

基于 SEER 数据库的乳腺癌研究的文献计量学分析

雷珍珍^{1,2}, 陈鹏³, 江思源¹, 于世博¹, 朱丽喆¹, 闫宇¹, 任予¹

(1.西安交通大学第一附属医院乳腺外科, 陕西 西安 710061;

2.商洛市中心医院乳甲外科, 陕西 商洛 726000;

3.三原县人民医院普外科, 陕西 三原 713800)

摘要:目的 分析近 10 年来 SEER 数据库中乳腺癌相关研究的科研产出情况及研究现状, 为今后的相关研究提供参考依据。方法 以 PubMed 数据库为统计来源, 统计 2010~2019 年基于 SEER 数据库的乳腺癌研究的相关文献, 对各年度乳腺癌相关研究的文献数量、国家、期刊、作者、高被引文献和研究热点进行分析。结果 共纳入 664 篇文献, 发表于 187 种期刊上, 近 10 年来文献发文量整体趋势呈波动性上升, 其中发文量最多的国家为美国 (444 篇, 占比 66.87%), 我国位居第 2 (150 篇, 占比 22.59%)。有 96 篇文章发表在 Breast Cancer Research And Treatment 上, 占总发文量的 14.46%, 大部分研究为回顾性研究, 研究热点主要集中于生存分析、临床研究、风险因素等方面。结论 美国是基于 SEER 数据库的乳腺癌相关研究领域的主导力量, 我国与美国相比还存在一定的差距, 我国的研究者应充分利用资源, 吸取国外优秀科研经验, 提高学术生产力, 增强我国的学术影响力。

关键词:乳腺癌; SEER 数据库; 学术影响力

中图分类号: R737.9; R311

文献标识码: A

DOI: 10.3969/j.issn.1006-1959.2021.01.005

文章编号: 1006-1959(2021)01-0015-06

A Bibliometric Analysis of Breast Cancer Research Based on SEER Database

LEI Zhen-zhen^{1,2}, CHEN Peng³, JIANG Si-yuan¹, YU Shi-bo¹, ZHU Li-zhe¹, YAN Yu¹, REN Yu¹

(1.Department of Breast Surgery, the First Affiliated Hospital of Xi'an Jiaotong University, Xi'an 710061, Shaanxi, China;

2.Department of Breast and Thyroid Surgery, Shangluo Central Hospital, Shangluo 726000, Shaanxi, China;

3.Department of General Surgery, Sanyuan County People's Hospital, Sanyuan 713800, Shaanxi, China)

Abstract: Objective To analyze the research output of breast cancer related research based on SEER database in the past 10 years, and to objectively analyze its research status, so as to provide reference for future related research. Methods Using PubMed database as the statistical source, statistics on breast cancer research related literature based on SEER database from 2010 to 2019, and analysis of the number of breast cancer related research literature, countries, journals, authors, highly cited literature and research hotspots in each year. Results A total of 664 articles were included in this study, which were published in 187 journals. In the past 10 years, the overall trend of the number of published papers showed a fluctuating rise. The United States (444 articles, accounting for 66.87%) was the largest number of published articles, and China ranked second (150 articles, accounting for 22.59%). 96 articles were published in the Breast Cancer Research And Treatment, accounting for 14.46% of the total number of articles published. Most of the studies were retrospective studies, and the research focuses were mainly on survival analysis, clinical research, risk factors, etc. Conclusion The United States is the leading force in breast cancer related research based on SEER database. There is still a certain gap between China and the United States. Therefore, Chinese researchers should make full use of resources, absorb foreign excellent scientific research experience, improve our academic productivity and enhance our academic influence.

Key words: Breast cancer; SEER database; Academic influence

乳腺癌(breast cancer)是女性最常见的恶性肿瘤之一,也是癌症死亡的主要原因之一。2018 年全球癌症统计显示,2018 年全球有超过 200 万例乳腺癌新发病例,居女性恶性肿瘤发病率首位,约占全年肿瘤新发病例的 11.6%,死亡病例高达 62 万^[1]。我国 2015 年女性乳腺癌新发病例达 26.8 万,死亡病例约 6.9 万^[2]。研究报道,乳腺癌已成为危害女性健康的发病率最高的恶性肿瘤^[3]。为减少人群癌症负担,美国国立卫生研究院下属的美国癌症研究所于 1973 年针对本国肿瘤患者建立了监测、流行病学和最终结果数据库(SEER),这是北美最具代表性的大型肿瘤数据库之一^[4]。SEER 数据库共有 18 个数据登记站,覆盖了大约 28%^[5]的美国人口,记录了美国 50

余年的已确诊肿瘤患者的人口统计数据 and 临床特征,并且不断更新随访结果,完善相关数据^[6]。因其纳入人群数量大、资料较为齐全、平台开放、操作便捷等优点,目前为止已为临床医师的循证实践及临床医学研究提供了系统的证据支持和宝贵的肿瘤疾病的相关资料,尤其是为恶性肿瘤和罕见肿瘤的研究提供了宽广的路径^[6]。近年来,国内外越来越多的研究者尝试利用统计学方法或机器学习方法对 SEER 数据库开展肿瘤的预后研究^[7]。在此背景下,本文以 PubMed 数据库刊载的文献为统计源,对 2010~2019 年基于 SEER 数据库的乳腺癌研究态势进行分析,为今后的相关研究提供参考依据。

1 资料与方法

1.1 数据来源 在 PubMed 数据库中检索 2010~2019 年基于 SEER 数据库的乳腺癌相关研究的文献。登录 PubMed 数据库 (<https://www.ncbi.nlm.nih.gov/PubMed>),进入高级检索,以文献标题、摘要中含“乳

作者简介:雷珍珍(1992.1-),女,陕西商南县人,硕士研究生,主要从事乳腺肿瘤研究

通讯作者:任予(1972.1-),男,陕西西安人,博士,主任医师,硕士生导师,主要从事乳腺疾病专业理论和学科发展研究

腺癌”及“SEER”相关词汇的文献作为研究对象,数据获取采用尽可能提高查全率的检索思路。检索式:((((Breast Cancer[Title/Abstract]) OR Breast Neoplasm [Title/Abstract]) OR Breast carcinoma [Title/Abstract]) OR Mammary Neoplasm [Title/Abstract]) AND SEER [Title/Abstract]) AND (“2010/01/01”[Date - Publication] : “2019/12/31”[Date - Publication])。检索时间为2020年3月10日,共检索出812篇文献。

1.2 纳入与排除标准 纳入标准:①基于SEER数据库的乳腺癌相关的临床研究;②能提供足够信息的摘要或可以获取全文资料的文献。排除标准:①不在研究时间内的文献;②无法获取作者的文献;③无法获取摘要的文献;④其他与研究目的不相符的文献。数据筛选是由2位作者独立从所有符合条件的文献中仔细提取的。

1.3 方法 通过Excel 2016对文献内容进行分类整理,对纳入文献各年度的发文量、国家、期刊、作者、高被引文献和研究热点等进行统计分析。使用VOSviewer生成基于SEER数据库的乳腺癌的相关研究的文本主题关键词的知识地图。

表1 2010~2019年基于SEER数据库乳腺癌研究相关文献发文量排名前10的国家

Rank	Country	Count	Percent (%)	Cumulative percentage (%)
1	United States	444	66.87	66.87
2	China	150	22.59	89.46
3	Canada	12	1.81	91.27
4	Egypt	7	1.05	92.32
5	France	6	0.90	93.22
6	Korea	6	0.90	94.12
7	Australia	5	0.75	94.87
8	Germany	4	0.60	95.47
9	Italy	4	0.60	96.07
10	England	4	0.60	96.67

2.3 文献期刊分析 664篇文献来源于187种期刊杂志,文献数排名前10的期刊见表2。近一半的文献发表于排名前10的期刊上(288篇,占比43.39%)。除了英国的Breast期刊,其余9种期刊均为美国出版的;排名前3的期刊分别是:Breast Cancer Research And Treatment(IF 2018=3.471),发表文献96篇,占总发文量的14.46%,Journal of Clinical Oncology(IF 2018=28.245),发表文献28篇,占总发文量的4.22%以及Oncotarget(IF 2018=0),发表文献27篇,占总发文量的4.07%。

2.4 文献作者分析 664篇文献的作者共有2842位,以第一作者和并列一作发文量为研究对象,依据总被引频次进行数据整理统计分析,结果得出发文量3篇以上的8位作者共发表文献29篇,占该领域所

2 结果

2.1 文献发表年限分析 最终纳入664篇文献,其年度发文量及变化趋势见图1。从各年度的文献量分布上可以看出,近10年来基于SEER数据库的乳腺癌研究总体趋势呈波动性增长。



图1 2010~2019年PubMed中基于SEER数据库乳腺癌研究相关文献的发文量

2.2 发文作者国家分析 对明确注明作者国家的文献进行分析,结果显示664篇文献主要分布于26个国家和地区,其中美国的发文量最多,其次为中国和加拿大,见表1。

有已发表文献的9.42%,见表3。8位作者中美国作者最多,共有6位,中国和埃及各1位。排名第1的作者为美国的Jagsi R,共发表了6篇文献,总被引次数为190次;排名第2和第3的作者为分别为埃及的Abdel-Rahman(4篇)和美国的Agarwal S(4篇)。

2.5 高被引文献分析 近10年被引量最高的10篇文献见表4。在排序前10的文献中,《US incidence of breast cancer subtypes defined by joint hormone receptor and HER2 status》以共计被引204次和年均被引60.90次位居第1位,《Differences in breast cancer stage at diagnosis and cancer-specific survival by race and ethnicity in the United States》共计被引124次居第3位,《Disparities in breast cancer characteristics and outcomes by race/ethnicity》以共计被

引 113 次位居第 3。

2.6 关键词研究热点分析 通过 VOSviewer 1.6.14 软件对检索到的 664 篇文献的文本主题中关键词的研究热点进行分析,在所有文章的标题和摘要中使用超过 20 次的关键词“危险因素”,见图 2。在“生存分析”数据集中包含了 49 个关键词,排序前 3 的关键词分别为 overall survival(153 次)、grade(110 次)和 prognosis(104 次);在“临床研究”数据集中包含了 41 个关键词,排序前 3 的关键词分别为 woman(350

次)、use(145 次)和 association(122 次);在“危险因素”数据集中包含了 35 个关键词,排序前 3 的关键词分别为 incidence(106 次)、period(83 次)和 invasive breast cancer(74 次)。另外,“overall survival”和“prognosis”作为关键词分别出现了 153 次和 104 次,这两个词都属于数据集“生存分析”,表明近年来研究者们越来越关注乳腺癌患者的生存状况。同时,关键字“woman”位于图 3 的中心,与其他词共出现 2988 次。密度图显示了关键字的引文集中区见图 4。

表 2 发文量排名前 10 的期刊

Rank	Journal	Count	IF 2018	Percent (%)	Cumulative percentage (%)
1	Breast Cancer Research And Treatment	96	3.471	14.46	14.46
2	Journal of Clinical Oncology	28	28.245	4.22	18.68
3	Oncotarget	27	0	4.07	22.75
4	Annals of Surgical Oncology	25	3.681	3.77	26.52
5	Clinical Breast Cancer	25	2.762	3.77	30.29
6	Cancer	21	6.102	3.16	33.45
7	PLoS One	18	2.776	2.71	36.16
8	Cancer Medicine	17	3.357	2.56	38.72
9	Breast	16	3.494	2.41	41.13
10	Breast Journal	15	2.433	2.26	43.39

表 3 2010~2019 年发文量 3 篇以上 8 位作者

Rank	Author	Author	Country	Cited times	Percent (%)
1	Jagsi R	6	United States	190	0.91
2	Abdel-Rahman	4	Egypt	17	0.60
3	Agarwal S	4	United States	44	0.60
4	Chavez-MacGregor M	3	United States	77	0.45
5	Churilla TM	3	United States	17	0.45
6	Katz SJ	3	United States	13	0.45
7	Reiner AS	3	United States	10	0.45
8	Wu Q	3	China	32	0.45

表 4 2010~2019 年被引次数排序前 10 位文献

Rank	Title
1	US incidence of breast cancer subtypes defined by joint hormone receptor and HER2 status.
2	Differences in breast cancer stage at diagnosis and cancer-specific survival by race and ethnicity in the United States.
3	Disparities in breast cancer characteristics and outcomes by race/ethnicity.
4	Breast Cancer Mortality After a Diagnosis of Ductal Carcinoma In Situ.
5	Characteristics associated with differences in survival among black and white women with breast cancer.
6	Incidence of breast cancer with distant involvement among women in the United States, 1976 to 2009.
7	Breast-Cancer Tumor Size, Overdiagnosis, and Mammography Screening Effectiveness.
8	Radiation-related mortality from heart disease and lung cancer more than 20 years after radiotherapy for breast cancer.
9	Underascertainment of radiotherapy receipt in Surveillance, Epidemiology, and End Results registry data.
10	Risk prediction model for heart failure and cardiomyopathy after adjuvant trastuzumab therapy for breast cancer.

表 4(续)

Rank	IF2018	First author	Journal	Year	Country	Total citations
1	10.211	Howlader N	J. Natl. Cancer Inst.	2014	United States	204
2	51.273	Iqbal J	JAMA	2015	Canada	124
3	3.471	Ooi SL	Breast Cancer Res. Treat.	2011	United States	113
4	22.416	Narod SA	JAMA Oncol	2015	Canada	109
5	51.273	Silber JH	JAMA	2013	United States	94
6	51.273	Johnson RH	JAMA	2013	United States	92
7	70.67	Welch HG	N. Engl. J. Med.	2016	Lebanon	85
8	5.416	Henson KE	Br. J. Cancer	2013	England	59
9	6.102	Jagsi R	Cancer	2012	United States	55
10	4.66	Ezaz G	J Am Heart Assoc	2014	United States	55

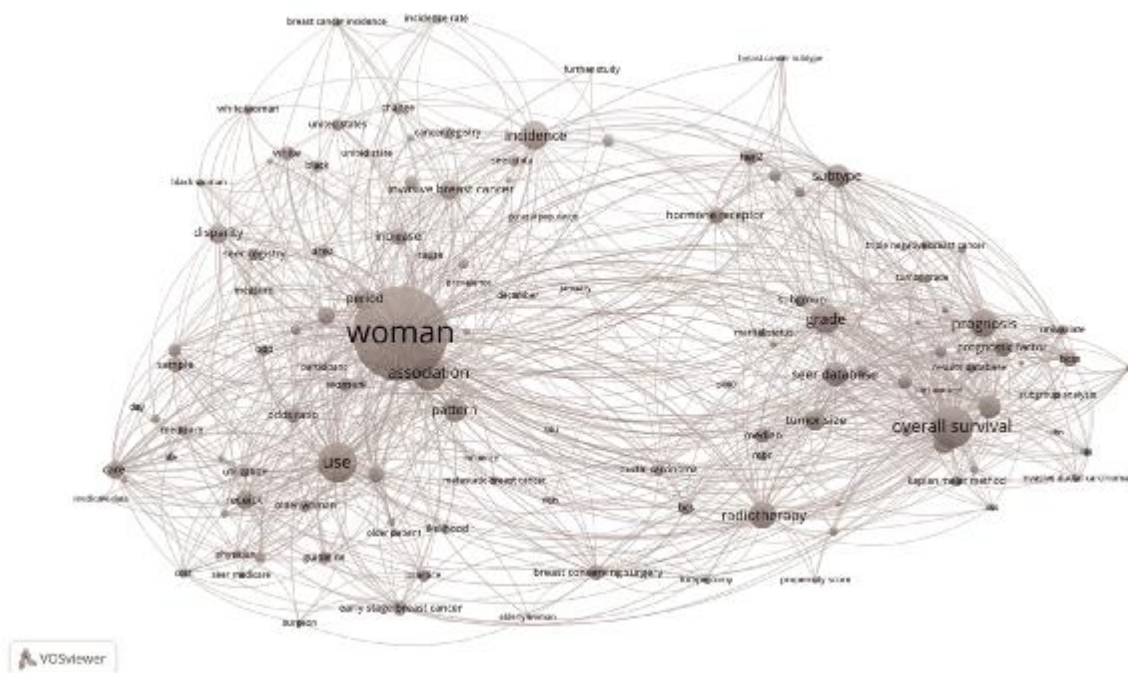


图 2 基于 SEER 数据库的乳腺癌研究领域论文中关键词的知识地图

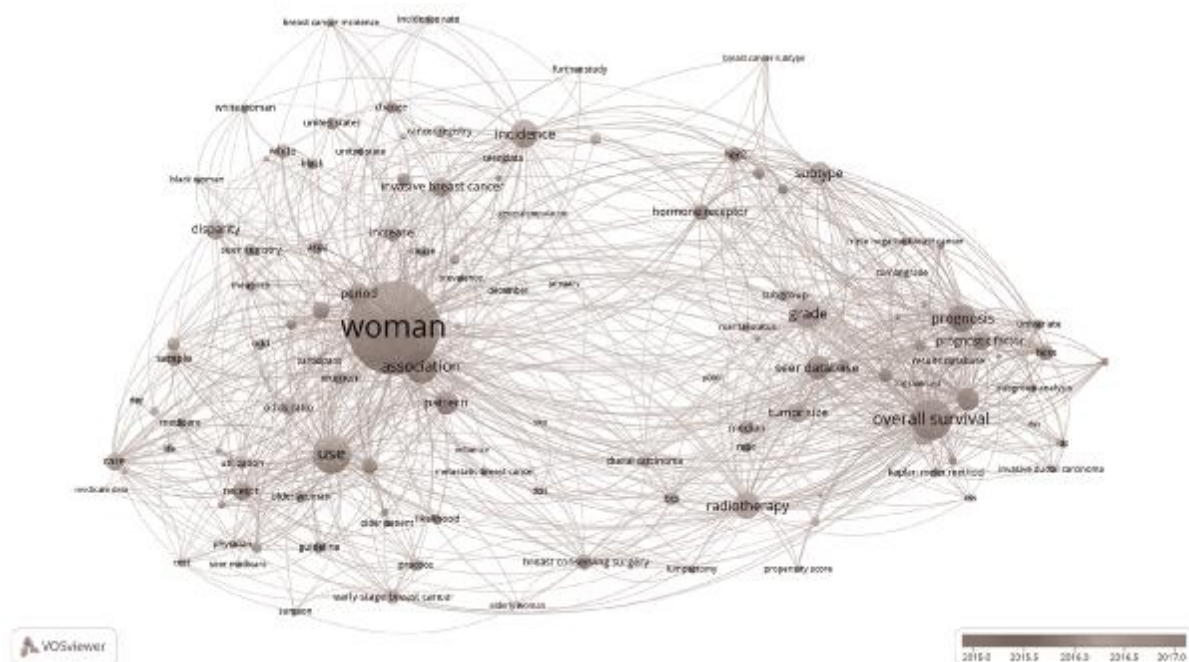


图 3 基于 SEER 数据库的乳腺癌研究领域论文中关键字的年份地图

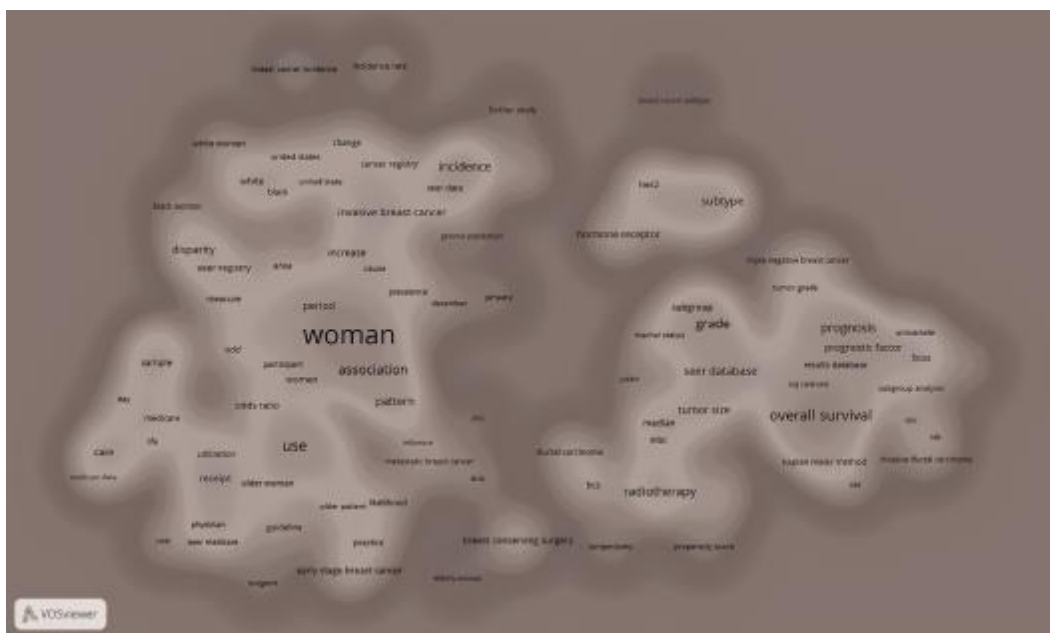


图 4 基于 SEER 数据库的乳腺癌研究领域论文中关键字的密度地图

3 讨论

文献计量学是一门运用数学和统计学方法对知识载体进行定量分析的交叉学科,利用文献计量学技术可以为医学研究人员提供一个快速、高效的获取医学文献信息的途径^[9]。PubMed 数据库由美国国家图书馆开发,自 1950 年以来已收集超过 2200 万份文献记录,涉及 40 多种语言,每年新增文献记录 60 多万条,被公认为世界上最权威的免费大型医学文献数据库^[9]。利用文献计量方法分析 PubMed 数据库所收录的基于 SEER 数据库的乳腺癌相关文献,可以较为全面准确的反映这一研究领域的现状和发展趋势。

从近 10 年 PubMed 数据库所收录的基于 SEER 数据库的乳腺癌相关文献的发文量看,总体呈波动性上升趋势,且近 3 年增长速度加快。这表明近年来科研人员对该领域的关注和研究越来越多,目前研究正处于稳定发展的阶段。由此推测,今后该研究方向文献仍将会保持一定的增长速度,进一步达到研究的成熟阶段。

从各国相关文献的发文量看,美国以发文量 444 篇,占比 66.87% 位居榜首,中国则以发文量 150 篇,占比 22.59% 位居第 2,虽然我国发文量已位居第 2,但与美国相比仍存在一定差距。分析其原因可能由于 SEER 数据库仅部分数据免费对外开放,这在一定程度上限制了国人对 SEER 数据库资源的使用;其次, PubMed 数据库仅收录我国乳腺癌研究领域影响力较高的期刊,其余发表在未被 PubMed 数据库收录的中文杂志中的文献无法统计在内,这也是本研究中我国学者发表相关文献所占比例低于美国的一个重要原因。但由于我国人口基数大,人均

发文量与美国相距甚远,因此我国的研究者应加大对该领域研究投入,吸取国外优秀科研经验,加强与国外权威科研机构合作,从而提高我国的学术生产力,增强我国的学术影响力。

该领域发文量排名前 10 的期刊相关发文量为 288 篇, 占总发文量的 43.39%, 可见发文期刊呈中心性聚集分布。该领域发文量最多的期刊为 **Breast Cancer Research And Treatment**, 有 96 篇, 占总发文量的 14.46%, 除了英国的 **Breast** 期刊, 其余 9 种期刊均为美国出版的, 说明美国在该领域的研究具有相当的地位和影响力。发文量 3 篇以上的 8 位作者共发表文献 29 篇, 占该领域所有已发表文献的 9.42%, 发文量 3 篇以上的 8 位作者中美国作者最多, 共有 6 位, 中国和埃及各 1 位, 说明中国学者在此领域仍需努力。

学术论文是学术成果的主要呈现方式,其被引用次数是衡量学术价值的一种尺度。通常情况下,一篇文章被引用次数越多,说明这篇文章的可参考度越高,其在领域内所具有的影响力往往也越大,受关注度越高^[10]。在高被引排序前 10 的文献中,我国未上榜,这再次表明,要提高我国研究的质量,需要给予更多的关注。

在科学评价计量研究方面,使用可视化知识图谱是众多学者的选择。Citespace 和 VOSviewer 是当前两款热门的知识图谱软件,研究发现 Citespace 软件在揭示学科的动态发展规律,发现学科的研究前沿上有一定的优势^[11],而对学科主题之间的关系进行清晰的呈现、或者数据量非常大的时候,可以选取由荷兰莱顿大学 Nees Jan van Eck 教授开发的 VOSviewer 软件绘制知识图谱^[12]。对文本主题中的主

题词及副主题词进行可视化分析,从可视化图中可以看出 2010~2019 年主要侧重于以下 3 个研究方向,“生存分析”“临床研究”及“危险因素”,表明近年来乳腺癌领域的研究热点集中于此。生存分析:乳腺癌是女性常见恶性肿瘤之一,其发病率呈逐年上升趋势,但乳腺癌总体预后较好,尤其是早期病例^[13]。本研究可以看出近年来越来越多的研究都在关注乳腺癌预后生存状况。临床研究:本研究显示,基于 SEER 数据库的乳腺癌早期研究中主要热点为乳腺癌患者的临床研究,关键词包括“woman”“retrospective cohort study”和“seer medicare data”等,这表明 SEER 数据库作为北美最大最具代表性的数据库之一,越来越多的受到医务工作者的重视,提示我们可以建立属于自己的数据库,为广大医务工作人员提供强有力的数据支持。危险因素:乳腺癌患者的危险因素关键词包括“incidence”“ethnicity”和“higher risk”等,以往的研究表明乳腺癌的危险因素与年龄^[14]、种族^[15]、婚姻^[16]、饮食习惯、肥胖^[17]、生育状况^[18]、精神压力^[19]等密切相关。

尽管 SEER 数据库为我们提供了大量宝贵的临床资料,但不可否认的是,由于 SEER 数据库及 PubMed 数据库自身的局限性,目前的研究尚存在着一些不足之处。一方面 SEER 数据库未能提供患者接受内分泌治疗以及详尽的化疗方案信息,有关 HER-2 表达情况于 2010 年开始加入,对于患者复发转移的情况没有公开等,这给研究者带来了一定的局限性,但也提示我们可以建立属于自己的数据库,为广大医务工作人员提供强有力的数据支持。另一方面 PubMed 是使用最广泛的文献计量分析数据源,但它并不包含所有生物医学期刊,偏向于英文期刊;此外,通讯作者的地址不一定与进行研究的国家有关,如许多来自美国以外的学者在美国发表论文。

综上所述,本研究发现,美国是基于 SEER 数据库的乳腺癌相关研究领域的主导力量,我国与美国相比还存在一定的差距,我国的研究者应充分利用资源,吸取国外优秀科研经验,提高学术生产力,增强我国的学术影响力。

参考文献:

- [1]Bray F,Ferlay J,Soerjomataram I,et al.Global cancer statistics 2018:GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries[J].CA,2018,68(6):394-424.
- [2]Siegel R,Miller K,Jemal A.Cancer statistics,2020 [J].CA,2020,70(1):7-30.
- [3]Pan Q, Ji J. Diagnostic value of ultrasound combined with magnetic resonance imaging in different stages of breast cancer [J].Oncology Letters,2019,17(1):209-214.
- [4]杨津,柳青青,耿辉,等. SEER 数据库的申请及数据提取方法

与流程[J].中国循证心血管医学杂志,2018(7):781-784.

- [5]Li Y,Chen M,Pardini B,et al.The role of radiotherapy in metaplastic breast cancer:a propensity score-matched analysis of the SEER database[J].Journal of Translational Medicine,2019,17(1):318.
- [6]Lin S,Liu C,Tao Z,et al.Clinicopathological characteristics and survival outcomes in breast carcinosarcoma:A SEER population-based study [J].Breast (Edinburgh,Scotland),2020 (49):157-164.
- [7]章鸣媛,张璇,郭欣,等.基于 SEER 数据库利用机器学习方法分析乳腺癌的预后因素[J].北京生物医学工程,2019(5):486-491,497.
- [8]刘秋霞,郑景辉,戴铭,等.基于 PubMed 数据库的传统医药治疗肝肿瘤文献计量分析[J].世界科学技术-中医药现代化,2018(9):1515-1522.
- [9]丁荣楣,王平,马丽君,等.2006-2015 年基于 PubMed 数据库的鼻咽癌文献计量分析[J].预防医学情报杂志,2017(6):607-612.
- [10]杨成会,戚厚兴,孙大卫,等.《中华诊断学电子杂志》2013 至 2019 年高被引论文特征分析 [J].中华诊断学电子杂志,2020 (1):68-72.
- [11]陈莹,黄春娟.国外知识可视化研究的现状与启示--基于 Citespace 和 Vosviewer 的知识图谱分析 [J].大学图书馆学报,2015(4):93-99.
- [12]付健,丁敬达.Citespace 和 VOSviewer 软件的可视化原理比较[J].农业图书情报,2019(10):31-37.
- [13]陈莉莉,石菊芳,刘玉琴,等.基于人群的乳腺癌预后参数研究现状[J].中华乳腺病杂志(电子版),2018(6):370-372.
- [14]段骄楠,李小龙,陈首英,等.女性乳腺癌危险因素研究新进展[J].中国妇幼保健,2016(12):2571-2573.
- [15]Dania V,Liu Y,Ademuyiwa F,et al.Associations of race and ethnicity with risk of developing invasive breast cancer after lobular carcinoma in situ[J].BCR,2019,21(1):120.
- [16]Hinyard L,Wirth L,Clancy J,et al.The effect of marital status on breast cancer-related outcomes in women under 65:A SEER database analysis[J].Breast(Edinburgh,Scotland),2017(32):13-17.
- [17]Befort C,Kimber B,Bantis L,et al.Effects of Weight Loss and Weight Regain on Circulating Biomarkers in Overweight/Obese Breast Cancer Survivors Enrolled in a Weight Loss Trial in the Rural Midwest[J].Cancer Epidemiology,Biomarkers&Prevention,2020,29(7):1321-1328.
- [18]Letourneau J,Wald K,Sinha N,et al.Fertility preservation before breast cancer treatment appears unlikely to affect disease-free survival at a median follow-up of 43 months after fertility-preservation consultation[J].Cancer,2020,126(3):487-495.
- [19]Cvetkovic J,Ivanovic Kovacevic S,Cvetkovic M,et al.Evaluation of the role of stress in patients with breast cancer and depression by paykel's life event and adaptive neuro-fuzzy approach[J].Brain and Behavior,2020,10(4):e01570.

收稿日期:2020-09-02;修回日期:2020-09-10

编辑/成森