

医疗保险数据与传统登记途径对肿瘤病例的捕获比较分析:以内蒙古队列人群为例

赖雪峰¹, 乔丽颖², 杜润茗¹, 杨腾¹, 吴宜瞳¹, 吕晓珍³, 董彬¹, 黄婧¹, 杨雷⁴, 金音子¹, 席云峰², 王胜锋¹

(1. 北京大学公共卫生学院, 北京 100191; 2. 内蒙古疾病预防控制中心慢性病预防控制所, 内蒙古呼和浩特 010031; 3. 北京大学第六医院, 北京 100191; 4. 北京大学肿瘤医院, 北京 100142)

摘要: [目的] 对比医疗保险数据和传统肿瘤登记数据捕获肿瘤病例的能力, 探索整合医疗保险数据收集人群肿瘤信息的可行性。 [方法] 研究基于内蒙古自治区的呼和浩特和通辽两市 70 010 名 40~74 岁的户籍居民队列人群, 分别采用传统肿瘤登记途径和同时期医疗保险数据途径收集肿瘤新发病例, 以两者从基线调查时间至 2019 年捕获的全部肿瘤病例作为全集, 分别计算 2 种捕获方式的捕获率及漏报率, 并对漏报患者的险种等特征进行比较。 [结果] 2 个数据库共捕获肿瘤患者 365 例, 使用捕获-再捕获法估计的患者总数为 571 例, 估计内蒙古自治区呼和浩特市和通辽市 2019 年肿瘤截缩发病率 (42~76 岁) 为 378.92/10 万。医疗保险和传统肿瘤登记数据分别捕获肿瘤患者 328 例和 86 例, 捕获率分别为 89.86% 和 23.56%。在肿瘤登记数据中, 享有城镇职工基本医疗保险的患者相较于享有城乡居民基本医疗保险的患者漏报率更高 (85.19% vs 72.73%, $P=0.007$), 而在医疗保险数据中, 享有城镇职工基本医疗保险的患者相对漏报更少 (2.47% vs 8.08%, $P=0.030$)。 [结论] 医疗保险数据途径可有效捕获肿瘤患者, 可作为传统肿瘤登记途径收集肿瘤新发病例的重要补充。

关键词: 医疗保险数据; 肿瘤登记; 漏报率; 内蒙古

中图分类号: R73-31 文献标识码: A 文章编号: 1004-0242(2022)03-0197-08

doi: 10.11735/j.issn.1004-0242.2022.03.A005

Comparative Analysis of Cancer Case Capture Based on Medical Insurance Claim Data and Cancer Registry Data: Based on a Population-based Cohort from Inner Mongolia

LAI Xuefeng¹, QIAO Liying², DU Runming¹, YANG Teng¹, WU Yitong¹, LYU Xiaozhen³, DONG Bin¹, HUANG Jing¹, YANG Lei⁴, JIN Yinzi¹, XI Yunfeng², WANG Shengfeng¹

(1. School of Public Health, Peking University, Beijing 100191, China; 2. Inner Mongolia Center for Disease Control and Prevention, Hohhot 010031, China; 3. The Sixth Hospital of Peking University, Beijing 100191, China; 4. Peking University Cancer Hospital, Beijing 100142, China)

Abstract: [Purpose] To investigate the the feasibility of integrating medical insurance claim data for cancer information of the whole population through comparison of claim data and cancer registry data to capture cancer cases. [Methods] The follow-up data of cancer registry and medical insurance claim data from baseline time to 2019 based on a cohort of 70 010 urban residents aged 40~74 years from Hohhot and Tongliao of Inner Mongolia Autonomous Region were collected. The cancer cases captured by the two databases were used as the complete set to calculate the capture rate and the omission rate, and the characteristics such as categories of insurance of under-report cases were analyzed. [Results] A total of 365 cancer patients were identified from the two databases from baseline time to 2019, and the estimated patients by capture-recapture was 571. The estimated truncate cancer incidence (42~76 years old) in Hohhot and Tongliao in 2019 was 378.92/10⁵. The claim data and cancer registry data identified 328 and 86 cancer patients, and the capture rates were 89.86% and 23.56%, respectively. In cancer registry data the omission rate of patients with basic medical insurance for urban employees was higher than that of patients with basic medical insurance for urban and rural residents (85.19% vs 72.73%, $P=0.007$), while in medical insurance claim data the missing report rate of patients with basic medical insurance for urban employees was lower (2.47% vs 8.08%, $P=0.030$). [Conclusion] Medical insurance claim data is an efficient source for identifying cancer cases and can be used as an important supplement to traditional cancer registry data to capture new cancer cases.

Key words: claim data; cancer registration; omission rate; Inner Mongolia

收稿日期: 2021-09-27; 修回日期: 2021-12-01

基金项目: 国家重点研发计划“重大慢性病非传染性疾病防控研究”重点专项 (2018YFC1311704); 中央引导地方科技发展资金项目 (2020ZY0015); 内蒙古自治区自然科学基金 (2021MS08039)

通信作者: 王胜锋, E-mail: shengfeng1984@126.com 席云峰, E-mail: xiyunfeng210@163.com

以人群为基础的肿瘤登记是规划和评价我国癌症预防、诊断与治疗成效的主要数据支持。2019年我国提出,要在2022年底前,建成覆盖全国所有区县的癌症病例登记系统^[1]。截至2020年底,我国肿瘤登记工作已覆盖1152个区县,覆盖人口达5.98亿人,约占全国人口的42.7%^[2]。

医疗保险数据通常由政府或医疗服务机构收集,包含患者人口学、疾病诊断、医疗服务及费用等丰富信息,收集成本较低且覆盖范围广^[3],与一般调查数据相比,具有样本量大、可避免选择偏倚及回忆偏倚^[4]、可估计发病率^[5]等优势。目前我国稳定在95%以上的参保率又保证了医疗保险数据的一般人群代表性^[6]。随着大数据时代的到来,利用医疗保险数据开展肿瘤登记工作或将是一种低成本、高效率的新模式^[7]。近年来,一些国家和地区已有不少研究应用医疗保险数据捕获肿瘤患者^[5,8-11]。尽管我国大陆地区医疗保险数据的构建和应用起步晚,也已有学者尝试利用医疗保险数据捕获肿瘤患者^[7,12-13]、估计肿瘤发病率^[14]和肿瘤经济负担^[15],并且发现新型农村合作医疗数据对肿瘤病例具有良好的捕获能力,提示医疗保险数据可作为传统肿瘤登记漏报信息的补充来源。但目前的研究局限于单一险种和农村地区,仍缺乏险种多元的城市地区的相关数据。

本研究基于内蒙古自治区呼和浩特市和通辽市的一项人群队列研究,对比医疗保险数据和传统肿瘤登记数据捕获肿瘤病例的情况,以了解我国医疗保险数据捕获肿瘤病例的能力,为探索整合医疗保险数据的人群肿瘤信息收集方式提供依据。

1 资料与方法

1.1 研究对象

本研究基于内蒙古自治区呼和浩特和通辽两市参与城市癌症早诊早治项目的70010位居民。该项目于2017—2018年在呼和浩特市和通辽市开展,采用整群随机抽样方法,随机抽取6个社区,以社区中所有40~74岁且在该地区居住3年及以上的常住居民为调查对象,排除已确诊恶性肿瘤的患者及有其他严重内外科疾病正在治疗的患者^[16]。项目收集居民的基本信息、既往疾病史和各种行为危险因素相关信息作为基线数据库。本研究纳入参与上述项

目,且享有当地医疗保险的居民。研究排除了医疗保险数据中出院诊断为肿瘤类疾病,但确诊肿瘤类疾病的报销记录对应的入院时间在基线时间之前的调查对象。

1.2 肿瘤结局的捕获

1.2.1 通过肿瘤登记捕获肿瘤病例

两市肿瘤登记的发病数据来源主要为各级医疗机构上报的肿瘤发病信息和疾病预防控制中心提供的居民死亡原因监测数据进行的死亡补发病信息,此外,也有部分旗县的登记点不定期地核对当地医疗保险数据进行补充。将两市自基线调查时至2019年期间肿瘤登记上报的病例通过问卷调查编码与基线队列人群进行关联,共关联到112例肿瘤患者的肿瘤诊断及死亡相关详细信息。

1.2.2 通过医疗保险数据捕获肿瘤病例

两市2010—2019年居民医疗保险数据通过姓名、性别、民族及出生日期作为唯一识别变量组合与基线队列人群进行匹配,将入院时间在2010年至基线调查时间期间的数据用于剔除基线前确诊的病例,以确保基线调查时间至2019年捕获的患者为新发肿瘤病例。所用居民医疗保险数据包含患者个人基本信息、就诊时间、出院诊断、医疗保险类型等信息,基线调查时间至2019年的就诊库共有14605例患者的报销记录。

由具备相关临床背景的数据处理人员根据《国际疾病分类》第10版(ICD-10)对疾病诊断进行标准化处理。将疾病诊断编码为“C”码或诊断名称中包含以下关键词的患者确定为肿瘤病例:①癌、恶性肿瘤或肿瘤;②淋巴瘤、淋巴肉瘤、白血病、骨髓瘤以及胸膜间皮瘤等恶性肿瘤;③中枢神经系统恶性肿瘤。符合以下诊断者确定为非肿瘤患者:①无癌症或肿瘤相关诊断;②诊断为良性肿瘤(除外中枢神经系统良性肿瘤);③诊断为血管瘤、错构瘤等先天性良性肿瘤或正常组织的类瘤样畸形。

上述关键词由肿瘤临床专家确定。为避免错误纳入,对两种数据源所识别出肿瘤病例,安排2名具备相关临床知识的人员平行核查,不一致处由第三人裁定。

1.3 统计学处理

医疗保险数据和肿瘤登记捕获的基线至2019年的肿瘤病例为全集。以全集为参照,将该数据库未

捕获的肿瘤病例定义为漏报,漏报病例数在全集中的占比定义为该数据库的漏报率。通过捕获-再捕获(capture-recapture)方法,以 Lincoln-Petersen 提出的公式($N = \left[\frac{(a+1)(b+1)}{c+1} \right] - 1$),其中,a 为数据库 1 捕获的病例数,b 为数据库 2 捕获的病例数,c 为 2 个数据库同时捕获的病例数,N 为目标人群肿瘤患者总数^[17-18],分别估计观察期间和 2019 年样本人群的新发恶性肿瘤病例总数,并以估计的 2019 年所有新发肿瘤病例数为分子,基线数据库中 2019 年初未患肿瘤的样本数为分母,计算 2019 年样本人群恶性肿瘤截缩发病率(42~76 岁)。本研究重点关注内蒙古城市地区分性别的发病率前 5 的恶性肿瘤,即肺癌、肝癌、结直肠癌、胃癌、食管癌、乳腺癌和甲状腺癌 7 种^[19-20]。报告医疗保险数据与肿瘤登记对于所有肿瘤及关注的 7 种肿瘤的漏报率,并比较漏报患者的险种等特征是否存在差异。

计量资料采用 $\bar{x} \pm s$ 表示,计数资料采用百分率(%)来描述。对于不同人口学和险种特征患者的漏报情况差异,分类变量使用卡方检验或 Fisher 精确概率检验,连续变量使用独立样本均数 t 检验。检验水平设为双侧 $\alpha=0.05$ 。统计软件使用 Stata 16.0。

2 结果

2.1 基本情况

呼和浩特市和通辽市共收到有效问卷 70 010 份,经去重和核查,并排除基线前确诊肿瘤的 316 例患者,最终剩余样本 69 590 人。其中,男性 32 260 人(46.36%),蒙古族 7 877 人(11.32%),年龄在 40~74 岁,平均年龄(55.66±9.01)岁(Table 1)。

2 种方式共捕获肿瘤患者 365 例(Figure 1),通过捕获-再捕获法估计病例总数为 571 例。2019 年全年新发肿瘤患者 183 例

(医保 175 例,肿瘤登记 23 例,其中共同捕获 15 例),估计 2019 年新发病例总数为 263 例,2019 年初未确诊肿瘤的居民为 69 408 人,利用该来源病例数估计 2019 年目标人群恶性肿瘤截缩发病率(42~76 岁)为 378.92/10 万。

2.2 医疗保险数据和肿瘤登记数据捕获结果对比

两种数据源共捕获的 365 例肿瘤患者中,通过医疗保险数据捕获 328 例,捕获率 89.86%,通过肿

Table 1 Basic characteristics of baseline survey of urban residents aged 40~74 years old in Hohhot and Tongliao[n(%)]

Characteristic	N
Total	69590(100.00)
Gender	
Female	37330(53.64)
Male	32260(46.36)
Mean age (years old)	55.66±9.01
Ethnicity	
Han	59982(86.19)
Mongolian	7877(11.32)
Others	1731(2.49)
Education	
Junior high and below	36828(52.91)
High school and above	32762(47.09)
Marriage	
Not in marriage	1976(2.84)
Married	67614(97.16)

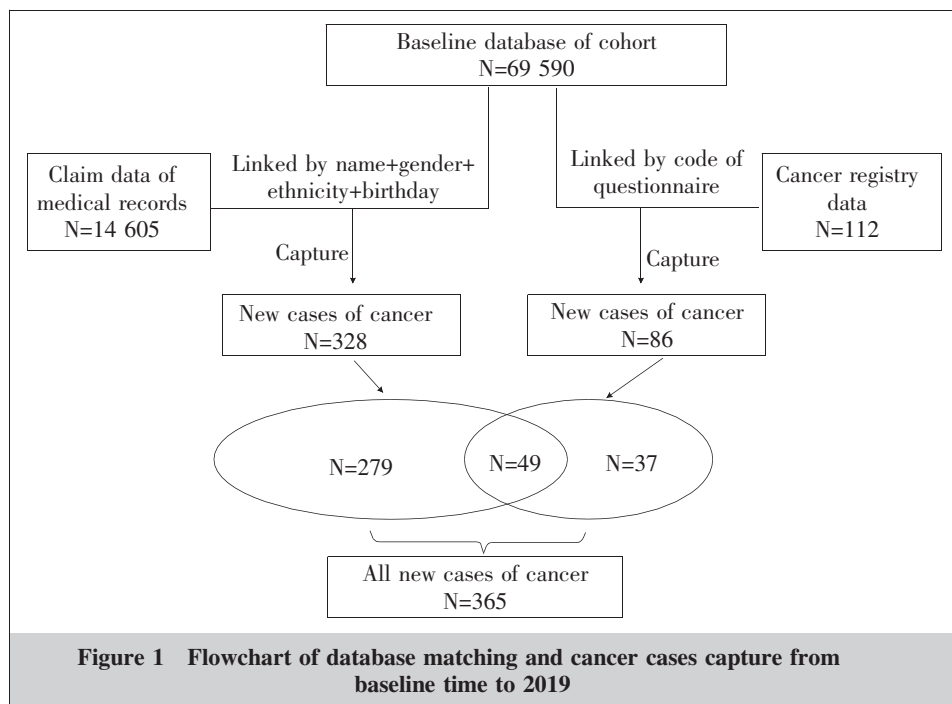


Figure 1 Flowchart of database matching and cancer cases capture from baseline time to 2019

瘤登记数据捕获 86 例,捕获率 23.56%。2 种数据源同时捕获肿瘤患者 49 例(13.42%)(Figure 1)。

在所有肿瘤病例中,肿瘤登记漏报 279 例,漏报率为 76.44%, 医疗保险数据漏报 37 例,漏报率为 10.14%。肿瘤登记对 7 种肿瘤的漏报率均在 50%及以上(Table 2)。

2.3 肿瘤登记数据和医疗保险数据对不同特征患者的捕获和漏报情况分析

肿瘤登记和医疗保险数据对肿瘤患者的捕获和漏报情况在不同人口学特征的人群中无显著差异,

但在不同医疗保险类型的患者中存在差异。肿瘤登记数据中,相较于城乡居民基本医疗保险,享有城镇职工基本医疗保险的患者漏报更多($P=0.007$)。而在医疗保险数据,享有城乡居民基本医疗保险的患者漏报更多($P=0.030$)(Table 3)。

3 讨论

自 2002 年全国肿瘤登记中心成立以来,基于人群的肿瘤登记工作逐步规范,成为我国收集肿瘤病例信息的主要方法,为肿瘤相关研究和国家公共卫生策略的制定提供了重要依据。随着医疗保险的全民覆盖和大数据时代的到来,国内外学者开始探索高效率、低成本的基于医疗保险数据的肿瘤信息收集模式,以完善肿瘤登记系统的建设。本研究通过比较呼和浩特市和通辽市医疗保险数据和传统肿瘤登记数据捕获肿瘤病例的情况,探讨医疗保险数据在识别肿瘤病例上的能力。结果显示,医疗保险数

Table 2 Omission of different cancers in cancer registry and claim data

Cancer category	Cases	Cancer registry		Claim data	
		Cases	Omission rate (%)	Cases	Omission rate (%)
All cancers	365	279	76.44	37	10.14
Lung cancer	59	46	77.97	6	10.17
Colorectal cancer	57	42	73.68	3	5.26
Breast cancer	57	44	77.19	3	5.26
Esophagus cancer	22	15	68.18	5	22.73
Thyroid cancer	29	22	75.86	3	10.34
Gastric cancer	19	17	89.47	1	5.26
Liver cancer	18	9	50.00	5	27.78

Table 3 Comparison of patients being captured and missed in cancer registry data and claim data[n(%)]

Basic characteristic	Cancer registry data			Claim data		
	Capture(n=86 ^a)	Omission(n=279)	P	Capture(n=328)	Omission(n=37 ^a)	P
Gender						
Female	47(21.56)	171(78.44)	0.272	201(92.20)	17(7.80)	0.071
Male	39(26.53)	108(73.47)		127(86.39)	20(13.61)	
Mean age (years old)	59.86±8.05	58.85±9.03	0.354 ^a	58.92±8.93	60.61±7.64	0.269 ^a
Education						
Junior high and below	48(24.12)	151(75.88)	0.783	176(88.44)	23(11.56)	0.325
High school and above	38(22.89)	128(77.11)		152(91.57)	14(8.43)	
Marriage						
Not in marriage	82(23.36)	269(76.64)	0.748 ^b	306(89.47)	36(10.53)	1.000 ^b
Married	4(28.57)	10(71.43)		13(92.86)	1(7.14)	
Occupation						
Technician/office worker	21(20.00)	82(80.00)	0.478	93(88.57)	12(11.43)	0.872
Worker/farmer/merchant	39(26.53)	108(73.47)		133(90.48)	14(9.52)	
Others	26(23.01)	87(76.99)		102(90.27)	11(9.73)	
Types of basic medical insurance						
For urban workers	36(14.81)	207(85.19)	0.007	230(97.53)	6(2.47)	0.030 ^b
For urban/rural residents	27(27.27)	70(72.73)		89(91.92)	8(8.08)	
Medical services						
Outpatient	2(20.00)	8(80.00)	0.378	10(100.00)	0	-
Inpatient	52(17.39)	247(82.61)		285(95.32)	14(4.68)	
Others	9(27.27)	24(72.73)		33(100.00)	0	

Notes: a; independent-samples *t* test; b; Fisher exact probability; *: among patients who were captured by cancer registry but missed by claim data, 23 patients had no reimbursement record, and information of medical insurance type and medical services was missing

据对于城市地区的肿瘤病例具有较高的捕获率。

本研究发现,呼和浩特市和通辽市传统肿瘤登记漏报率较高,漏报情况与一些较为完善的登记点早期的数据相比尚有很大差距^[21-23]。我国传统的登记信息采集经历较多人工环节,容易造成数据失实和缺漏,加之部分肿瘤登记地区仅收集住院患者的资料^[24],从而导致漏报。既往文献表明,当年的肿瘤新发病例数据需要经过3年的补充漏报才能稳定^[25]。此外,我国肿瘤登记系统最初多集中于肿瘤高发和资源较好的地区,西部地区由于社会经济发展和医疗资源的限制,肿瘤登记发展较慢,普遍存在数据采集、识别和提取手段滞后^[26],数据时效性差,随访效率低下^[25]等方面的问题。而内蒙古自治区的肿瘤登记工作从2009年起陆续开展,本研究关注的呼和浩特市和通辽市的肿瘤登记于2017年方才起步,尚处于发展初期,漏报率相对更高。为减少错报、漏报,既往研究提示,可通过提高报告过程的自动化程度改善数据质量,同时降低昂贵的劳动力成本^[27]。此外,在肿瘤登记较不完善的地区,通过肿瘤登记数据与医疗保险数据、死因登记或病理系统等其他数据库连接,从而补充患者的肿瘤相关信息、减少漏报,也是完善传统肿瘤登记工作的重要思路^[26]。

本研究以估计的2019年新发肿瘤病例数估计当年目标人群截缩发病率(42~76岁)为378.92/10万,稍高于2016年内蒙古城市地区40~69岁人群肿瘤粗发病率(330.66/10万),可能原因是本研究人群的年龄和年龄范围都更大,相对更多的纳入了肿瘤高发的老年人群。在本研究中,以捕获-再捕获法估计的2017年、2018年和2019年肿瘤患者总数计算的发病密度分别为510.70/10万、385.18/10万和375.26/10万人年,以通过医疗保险数据识别的2017年、2018年和2019年新诊断肿瘤患者计算的发病密度分别为283.04/10万、211.35/10万和249.70/10万人年,结果显示2017—2019年期间两者估计的发病密度逐渐接近,说明随着医疗保险全面覆盖工作的推进以及医保数据管理的逐渐规范化^[28],医疗保险数据的完整性不断提高,对肿瘤病例的识别能力也逐渐提升,提示在稳定的医疗保险数据库中采用主动抓取的方式捕获肿瘤病例是较为可靠的手段。此外,研究结果显示,相比于传统肿瘤登记周期性开展被动随访的肿瘤病例捕获模式,医疗保险数据可

通过患者的就诊和报销记录逐月收集患者信息,时效性更好。若结合2种模式开展肿瘤监测工作,通过医疗保险数据常规获取患者信息,同时开展肿瘤登记工作互为补充,就可在原来的基础上延长开展肿瘤登记被动随访的周期,从而降低成本,提高工作效率。

本研究结果显示,多种来源的医疗保险数据对城市地区肿瘤患者有着较高的捕获率。进一步分析发现,医疗保险数据和传统肿瘤登记对不同险种患者的捕获情况存在差异,医疗保险数据对城镇职工基本医疗保险参保患者的捕获率高于城乡居民基本医疗保险的参保患者,而传统肿瘤登记对城镇职工基本医疗保险参保患者的捕获率低于城乡居民基本医疗保险的参保患者。从不同医疗保险本身的特征来看,城镇职工基本医疗保险和城乡居民基本医疗保险面向的人群和待遇标准不同,前者主要面向有工作单位或从事个体经济的在职职工和退休人员,其平均经济水平高于后者的参保人群,加之职工基本医疗保险补偿额度高^[29],参保患者异地就医的倾向性更高,而异地就医的患者可由医疗保险数据捕获,却可能会被肿瘤登记遗漏,从而导致2种险种患者捕获情况出现差异。

本研究提取呼和浩特市和通辽市城镇职工基本医疗保险和城乡居民基本医疗保险数据开展研究,为我国应用医疗保险数据识别肿瘤病例进行资料补充,同时填补了险种多元的城市地区相关研究的空白^[30]。美国等发达国家的医疗保险数据发展相对完善,基于医疗保险数据识别肿瘤病例的准确性研究也较多^[31-34]。近几年,我国也开始出现使用医疗保险数据进行肿瘤病例识别及流行病学相关研究^[7,12,14-15]。河南滑县的一项研究^[12]对比了新型农村合作医疗数据和肿瘤登记数据捕获肿瘤病例的差异,发现新型农村合作医疗数据捕获率高达95.60%,证实了基于医疗保险数据捕获肿瘤病例这一新模式在我国的可行性。本研究在西部地区的2座城市开展进一步探索,发现医疗保险数据对城市地区肿瘤病例的捕获率较高,但与已有研究相比捕获率仍有进步空间。本研究队列人群未纳入有严重内外科疾病正在治疗的患者,这些疾病与肿瘤的发生有较多的共同危险因素^[35-36],因此该人群可能是肿瘤高发人群,并且频繁就医和进行各项检查会增加肿瘤早期患者的诊出可

能性^[37-38],将其剔除也使得医疗保险数据的捕获率降低。不过基于医疗保险数据捕获肿瘤病例也存在不足,除了极少数未参保人群医疗保险数据无法覆盖之外,由于疾病发现较晚未住院治疗即死亡,被误诊为良性肿瘤或其他疾病的患者^[15]也有可能被医疗保险漏报。因此,在继续推进全民医保的同时,也需提高医疗保险记录的原始数据的内容详尽度和质量,如美国从20世纪90年代就开始倡导准确记录疾病诊断和治疗操作的各项编码,并使用国际统一编码取缔地方编码等^[39-40],以促进医疗保险数据的广泛应用。

基于人群的肿瘤登记在肿瘤监测工作中的重要性毋庸置疑,因此,在全国范围内增加肿瘤登记点的数量、提高登记数据的质量一直都是我国癌症防控计划的重要目标^[41]。目前,我国肿瘤监测及随访网络已基本建成,覆盖全国人口的40%以上,但与韩国、澳大利亚等肿瘤登记全民覆盖的发达国家还有很大差距^[42]。此外,我国肿瘤登记数据的质量也亟待提高^[25]。国际癌症研究署/国际癌症登记协会(IARC/IACR)对纳入《五大洲癌症发病率》的登记数据的质量提出要求,需要审核多项指标^[43-44]。目前我国多个登记点的数据质量仍难以满足评估中的完整性和可靠性要求^[44],在2017年发布的《五大洲癌症发病率》第11卷中,全国仅有36个登记点的发病数据质量达标被收录^[45]。要减少数据质量问题造成的漏报,改善全国尤其西部资源匮乏地区的报告资料质量,还需要国家大量的资金投入和长时间政策支持。《“健康中国2030”规划纲要》《“十三五”卫生与健康规划》等重要政策文件都将癌症防控作为重要内容,在国家层面进行健康战略部署^[46-47],相关工作主要考核指标的完成和落实,以及癌症防控效果的评价都依赖于及时、准确的肿瘤相关的人群数据。医疗保险数据具有覆盖人群广、资料收集所需时间短、成本低的显著优势,本研究结果提示,在一些肿瘤登记工作还较为落后的城市地区,通过医疗保险捕获肿瘤病例,并与其他数据库关联进行信息补充,是极具潜力的人群肿瘤信息收集模式。但更进一步的研究还有待开展,如2种模式在儿童、青壮年等肿瘤低危人群中的捕获效果和成本效益的相关资料仍待补充。

本研究也存在以下不足:(1)不同肿瘤登记点随访工作完善程度存在差异,本研究仅关注内蒙古2

个市的肿瘤登记数据,不能代表全国各登记点的水平,但本研究目的在于为我国肿瘤登记系统的完善提供思路,提示在肿瘤登记发展较落后、登记数据质量仍有待提高的地区使用医疗保险数据识别肿瘤病例是较为高效的人群肿瘤信息收集方式。(2)不同级医疗机构对医疗保险数据的诊断要求不一,因此存在出院诊断填写“恶性肿瘤”而未写明部位的情况,可能低估了医疗保险数据识别特定类别肿瘤病例的数量,提示医疗保险数据的质量仍需提升。(3)本研究仅关注恶性肿瘤及中枢神经系统良性肿瘤,但对于真性红细胞增多症等曾经被认为是良性的恶性肿瘤未予以纳入,后续研究可纳入这些特殊病种以保证研究结果的代表性。(4)本研究选择现成的筛查数据,探索性观察医疗保险数据对于40~74岁这一肿瘤高发群体的捕获效果,目标人群未涵盖75岁及以上老年人群与儿童及青壮年群体,研究结果对这些特殊群体的代表性有所欠缺。但儿童和青少年肿瘤发病率相对低,通过队列随访的形式捕获患者成本高、获益小,因此相比于本研究人群,在儿童及青壮年群体中通过医疗保险捕获肿瘤病例具有更重要的现实意义。

综上,本研究表明,在我国城市地区通过医疗保险数据捕获肿瘤病例在减少漏报方面极具优势,相关部门应采取措施进一步提高医疗保险资料中诊断相关的原始数据录入准确性,以探索更高效的全人群肿瘤登记监测的发病数据收集模式。

志谢:感谢国家癌症中心/国家肿瘤临床研究中心/中国医学科学院北京协和医学院肿瘤医院魏文强教授为本文提供了宝贵建议!

参考文献:

- [1] 健康中国行动——癌症防治实施方案(2019—2022年)[J]. 中国肿瘤,2019,28(11):803-806.
Cancer prevention and treatment implementation plan of Healthy China Action(2019—2022)[J]. China Cancer,2019,28(11):803-806.
- [2] 魏文强,张思维,李敏娟. 中国肿瘤登记发展历程[J]. 中国肿瘤,2021,30(9):641-647.
Wei WQ,Zhang SW,Li MJ. The history,present and prospect of cancer registration in China[J]. China Cancer 2021,30(9):641-647.
- [3] Rimland JM,Abraha I,Luchetta ML,et al. Validation of chronic obstructive pulmonary disease(COPD) diagnoses in

- healthcare databases: a systematic review protocol[J]. *BMJ Open*, 2016, 6(6):e011777.
- [4] Benchimol EI, Manuel DG, To T, et al. Development and use of reporting guidelines for assessing the quality of validation studies of health administrative data[J]. *J Clin Epidemiol*, 2011, 64(8): 821–829.
- [5] Kim M, Chae KH, Chung YJ, et al. The effect of the look-back period for estimating incidence using administrative data[J]. *BMC Health Serv Res*, 2020, 20(1): 166.
- [6] 国家医疗保障局. 2019年全国医疗保障事业发展统计公报 [EB/OL]. (2020-06-24). http://www.nhsa.gov.cn/art/2020/6/24/art_7_3268.html, National Healthcare Security Administration. Communique of development statistics of national healthcare security [EB/OL]. (2020-06-24). http://www.nhsa.gov.cn/art/2020/6/24/art_7_3268.html.
- [7] Tian H, Xu R, Li F, et al. Identification of cancer patients using claims data from health insurance systems: a real-world comparative study [J]. *Chin J Cancer Res*, 2019, 31(4): 699–706.
- [8] Yuk JS. The incidence rates of endometrial hyperplasia and endometrial cancer: a four-year population-based study[J]. *Peer J*, 2016, 4: e2374.
- [9] Lin YH, Chen KK, Chiu JH. Prevalence, patterns, and costs of Chinese medicine use among prostate cancer patients: a population-based study in Taiwan[J]. *Integr Cancer Ther*, 2010, 9(1): 16–23.
- [10] Sathiakumar N, Delzell E, Yun H, et al. Accuracy of medicare claim-based algorithm to detect breast, prostate, or lung cancer bone metastases [J]. *Med Care*, 2017, 55(12): e144–e149.
- [11] Krensell M, Petersen J, Mohr P, et al. Estimating prevalence and incidence of skin cancer in Germany[J]. *J Dtsch Dermatol Ges*, 2019, 17(12): 1239–1249.
- [12] Shi C, Liu M, Liu Z, et al. Using health insurance reimbursement data to identify incident cancer cases[J]. *J Clin Epidemiol*, 2019, 114: 141–149.
- [13] 许璐, 刘扬, 赖雪峰, 等. 我国孤立性浆细胞瘤的患病率调查: 基于全国 2016 年城镇医疗保险数据的测算[J]. *中华血液学杂志*, 2020, 41(6): 451–455.
- Xu L, Liu Y, Lai XF, et al. Prevalence investigation of solitary plasmacytoma in China: a calculation based on national urban medical insurance in 2016 [J]. *Chinese Journal of Hematology*, 2020, 41(6): 451–455.
- [14] Tian HR, Yang W, Hu YJ, et al. Estimating cancer incidence based on claims data from medical insurance systems in two areas lacking cancer registries in China[J]. *E Clinical Medicine*, 2020, 20: 100312.
- [15] Tian H, Hu Y, Li Q, et al. Estimating cancer survival and prevalence with the Medical-Insurance-System-based Cancer Surveillance System(MIS-CASS): an empirical study in China[J]. *E Clinical Medicine*, 2021, 33: 100756.
- [16] 席云峰, 张星光, 乔丽颖, 等. 内蒙古城市社区居民 2017—2018 年肺癌风险评估及筛查结果分析 [J]. *中国公共卫生*, 2020, 36(1): 25–29.
- Xi YF, Zhang XG, Qiao LY, et al. Lung cancer risk and screening outcome among 40–74 years old urban community residents in Inner Mongolia Autonomous[J]. *Chinese Journal of Public Health*, 2020, 36(1): 25–29.
- [17] Hook EB, Regal RR. Capture-recapture methods in epidemiology: methods and limitations[J]. *Epidemiol Rev*, 1995, 17(2): 243–264.
- [18] Barbadoro P, Luciani A, Ciotti M, et al. Two-source capture-recapture method to estimate the incidence of acute flaccid paralysis in the marches region(Italy)[J]. *Int J Environ Res Public Health*, 2020, 17(24): 9400.
- [19] 高建琼. 内蒙古 2015 年恶性肿瘤流行现状与 2011–2015 年流行趋势分析[D]. 呼和浩特: 内蒙古医科大学, 2019.
- Gao JQ. Prevalence of malignant tumors in inner mongolia in 2015 and prevalence trend analysis from 2011 to 2015 [D]. Hohhot: Inner Mongolia Medical University, 2019.
- [20] 陈文婕, 董玮琪, 乔丽颖, 等. 2015 年内蒙古自治区女性乳腺癌发病与死亡情况及 2010—2015 年趋势分析[J]. *中国肿瘤*, 2020, (10): 736–743.
- Chen WJ, Dong WQ, Qiao LY, et al. Incidence and mortality of female breast cancer in Inner Mongolia Autonomous Region in 2015 and trend analysis in 2010 and 2015[J]. *China Cancer*, 2020, (10): 736–743.
- [21] 王保山, 贺淑梅, 张世来. 德州市德城区 2014 年肿瘤登记漏报调查报告[J]. *中国保健营养*, 2017, 27(28): 280.
- Wang BS, He SM, Zhang SL. Investigation report of tumor registration omission in Dezhou City in 2014[J]. *China Health Care & Nutrition*, 2017, 27(28): 280.
- [22] 孙婷婷, 信振江, 谢俊卿, 等. 2010 年北京市丰台区医疗机构住院肿瘤病例漏报调查分析 [J]. *中国肿瘤*, 2013, (4): 270–273.
- Sun TT, Xin ZJ, Xie JQ, et al. An analysis on missed report of inpatients with cancer in Fengtai District, Beijing, 2010 [J]. *China Cancer*, 2013, (4): 270–273.
- [23] 张木瑾, 李树业, 高海燕, 等. 宝鸡市肺癌监测点漏报情况调查分析[J]. *陕西肿瘤医学*, 2001, 9(1): 45–46.
- Zhang MJ, Li SY, Gao HY, et al. Investigation and analysis of omission of lung cancer monitoring sites in Baoji City[J]. *Shaanxi Oncology Medicine*, 2001, 9(1): 45–46.
- [24] 陈万青, 梁智恒, 岑惠珊, 等. 中国肿瘤登记现状及发展 [J]. *中国医学前沿杂志: 电子版*, 2016, 8(7): 1–5.
- Chen WQ, Liang ZH, Cen HS, et al. Current status and development of tumor registration in China[J]. *Chinese Journal of Medical Frontiers (Electronic Edition)*, 2016, 8(7): 1–5.
- [25] 张文彬, 刘潇霞, 乔良, 等. 探索大数据背景下肿瘤登记信息平台建设的新模式 [J]. *肿瘤预防与治疗*, 2019, 32(11): 951–954.

- Zhang WB, Liu XX, Qiao L, et al. Exploring a new model for the construction of cancer registries in the era of big data[J]. *Journal of Cancer Control and Treatment*, 2019, 32(11):951-954.
- [26] 魏文强, 赫捷. 大数据信息化背景下我国肿瘤登记工作的思考[J]. *中华肿瘤杂志*, 2019, (1):15-18.
Wei WQ, He J. Some thoughts on cancer registry in China: in the era of big data and informatization[J]. *Chinese Journal of Oncology*, 2019, (1):15-18.
- [27] Tangka FK, Subramanian S, Edwards P, et al. Resource requirements for cancer registration in areas with limited resources: analysis of cost data from four low- and middle-income countries [J]. *Cancer Epidemiol*, 2016, 45 (Suppl 1):S50-S58. 2016, 45(Suppl 1):S50-S58.
- [28] 孔德明, 曾丽芳, 易懿, 等. 公立医院医保管理智能审核系统应用[J]. *中华医院管理杂志*, 2018, 34(11):936-939.
Kong DM, Zeng LF, Yi Y, et al. Application of intelligent audit system for medical insurance management in public hospitals [J]. *Chinese Journal of Hospital Administration*, 2018, 34(11):936-939.
- [29] 薄志红. 城镇职工医疗保险和城镇居民医疗保险之四大区别[J]. *金融经济(市场版)*, 2014, (10):38.
Bo ZH. The four big differences between medical insurance of urban worker and urban residents[J]. *Finance & Economy*, 2014, (10):38.
- [30] 仇雨临. 中国医疗保障 70 年: 回顾与解析[J]. *社会保障评论*, 2019, 3(1):89-101.
Qiu YL. 70 Years of medical security in China: review and analysis[J]. *Social Security Review*, 2019, 3(1):89-101.
- [31] Melody J, Eide RK, Dayna J, et al. Identification of patients with nonmelanoma skin cancer using health maintenance organization claims data[J]. *Am J Epidemiol*, 2010, 171(1):123-128.
- [32] Ramsey SD, Scoggins JF, Blough DK, et al. Sensitivity of administrative claims to identify incident cases of lung cancer: a comparison of 3 health plans[J]. *J Manag Care Pharm*, 2009, 15(8):659-668.
- [33] McClish DK, Penberthy L, Whittemore M, et al. Ability of medicare claims data and cancer registries to identify cancer cases and treatment[J]. *Am J Epidemiol*, 1997, 145(3):227-233.
- [34] Gold HT, Do HT. Evaluation of three algorithms to identify incident breast cancer in Medicare claims data [J]. *Health Serv Res*, 2007, 42(5):2056-2069.
- [35] Keum N, Giovannucci E. Global burden of colorectal cancer: emerging trends, risk factors and prevention strategies [J]. *Nat Rev Gastroenterol Hepatol*, 2019, 16(12):713-732.
- [36] Tzoulaki I, Elliott P, Kontis V, et al. Worldwide exposures to cardiovascular risk factors and associated health effects: current knowledge and data gaps [J]. *Circulation*, 2016, 133(23):2314-2333.
- [37] 中华医学会消化病学分会. 中国大肠肿瘤筛查、早诊早治和综合预防共识意见 [J]. *胃肠病学和肝病学杂志*, 2011, 20(11):979-996.
Chinese Society of Gastroenterology. Consensus on colorectal cancer screening, early diagnosis and treatment and comprehensive prevention in China[J]. *Chinese Journal of Gastroenterology and Hepatology*, 2011, 20(11):979-996.
- [38] 蒋绚, 袁一平, 徐定婷, 等. 结直肠癌就诊模式对肿瘤早期诊断的影响[J]. *中华医学杂志*, 2011, 91(41):2886-2890.
Jiang X, Yuan YP, Xu DT, et al. Influences of diagnostic modes on an early diagnosis of colorectal cancer [J]. *Chinese Medical Journal*, 2011, 91(41):2886-2890.
- [39] Leonard CE, Brensinger CM, Nam YH, et al. The quality of Medicaid and Medicare data obtained from CMS and its contractors: implications for pharmacoepidemiology [J]. *BMC Health Serv Res*, 2017, 17(1):304.
- [40] Parente ST, Weiner JP, Garnick DW, et al. Developing a quality improvement database using health insurance data: a guided tour with application to Medicare's national claims history file[J]. *Am J Med Qual*, 1995, 10(4):162-176.
- [41] 魏文强. 中国肿瘤登记工作及其在落实全民健康战略中的作用[J]. *中国肿瘤*, 2020, 29(10):721-724.
Wei WQ. Cancer registration in China and its role in implementing the National Health Strategy[J]. *China Cancer* 2020, 29(10):721-724.
- [42] Wei WQ, Zeng HM, Zheng RS, et al. Cancer registration in China and its role in cancer prevention and control[J]. *Lancet Oncol*, 2020, 21(7):e342-e349.
- [43] Bray F, Parkin DM. Evaluation of data quality in the cancer registry: principles and methods. Part I: comparability, validity and timeliness [J]. *Eur J Cancer*, 2009, 45(5):747-755.
- [44] Parkin DM, Bray F. Evaluation of data quality in the cancer registry: principles and methods. Part II: completeness[J]. *Eur J Cancer*, 2009, 45(5):756-764.
- [45] Bray F, Colombet M, Mery L, et al. Cancer incidence in five continents, Vol. XI[M]. IARC:2017.
- [46] 中国共产党中央委员会, 中华人民共和国国务院. “健康中国 2030”规划纲要[R/OL]. (2016-10-25). http://www.gov.cn/xinwen/2016-10/25/content_5124174.htm.
Central Committee of Communist Party of China, State Council of the People's Republic of China. An outline for the Healthy China 2030 Initiative[R/OL]. (2016-10-25). http://www.gov.cn/xinwen/2016-10/25/content_5124174.htm.
- [47] 中华人民共和国国务院. “十三五”卫生与健康规划[R/OL]. (2016-12-27). http://www.gov.cn/zhengce/content/2017-01/10/content_5158488.htm.
The State Council of the People's Republic of China. The 13th 5-year plan for health and wellness [R/OL]. (2016-12-27). http://www.gov.cn/zhengce/content/2017-01/10/content_5158488.htm.