management; Orange, Epidemiology and cancer risk study; Red, Prognostic study. The size of circle represents the number of papers in each topic, the thickness of line represents the weight of connection between each topic.

relatively independent. However, the reported cases associated with liver metastasis were strongly related to the cluster of Radiotherapy and chemotherapy intervention (purple area).

In the cluster of Epidemiology and cancer risk study (orange area), there were three major topics: Age and cancer risk, Precancerous lesion and Rectal and colon cancer risk. It is noteworthy that although the number of publications studying inflammation was rarely small, inflammation showed strong correlation with precancerous lesions. In the view of the relationships among the five clusters, the cluster of Epidemiology and cancer risk study shows strong correlation with the cluster of Prognostic study (red area).

In the cluster of Radiotherapy and chemotherapy intervention (purple area), four main topics were Chemotherapy toxicity, Treatment review and literature, Comparison of radiotherapy and surgery and Adjuvant chemoradiotherapy. The studies on comparison of radiotherapy and surgery were closely connected with the other three topics, suggesting that radiotherapy and surgery was a very important therapy for RC. Furthermore, this cluster is strongly related with the other four clusters, demonstrating that this cluster was a key field in last 25 years.

Among the Prognostic study cluster (red area), both Survival analysis and Neoadjuvant chemoradiotherapy are the most studied topics. Although there were fewer than 350 papers on the topic of magnetic resonance imaging, this topic was closely associated with multiple topics, suggesting that MRI was a research hotspot in this area. In addition, this cluster also contained a variety of

topics such as gene mutation, tumour related protein expression, risk factors of invasion and so on, which suggested that studies in this cluster were very comprehensive. The small number of basic research papers on RC indicates that there are a lot of blanks in basic research on RC. This cluster also showed close relationship with the other four clusters.

Discussion

For the first time, 23,492 literatures on RC published during 1994–2018 were systematically analysed by machine learning and semantic analysis. The average number of annual publications was 603 from 1994 to 2002, and increased by 74% from 2002 to 2007 and 87% from 2008 to 2014. The numbers of papers declined in 2017 as well as 2018, which may be due to index delay, rather than a decline in publications. Studies on RC over 25 years have mainly focused on five fields: surgical intervention, radiotherapy and chemotherapy intervention, clinical case management, epidemiology and cancer risk study, and prognosis study. The relationships among the clusters and the topics were identified.

Surgical intervention of RC has been a hot and controversial topic for the last 25 years. Our study found that surgical treatment strategies kept improving. The operation mode was developed from open surgery to laparoscopic, robotic surgery. Total mesorectal excision (TME) was gradually improved into transanal total mesorectal excision (TaTME). Heald *et al.* put forward TME for the first time in 1982.[17] Jacobs firstly completed laparoscopic radical resection in 1991.[18] However, it was not until the results of the COREAN trail were published in 2013 that the National Cancer Network Guidelines (NCCNs) had a more positive attitude toward laparoscopic radical surgery.[18,19] Sylla launched the world's first transanal endoscopic microsurgery-based TaTME surgery in 2010.[20] However, TaTME is still a relatively new surgery procedure with many problems needing to be solved, especially in peritoneal exploration, mesenteric lymph node dissection and left colon vascular management.[21] However, our previously study found that the improvement of RC prognosis might mostly be attributed to radiotherapy and chemotherapy rather than surgery improvement.[22] Therefore, newly RC surgical strategies still need to be supported by more RCTs and clinical data.

Meanwhile, we find that metastatic RC treated by surgical methods was scarcely reported over the past 25 years. Recent studies have shown that there are many advantages of surgery, combined with neoadjuvant radiotherapy and chemotherapy for liver metastasis of RC.[23] Surgical intervention for liver metastasis of RC may suggest that other tissues' metastasis of RC potentially achieve good prognosis under surgical intervention. Therefore, surgical intervention for metastatic cancer may become one of the research focuses in the future. It is well-known that the complications of RC treated by surgical or radio therapy are very painful for patients due to the rectal physical function. However, our study showed that there were few publications on studies of anal function protection and patients' quality of life over the past years. These two topics account for less than 8% of the total amount of research related to surgery intervention. Therefore, more attention should be paid to anal function protection and patients' quality of life by using advanced surgical strategies in the future.

Local recurrence was not only the most studied topic in the field of surgery intervention, but was also strongly associated with the other four clusters. We found that the rate of local recurrence has been gradually decreased from 20–30% to 5–10% over the past 25 years.[24] Neoadjuvant chemoradiotherapy, especially radiotherapy, was reported to play an important role in treatment of local recurrence.[25] Furthermore, the topic of comparison of radiotherapy and surgery shows a strong connection with local recurrence. Some retrospective studies demonstrated that compared with surgery, preoperative radiotherapy of low RC and local advanced RC could increase survival rate, control distance metastasis, improve patient's life quality, preserve anus and secondary reconstruct anus. Therefore, the comparison between radiotherapy and surgery might still be a research focus and will stay controversial in the future.[26]

Epidemiology and cancer risk of RC is another hot research field. Along with our understanding of characteristics of sporadic and hereditary CRC, many effective methods for diagnosis and prevention of the cancer in early stage have been developed.[27] In the MeSH analysis, we found that many methods were developed for early diagnosis, including biomarker detection, occult blood test, endoscopy examination and faecal DNA

sequencing.[28] Another interesting finding was the topic of inflammation in this cluster. There were only 622 (2.6%) articles related to inflammation out of 23,492 publications. The weight between inflammation and precancerous lesions is 212, suggesting strong correlation between them. We also notice that inflammation has little correlation with study on epidemiology and cancer risk, corroborating that the current guidelines on cancer prevention do not clearly indicate the treatment of inflammation.[29] Basically, our study did not find that gut microbiota was a research hotspot, because of the small number of relative articles. However, recent studies have revealed that gut microbiota played an important role in CRC. Li demonstrated that gut microbiota-stimulated cathepsin K secretion mediated TLR4-dependent M2 macrophage polarization and promoted tumour metastasis in CRC, which could cause a higher death rate.[30–32] Therefore, it is reasonable that past researches on the role of inflammation in rectal carcinogenesis and development were insufficient. Inflammation is likely to be a hot studied topic in the future.[33]

The area of prognostic study contained many topics, not only basic medicine researches of RC but also clinic and diagnostic strategies. These themes showed that prognosis of RC is influenced by many factors. It was interesting that neoadjuvant chemoradiotherapy was incorporated into the area of prognostic study rather than the radiotherapy and chemotherapy cluster. Neoadjuvant therapy was firstly proposed as a MeSH term in 1995 and became one of the research's focuses over the following years. Neoadjuvant therapy was firstly introduced in the 2007 edition of the NCCN of RC. Based on the findings of two clinical trials, MRCCR07 and CAO/ARO/AIO-94, preoperative neoadjuvant therapy combined with radical surgery is considered to be the gold standard for the treatment of locally advanced RC.[34,35] Many clinical studies demonstrated that neoadjuvant chemotherapy assisting TME shows a positive effect on the prognosis of advanced RC. Another hotspot is MRI. The number of papers concerned with MRI was 1253 (5.3%), but the total combined weights with other studies was 396, suggesting that MRI was closely related to neoadjuvant chemoradiotherapy, ultrasound and endoscopy et cetera. Pelvis MRI was first completed in 1984.[36] Recent studies using developed MRI have shown that MRI plays an important role in the staging of RC. In the different guidelines of RC, MRI is considered an indispensable examination.[37] In this cluster, the basic research of RC is insufficient. We found that the proportion of basic research such as gene mutation is less than 3% of total publications, which may be related to the fact that most of the relevant RC journals are clinical journals. At present, even the cell line model of RC has not been established. RC is still studied together with colon cancer in many studies. However, due to the different biological behaviours of these two cancers, they have been gradually separated for research. Therefore, the basic research of RC will be an indispensable research content in the future.

In the past 25 years, studies on the cluster of Radiotherapy and chemotherapy intervention of RC showed considerable progress. Topics including Radiotherapy, Comparison of radiotherapy and surgery, Intensity of chemotherapy and Review of treatment are research hotpots. But some contradictions among these topics were found. Although 5Fu-based chemotherapy has brought survival benefits since 1990s,[38] there was a tendency to deal with RC by de-chemotherapeutic and de-radiotherapeutic methods by evaluation of different risk factors first.[39] We found that citation rates of adjuvant radiotherapy and chemotherapy have been gradually reduced, which is consistent with the mentioned concepts. Accurate assessment can benefit early patients and reduce the risk of radiotherapy and chemotherapy. Patients with advanced RC tend to be treated with neoadjuvant radiotherapy and chemotherapy with TME,[40] which also corresponded with our research results.

For the first time in our research, with the help of ML and NLP, we analysed over 20,000 publications on RC and have shown the study characteristics, deficiencies of research and possible research directions in the future. In the past 25 years, great progress has been made in the diagnosis and treatment of RC. However we find several things in our research. First, basic research on RC accounted for a small proportion. There are few studies on the basic mechanism of carcinogenesis and cancer development. Second, there is insufficient attention paid to functional protection, fistula and quality of life. Publications on these three topics comprised about 2000 (8.5%) articles, and many of them were related to surgical methods. Third, there were few studies on cost-effective research. Only about 350 (1.5%) papers described the

economic burden of cancer. We predicted that basic research, inflammation and some other research fields might become hotspots in the future. Our study proves that neoadjuvant radiotherapy and chemotherapy might be a more important factor for the improvement of survival prognosis than surgery. We found that several research fields, such as inflammation with RC, are likely to be novel research fields in the future and anticipate it will bring positive effect in understanding the pathogenicity and treatment of RC. It is reasonable that machine learning will play an active role in clinical practice in the future with sufficient medical record text databases and improved algorithms. Our finding suggests that Machine Learning and NLP could be a tremendous tool for scientists, who will be able to draw objective and comprehensive clues from huge numbers of data.

As for limitations, first, there are several other databases besides PubMed available for bibliometric research, including Scopus, Web of Science and Embase. Although PubMed contains the highest quality of peer-reviewed research and excludes irrelevant, non-peer-reviewed publications, detailed and comprehensive knowledge will be provided if other databases are explored meantime. Second, some publications may not have appeared in our study because they had not yet been indexed by MeSH words. Last but not least, the publications may include more positive results and less intuitive experience of clinical practice, which could cause bias in our research.

Acknowledgement

We would like to express our gratitude to Mr Wen Yan, who helped us in the programming and coding.

Author contribution

Conceptualization, KW, PH; methodology, KW, CF, WY; formal analysis, KW, YW, CF; investigation, ML, CF; writing – original draft preparation, KW, CF; writing, reviewing and editing, ML, HP, HZ, FT, QY, PH, PQ, XP; supervision, PH.

Conflict of interest statement

The authors declare that there is no conflict of interest.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: this research was funded by The Nature Scientific Foundation of China (81702956); The Strategy-Oriented Special Project of Central South University in China (ZLXD2017003); The XiangYa-Peking University Wei Ming Clinical and Rehabilitation Research Fund (xywm2015I21); General programme of Natural Science Foundation of Hunan Province (2019JJ40490).

ORCID iDs

Kangtao Wang https://orcid.org/0000-0003-0948-0277

Yuqiang Li https://orcid.org/0000-0003-1517-3687

Supplemental material

The following are available online: the R code is publicly available on GitHub (https://github.com/christopherBelter/pubmedXML). Relevant Python code on GitHub (https://github.com/yanwen0614/Medicine-Bibliometric-Analysis). The raw XML file is on (https://pan.baidu.com/s/1ulMJeIds2UvvDGu2m8pJDQ), code:2rca.

References

1. Siegel RL, Miller KD and Jemal A. Cancer statistics, 2020. *CA Cancer J Clin* 2020; 70: 7–30.
2. Kinzler KW and Vogelstein B. Landscaping the cancer terrain. *Science* 1998; 280: 1036–1037.
3. DuBois RN. COX-2 in large bowel cancer: a one-sided story. *Gut* 1999; 45: 636–637.
4. Singh GK and Jemal A. Socioeconomic and racial/ethnic disparities in cancer mortality, incidence, and survival in the United States, 1950–2014: over six decades of changing patterns and widening inequalities. *J Environ Public Health* 2017; 2017: 2819372.
5. Singh GK, Miller BA and Hankey BF. Changing area socioeconomic patterns in U.S. cancer mortality, 1950–1998: part II–lung and colorectal cancers. *J Natl Cancer Inst* 2002; 94: 916–925.
6. Keller DS, Berho M, Perez RO, *et al.* The multidisciplinary management of rectal cancer. *Nat Rev Gastroenterol Hepatol*. Epub ahead of print 12 March 2020. DOI: 10.1038/s41575-020-0275-y.
7. Ma Y, Dong M, Mita C, *et al.* Publication analysis on insomnia: how much has been done in the past two decades? *Sleep Med* 2015; 16: 820–826.

8. Wrafter PF, Connelly TM, Khan J, *et al*. The 100 most influential manuscripts in colorectal cancer: a bibliometric analysis. *Surgeon* 2016; 14: 327–336.
9. Blei D, Carin L and Dunson D. Probabilistic Topic models: a focus on graphical model design and applications to document and image analysis. *IEEE Signal Process Mag* 2010; 27: 55–65.
10. Blei DM, Ng AY, Jordan MI, *et al*. Latent Dirichlet allocation. *J Mach Learn Res* 2003; 3: 993–1022.
11. Chen X, Xie H, Wang FL, *et al*. A bibliometric analysis of natural language processing in medical research. *BMC Med Inform Decis Mak* 2018; 18(Suppl. 1): 14.
12. Neveol A and Zweigenbaum P. Clinical natural language processing in 2015: leveraging the variety of texts of clinical interest. *Yearb Med Inform* 2016; 1: 234–239.
13. Gal D, Thijs B, Glanzel W, *et al*. Hot topics and trends in cardiovascular research. *Eur Heart J* 2019; 40: 2363–2374.
14. Stout NL, Alfano CM, Belter CW, *et al*. A bibliometric analysis of the landscape of cancer rehabilitation research (1992–2016). *J Natl Cancer Inst* 2018; 110: 815–824.
15. Feng C, Wu Y, Gao L, *et al*. Publication landscape analysis on gliomas: how much has been done in the past 25 years? *Front Oncol* 2019; 9: 1463.
16. Jacomy M, Venturini T, Heymann S, *et al*. ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. *PLoS One* 2014; 9: e98679.
17. Heald RJ, Husband EM and Ryall RD. The mesorectum in rectal cancer surgery–the clue to pelvic recurrence? *Br J Surg* 1982; 69: 613–616.
18. Jacobs M, Verdeja JC and Goldstein HS. Minimally invasive colon resection (laparoscopic colectomy). *Surg Laparosc Endosc* 1991; 1: 144–150.
19. Jeong SY, Park JW, Nam BH, *et al*. Open versus laparoscopic surgery for mid-rectal or low-rectal cancer after neoadjuvant chemoradiotherapy (COREAN trial): survival outcomes of an open-label, non-inferiority, randomised controlled trial. *Lancet Oncol* 2014; 15: 767–774.
20. Sylla P, Rattner DW, Delgado S, *et al*. NOTES transanal rectal cancer resection using transanal endoscopic microsurgery and laparoscopic assistance. *Surg Endosc* 2010; 24: 1205–1210.
21. Marks JH, Myers EA, Zeger EL, *et al*. Long-term outcomes by a transanal approach to total mesorectal excision for rectal cancer. *Surg Endosc* 2017; 31: 5248–5257.
22. Li Y, Zhao L, Gungor C, *et al*. The main contributor to the upswing of survival in locally advanced colorectal cancer: an analysis of the SEER database. *Therap Adv Gastroenterol* 2019; 12: 1756284819862154.
23. Sadot E, Groot KB, Leal JN, *et al*. Resection margin and survival in 2368 patients undergoing hepatic resection for metastatic colorectal cancer: surgical technique or biologic surrogate? *Ann Surg* 2015; 262: 476–85; discussion 483–485.
24. Sao JG, Habr-Gama A, Vailati BB, *et al*. New strategies in rectal cancer. *Surg Clin North Am* 2017; 97: 587–604.
25. Enriquez-Navascues JM, Borda N, Lizerazu A, *et al*. Patterns of local recurrence in rectal cancer after a multidisciplinary approach. *World J Gastroenterol* 2011; 17: 1674–1684.
26. Sideris M, Donaldson AN, Hanrahan J, *et al*. Radiotherapy may offer a recurrence and survival benefit in rectal cancers treated surgically with transanal endoscopic microsurgery: a systematic review and meta-analysis. *Anticancer Res* 2018; 38: 1879–1895.
27. Peery AF, Cools KS, Strassle PD, *et al*. Increasing rates of surgery for patients with non-malignant colorectal polyps in the United States. *Gastroenterology* 2018; 154: 1352–1360.e3.
28. Song M and Giovannucci E. Preventable incidence and mortality of carcinoma associated with lifestyle factors among white adults in the United States. *JAMA Oncol* 2016; 2: 1154–1161.
29. Zandberg DP, Bhargava R, Badin S, *et al*. The role of human papillomavirus in nongenital cancers. *CA Cancer J Clin* 2013; 63: 57–81.
30. Li R, Zhou R, Wang H, *et al*. Gut microbiota-stimulated cathepsin K secretion mediates TLR4-dependent M2 macrophage polarization and promotes tumor metastasis in colorectal cancer. *Cell Death Differ* 2019; 26: 2447–2463.
31. Arthur JC, Perez-Chanona E, Muhlbauer M, *et al*. Intestinal inflammation targets cancer-inducing activity of the microbiota. *Science* 2012; 338: 120–123.
32. Wang T, Cai G, Qiu Y, *et al*. Structural segregation of gut microbiota between colorectal cancer patients and healthy volunteers. *ISME J* 2012; 6: 320–329.

33. Wirbel J, Pyl PT, Kartal E, *et al*. Meta-analysis of fecal metagenomes reveals global microbial signatures that are specific for colorectal cancer. *Nat Med* 2019; 25: 679–689.

34. Rolf S, Heinz B, Werner H, *et al*. Preoperative versus postoperative chemoradiotherapy for rectal cancer. *J Clin Oncol* 2004; 351: 1926–1933.

35. Park JH, Yoon SM, Yu CS, *et al*. Randomized phase 3 trial comparing preoperative and postoperative chemoradiotherapy with capecitabine for locally advanced rectal cancer. *Cancer* 2011; 117: 3703–3712.

36. Stutley JE and Conway WF. Magnetic resonance imaging of the pelvis and hips. *Orthopedics* 1994; 17: 1053–1062.

37. Krdzalic J, Maas M, Gollub MJ, *et al*. Guidelines for MR imaging in rectal cancer: Europe versus United States. *Abdom Radiol (NY)* 2019; 44: 3498–3507.

38. Simmonds PC. Palliative chemotherapy for advanced colorectal cancer: systematic review and meta-analysis. Colorectal Cancer Collaborative Group. *BMJ* 2000; 321: 531–535.

39. Pei Q, Zhu H, Tan F, *et al*. Intravascular emboli is an independent risk factor for the prognosis of stage III colorectal cancer patients after radical surgery. *Oncotarget* 2014; 7: 57268–57276.

40. Glynne-Jones R, Wyrwicz L, Tiret E, *et al*. Rectal cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up [published correction appears in Ann Oncol. 2018 Oct 1;29(Suppl 4):iv263] [published correction appears in Ann Oncol. 2018 Oct;29 Suppl 4:iv263]. *Ann Oncol*. 2017; 28 (suppl_4):iv22-iv40. doi:10.1093/annonc/mdx224