Original Article

Identification of cancer patients using claims data from health insurance systems: A real-world comparative study

Hongrui Tian[1], Ruiping Xu[2], Fenglei Li[3], Chuanhai Guo[1], Lixin Zhang[2], Zhen Liu[1], Mengfei Liu[1], Yaqi Pan[1], Zhonghu He[1], Yang Ke[1]

[1]Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education/Beijing), Laboratory of Genetics, Peking University Cancer Hospital & Institute, Beijing 100142, China; [2]Anyang Cancer Hospital, Anyang 455000, China; [3]Hua County People's Hospital, Anyang 456400, China

Correspondence to: Yang Ke. Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education/Beijing), Laboratory of Genetics, Peking University Cancer Hospital & Institute, No. 52 Fucheng Road, Beijing 100142, China. Email: keyang@bjmu.edu.cn; Zhonghu He. Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education/Beijing), Laboratory of Genetics, Peking University Cancer Hospital & Institute, No. 52 Fucheng Road, Beijing 100142, China. Email: zhonghuhe@foxmail.com.

Abstract

Objective: To evaluate the accuracy of identifying cancer patients by use of medical claims data in a health insurance system in China, and provide the basis for establishing the claims-based cancer surveillance system in China.

Methods: We chose Hua County, Henan Province as the study site, and randomly selected 300 and 1,200 qualified inpatient electronic medical records (EMRs) as well as the New Rural Cooperative Medical Scheme (NCMS) claims records for cancer patients in Hua County People's Hospital (HCPH) and Anyang Cancer Hospital (ACH) in 2017. Diagnostic information for NCMS claims was evaluated on an individual level, and sensitivity and positive predictive value (PPV) were calculated taking the EMRs as the gold standard.

Results: The sensitivity of NCMS was 95.2% (93.8%−96.3%) and 92.0% (88.3%−94.8%) in ACH and HCPH, respectively. The PPV of the NCMS was 97.8% (96.7%−98.5%) in ACH and 89.0% (84.9%−92.3%) in HCPH. Overall, the weighted and combined sensitivity and PPV of NCMS in Hua County was 93.1% and 92.1%, respectively. Significantly higher sensitivity and PPV in identifying patients with common cancers than non-common cancers were detected in HCPH and ACH separately ($P<0.01$).

Conclusions: Identification of cancer patients by use of the NCMS is accurate on individual level, and it is therefore feasible to conduct claims-based cancer surveillance in areas not covered by cancer registries in China.

Keywords: NCMS; claims data; cancer surveillance; sensitivity; positive predictive value

Submitted Jun 23, 2019. Accepted for publication Aug 13, 2019.
doi: 10.21147/j.issn.1000-9604.2019.04.13
View this article at: https://doi.org/10.21147/j.issn.1000-9604.2019.04.13

Introduction

Cancer is the second most frequent cause of death in the world (1). According to the estimate by the International Agency for Research on Cancer, in 2018, 4.29 million new cancer cases were diagnosed in China, accounting for 23.7% of the total in the world (2). Accurate and real-time surveillance of population-level cancer incidence is requisite for effective cancer control and provision of cancer services (3).

Medical claims data from health insurance systems constitute a potential means for identification of cancer cases, and the performance of this method has been evaluated in several previous studies (4-13). For example, Medicare claims were reported to be accurate in capturing cases of breast, colorectal and endometrial cancer (6,8).

The HealthCore Integrated Research Database was found to be sensitive in identifying patients with non-small cell lung cancer (13). For these studies, effective identification of cancer cases can be guaranteed by qualified and accurate diagnostic information in medical claims.

In China, the New Rural Cooperative Medical Scheme (NCMS) is one of three major nationwide medical insurance systems. In the late 1970s, fiscal decentralization led to the collapse of the rural health insurance system in China. As a result, 640 million rural residents accounting for more than half of the Chinese population became uninsured, and were thus exposed to household financial risk in the 1980s and 1990s (14). To deal with this crisis, the China central government launched the NCMS in 2003. This is a government-run, voluntary, county-based and cost-sharing medical insurance program, which aims to improve access to healthcare services for rural residents (15). By 2012, the NCMS had become available to almost all residents in rural China (16).

For patients insured by NCMS, diagnosis-related information for each inpatient visit is entered and stored in the medical insurance system once the reimbursement process is completed, which makes it theoretically feasible to identify hospitalized cancer patients. A recent study used NCMS claims data as one means of following up subjects in a screening trial cohort for newly diagnosed cancer cases, supporting the potential of NCMS-claims-based cancer incidence surveillance (17). However, before general application of this method, evaluating the accuracy of cancer diagnoses in NCMS claims at the individual level based on real-world cancer patients' data in a larger population is needed.

In the present study, we evaluated the accuracy of diagnostic information in NCMS claims by comparing the electronic medical records (EMRs) of cancer patients in two main hospitals in a high-risk area for esophageal cancer in northern China, in order to examine the potential for identifying cancer patients using claims data in the health insurance system in China.

Materials and methods

Study site

Hua County, Henan Province is an underdeveloped agricultural region in northern China (18), with a resident population of 1.1 million and a relatively low gross domestic product per capita ($3,411 in 2017) (19-21). As the sole medical insurance system for this rural population, the NCMS in Hua County has had coverage of over 99% since 2010 (22). In terms of real-time reimbursement, both expense details directly related to reimbursement policies and information regarding diagnoses are recorded in the NCMS system.

Hua County People's Hospital (HCPH) is a secondary general hospital in northwestern Hua County, and Anyang Cancer Hospital (ACH), located in Anyang City 40 miles from Hua County, is a tertiary hospital which specializes in cancer treatment. For cancer patients in rural Hua County, HCPH and ACH are the two major hospitals which contributed to 43% of the total cancer-related inpatient visits in 2017 (*Supplementary Figure S1*). On this basis, we chose these two hospitals to assess the accuracy of NCMS claims for cancer patients.

Data source and sampling

In this study, the inpatient EMR systems in HCPH and ACH were regarded as the reference for diagnosis. We used the time period from January 1st, 2017 to December 31st, 2017 based on patient admission date, and exported EMRs of cancer patients in this period. With the approval of the Hua County NCMS Management Office, we obtained all of the inpatient NCMS claims records for cancer patients admitted to HCPH and ACH in 2017. Cases were identified from the records as cancer if the International Classification of Diseases (10th revision, ICD-10) codes ranged from C00 to C97 (23), or if text-based diagnoses contained any of cancer-related keywords [After thoroughly reviewing the ICD-10 (Chinese version), we extracted all the keywords indicating cancer]. For the purpose of guaranteeing data security and privacy in both the EMR systems and the NCMS system, we abstracted only variables including name, gender, birth year, admission date, text-based diagnosis and ICD-10 code. Name, gender, birth year and admission date were jointly used to match the medical record and corresponding claims record for each visit.

We obtained the EMRs for a total of 416 and 1,475 cancer inpatients admitted in 2017 to the HCPH and ACH, respectively. Quality control was performed in the beginning, and those EMRs with incomplete or obscure diagnosis-related information were excluded. Of the remaining 392 EMRs from the HCPH and 1,432 EMRs from the ACH, we randomly selected 300 and 1,200 EMRs in the HCPH and ACH, respectively. We then matched these EMRs with corresponding NCMS claims records in preparation for verification of claims-based diagnoses. We also randomly selected 300 and 1,200 NCMS claims records for NCMS-identified cancer patients in Hua County admitted to the HCPH and the ACH, respectively in 2017. To assess the accuracy of cancer diagnoses from NCMS claims, we were permitted to review the corresponding EMRs in these two hospitals.

Statistical analysis

For all analysis, the diagnoses from the EMR systems served as the gold standard. Sensitivity and positive predictive value (PPV) were used to determine the ability for correctly identifying cancer cases from claims data. Sensitivity was defined as the percentage of EMR-identified cancer patients who were correctly identified using NCMS claims data. PPV was defined as the percentage of NCMS-identified cancer patients who were correctly identified according to the EMR system.

$$\text{Sensitivity} = \frac{\text{Cancer cases correctly identified using claims data}}{\text{Total cancer cases in the EMRs}} \times 100\% \qquad (1)$$

$$\text{PPV} = \frac{\text{Cancer cases correctly identified according to the EMR system}}{\text{Total cancer cases in claims data}} \times 100\% \qquad (2)$$

We calculated the sensitivity and PPV of claims-based diagnoses by hospital with a 95% exact binomial confidence interval (95% CI) (24). These parameters were also separately reported by gender, age, and cancer site. We classified patient cancer sites into two categories as “Common” and “Other”. “Common” sites included lung cancer, stomach cancer, esophageal cancer, liver cancer and breast cancer which were the five most common cancers in Henan Province according to the latest annual report (25). Comparisons between categorical variables were performed using a general Chi-square test and Fisher’s exact test as appropriate. We also enumerated the reasons for underreporting or misreporting cancer patients using NCMS claims data. For all tests, a two-sided $P<0.05$ was considered statistically significant. All the analysis was performed using Stata (Version 15; StataCorp LLC, Texas, USA).

Ethical statement

Research protocols were approved by the Institutional Review Board of Peking University School of Oncology, Beijing, China. Prior to data export, the data provider masked private information including identity card number, phone number, address, and so on.

Results

Of all the cancer patients in the ACH which were sampled, 62.8% were less than 65 (mean age 59) years old; 58.7% were female; and 75.4% were common cancers (lung cancer, stomach cancer, esophageal cancer, liver cancer or breast cancer). Of all the cancer patients sampled in the HCPH, 51.0% were less than 65 (mean age 62) years old; 58.0% were female; and 54.0% were common cancers (*Table 1*).

Calculated by weighting the proportion of medical claims for cancer patients in Hua County in these two hospitals (*Supplementary Figure S1*), the combined sensitivity and PPV of NCMS was 93.1% and 92.1%, respectively. Specifically, the sensitivity of the NCMS was 95.2% (93.8%−96.3%) for the ACH and 92.0% (88.3%−94.8%) for the HCPH, respectively. For the ACH, no significant difference was observed between common cancers and other cancers. In the HCPH, the NCMS had a significantly higher sensitivity for identification of common cancer patients than other cancer patients (P=0.003) (*Table 1*). In addition, the PPV of the NCMS was 97.8% (96.7%−98.5%) in the ACH and 89.0% (84.9%−92.3%) in the HCPH. For the ACH, the PPV of NCMS claims for common cancers was significantly higher than that for other cancers (P<0.001). For the HCPH, the PPV of NCMS claims tended to be higher for older male patients who suffered from common cancers (although the PPV was not significantly higher) (*Table 2*).

To determine the reasons for false negative and false positive results, the medical records of corresponding patients were further reviewed. There were 58 (4.8%) and 24 (8.0%) records of underreported cancer cases in the ACH and HCPH respectively. For the NCMS in the ACH, miscoding as unspecified diagnosis (such as abnormal findings on diagnostic imaging, or localized mass

Table 1 Sensitivity of NCMS claims in identifying cancer patients in 2017

Variables	Anyang Cancer Hospital				Hua County People's Hospital			
	n (%)	Correctly identified	Sensitivity [% (95% CI)]	P	n (%)	Correctly identified	Sensitivity [% (95% CI)]	P
Total	1,200 (100.0)	1,142	95.2 (93.8, 96.3)		300 (100.0)	276	92.0 (88.3, 94.8)	
Age (year)				0.345				0.454
<65	753 (62.8)	720	95.6 (93.9, 97.0)		153 (51.0)	139	90.8 (85.1, 94.9)	
≥65	447 (37.2)	422	94.4 (91.9, 96.3)		147 (49.0)	137	93.2 (87.8, 96.7)	
Gender				0.994				0.184
Male	496 (41.3)	472	95.2 (92.9, 96.9)		126 (42.0)	119	94.4 (88.9, 97.7)	
Female	704 (58.7)	670	95.2 (93.3, 96.6)		174 (58.0)	157	90.2 (84.8, 94.2)	
EMR cancer site				0.242				0.003
Common*	905 (75.4)	865	95.6 (94.0, 96.8)		162 (54.0)	156	96.3 (92.1, 98.6)	
Other	295 (24.6)	277	93.9 (90.5, 96.3)		138 (46.0)	120	87.0 (80.2, 92.1)	

EMR, electronic medical records; 95% CI, 95% confidence interval; *, Lung cancer, stomach cancer, esophageal cancer, liver cancer and breast cancer are classified as common cancer sites.

Table 2 PPV of NCMS claims in identifying cancer patients in 2017

Variables	Anyang Cancer Hospital				Hua County People's Hospital			
	n (%)	Correctly identified	PPV [% (95% CI)]	P	n (%)	Correctly identified	PPV [% (95% CI)]	P
Total	1,200 (100.0)	1,173	97.8 (96.7, 98.5)		300 (100.0)	267	89.0 (84.9, 92.3)	
Age (year)				0.975				0.379
<65	759 (63.3)	742	97.8 (96.4, 98.7)		142 (47.3)	124	87.3 (80.7, 92.3)	
≥65	441 (36.7)	431	97.7 (95.9, 98.9)		158 (52.7)	143	90.5 (84.8, 94.6)	
Gender				0.879				0.160
Male	506 (42.2)	495	97.8 (96.1, 98.9)		125 (41.7)	115	92.0 (85.8, 96.1)	
Female	694 (57.8)	678	97.7 (96.3, 98.7)		175 (58.3)	152	86.9 (80.9, 91.5)	
NCMS cancer site				<0.001				0.097
Common*	896 (74.7)	884	98.7 (97.7, 99.3)		202 (67.3)	184	91.1 (86.3, 94.6)	
Other	304 (25.3)	289	95.1 (92.0, 97.2)		98 (32.7)	83	84.7 (76.0, 91.2)	

PPV, positive predictive value; NCMS, New Rural Cooperative Medical Scheme; 95% CI, 95% confidence interval; *, Lung cancer, stomach cancer, esophageal cancer, liver cancer and breast cancer are classified as common cancer sites.

and lump), other malignant tumors or symptoms (such as hemorrhage and anemia) were common reasons for false negative results. For the NCMS in the HCPH, miscoding as symptoms ranked first among underreported cancer cases, followed by miscoding as unspecified diagnoses (*Table 3*). In terms of false positive results, 27 of 1,200 NCMS-identified cancer cases were misreported in the ACH, with miscoding as other types of malignant tumors as the predominant reason. In the HCPH, 33 of 300 NCMS-identified cancer cases were misreported, mainly

Table 3 Reasons for underreporting cancer patients using NCMS claims

Hospital	Reason	n (%)
Anyang Cancer Hospital		
	Miscoded as unspecified diagnosis*	19 (32.8)
	Miscoded as other malignant tumors	12 (20.7)
	Coded according to symptoms	11 (19.0)
	Miscoded as uncertain/unknown behavior	8 (13.8)
	Miscoded as other diseases (same site)	3 (5.2)
	Miscoded as benign tumors	2 (3.5)
	Miscoded as other diseases (other sites)	2 (3.5)
	Unreimbursed	1 (1.7)
	Total	58 (100.0)
Hua County People's Hospital		
	Coded according to symptoms	8 (33.3)
	Miscoded as unspecified diagnosis*	6 (25.0)
	Miscoded as other diseases (other sites)	3 (12.5)
	Miscoded as uncertain/unknown behavior	2 (8.3)
	Miscoded as other diseases (same site)	2 (8.3)
	Others	3 (12.5)
	Total	24 (100.0)

NCMS, New Rural Cooperative Medical Scheme; *, Miscoding as unspecified diagnosis refers to abnormal findings on diagnostic imaging, and a localized mass and lump.

due to miscoding as other types of malignant tumors, benign tumors, or other disease at the same site (*Table 4*).

Discussion

Traditionally, population-based cancer registries (PBCR) and active follow-up by face-to-face interview or via telephone are the two main methods for collecting cancer cases in prospective studies (26-28). PBCR serves as the gold standard for determination of incident cancer cases (29). In previous studies, the accuracy of identification of cancer cases using medical claims in the health insurance system has been evaluated in several other countries, such

Table 4 Reasons for misreporting cancer patients using NCMS claims

Hospital	Reason	n (%)
Anyang Cancer Hospital		
	Other malignant tumors	24 (88.9)
	Benign tumors	1 (3.7)
	Other diseases (same site)	1 (3.7)
	Others	1 (3.7)
	Total	27 (100.0)
Hua County People's Hospital		
	Other malignant tumors	13 (39.4)
	Benign tumors	10 (30.3)
	Other diseases (same site)	9 (27.3)
	Others	1 (3.0)
	Total	33 (100.0)

NCMS, New Rural Cooperative Medical Scheme.

as the USA (sensitivity 91%, PPV 82%, prostate cancer) (30), Canada (sensitivity 87%, PPV 79%, pancreatic cancer) (31), and Japan (sensitivity 90%, PPV 87%, breast cancer) (32). For areas not covered by PBCRs in China, claims data have the potential to provide a basis for cancer surveillance, but the accuracy of this method must first be evaluated. Based on the above, we conducted the present comparative study in a high-risk region for cancer. Through individual-level comparison of diagnoses from NCMS claims and EMRs (gold standard), we observed that NCMS claims data were accurate in identifying cancer patients.

Of all the EMRs sampled, only one had no corresponding claims record, with name, gender, birth year and admission date as matching variables. This indicates that the reimbursement rate for NCMS-insured cancer patients in Hua County is nearly 100%. In addition, as the two major hospitals for cancer patients in Hua County, the ACH (a tertiary specialized cancer hospital) and the HCPH (a secondary general hospital) respectively represent the upper and lower limits of professional level of clinical oncologists and NCMS staffs to a great extent, which largely determines the claims data quality. Thus, for NCMS claims in Hua County, the point estimate of sensitivity for identification of cancer patients is likely within the interval 92%−95%, with PPV ranging from 89% to 98%. Although the present study simply evaluated the accuracy of claims-based diagnoses for cancer patients (prevalent cases), which in terms of study design differed from the studies cited above verifying claims-based incident cases, our results are to some extent comparable with previous studies. Hence, the sensitivity and PPV for NCMS claims for cancer patients, whether from the ACH or in HCPH, are relatively high.

As mentioned, we observed variation in the accuracy of NCMS claims by hospital type. NCMS claims for the ACH were more accurate than those for the HCPH, with a 3% increase in sensitivity and a 9% increase in PPV. With further stratified analysis, we found that the difference in NCMS accuracy among cancer sites might for the most part account for this variation by hospital type. NCMS claims data for common cancer patients were more accurate, with similar sensitivity in these two hospitals. For less common cancers, NCMS accuracy in the HCPH was obviously lower than that in the ACH. This may be due to limited knowledge regarding cancer of the NCMS staff in lower grade hospitals, especially for less common cancers. After reviewing the EMRs of those incorrectly identified patients, we found that most of the errors could have been avoided. For example, during the reimbursement process, some cancer patients were recorded in the reimbursement system according to certain symptoms, such as hemorrhage and anemia, which might be related to the severity of cancer stage or the treatment cancer patients had received. This indicates that it is necessary to raise awareness of cancer diagnosis among the NCMS staff, enhance staff training and strengthen quality control procedures, especially for lower grade hospitals.

In practice, this claims-based method to identify cancer patients has obvious advantage due to its low cost as the diagnostic data are collected and restored in real-time once patients are reimbursed. As such, NCMS claims data perform well in monitoring cancer burden, and can also serve as a supplementary source of newly diagnosed cancer cases for cancer registry. For rural areas in China without cancer registries, the NCMS has the potential to play an increasingly important role in cancer incidence surveillance, where further claims-based studies are warranted.

There are several limitations to this study. Firstly, as a severe chronic disease, cancer is often fatal because it is late stage at the time of clinical presentation (33). Hospitalization is thus the predominant path chosen by cancer patients for treatment, and can theoretically reflect cancer burden in Hua County. Based on this, we evaluated the accuracy of claims-based diagnosis for cancer patients, using inpatient EMRs only as a gold standard. However, there may still be a small proportion of cancer patients who are never hospitalized due to the severity of disease, which indicates that further studies are needed to evaluate the impact of these missed cases on the accuracy of NCMS claims. Secondly, since abstracted EMR data were provided through a request for information of cancer patients only, those patients with non-neoplastic disease were unavailable, and there was thus inadequate information for calculation of specificity (5,34). However, from a practical point of view, it is difficult to determine and obtain non-cancer disease data in view of the massive amount of data and limited permission for access. In addition, sensitivity and PPV can comprehensively reflect the ability of NCMS claims data for identifying cancer cases. Thirdly, though HCPH and ACH could respectively represent the lower and upper limit of the accuracy of NCMS in Hua County, we still propose the necessity of sampling cancer patients from all hospitals in Hua County in the future, provided that all the EMRs are available.

Conclusions

This comparative study has shown that it is feasible and accurate to conduct cancer surveillance in China using government-run medical insurance systems such as the NCMS. Continuous effective measures should be taken to further enhance the quality of original claims data in terms of cancer diagnosis and data entry, especially in those non-specialized hospitals of lower grade. Medical claims databases in China, such as the NCMS system, could be broadly applied for accurate and timely reflection of local cancer burden.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (No. 30930102, 81473033), the National Key R&D Program of China (No. 2016YFC0901404), the Digestive Medical Coordinated Development Center of Beijing Hospitals Authority (No. XXZ0204), the Science Foundation of Peking University Cancer Hospital (No. 2017-4), and the Open Project funded by the Key Laboratory of Carcinogenesis and Translational Research, Ministry of Education/Beijing (No. 2017-10).

Footnote

Conflicts of Interest: The authors have no conflicts of interest to declare.

References

1. World Health Organization. World Health Statistics 2018: Monitoring health for the SDGs. Geneva: World Health Organization, 2018.
2. International Agency for Research on Cancer. Cancer Today. 2018. Available online: http://gco.iarc.fr/
3. Boyle P, Levin B. World Cancer Report 2008. Lyon: International Agency for Research on Cancer, 2018.
4. McClish D, Penberthy L, Pugh A. Using Medicare claims to identify second primary cancers and recurrences in order to supplement a cancer registry. J Clin Epidemiol 2003;56:760-7.
5. 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 Manag Care Pharm 2009;15:659-68.
6. Deshpande AD, Schootman M, Mayer A. Development of a claims-based algorithm to identify colorectal cancer recurrence. Ann Epidemiol 2015;25:297-300.
7. Funch D, Ross D, Gardstein BM, et al. Performance of claims-based algorithms for identifying incident thyroid cancer in commercial health plan enrollees receiving antidiabetic drug therapies. BMC Health Serv Res 2017;17:330.
8. Cooper GS, Yuan Z, Stange KC, et al. The sensitivity of Medicare claims data for case ascertainment of six common cancers. Med Care 1999;37:436-44.
9. Nattinger AB, Laud PW, Bajorunaite R, et al. An algorithm for the use of Medicare claims data to identify women with incident breast cancer. Health Serv Res 2004;39:1733-49.
10. Fenton JJ, Onega T, Zhu W, et al. Validation of a Medicare claims-based algorithm for identifying breast cancers detected at screening mammography. Med Care 2016;54:e15-22.
11. Mahnken JD, Keighley JD, Girod DA, et al. Identifying incident oral and pharyngeal cancer cases using Medicare claims. BMC Oral Health 2013;13:1.
12. McClish DK, Penberthy L, Whittemore M, et al. Ability of Medicare claims data and cancer registries to identify cancer cases and treatment. Am J Epidemiol 1997;145:227-33.
13. Turner RM, Chen YW, Fernandes AW. Validation of a case-finding algorithm for identifying patients with non-small cell lung cancer (NSCLC) in administrative claims databases. Front Pharmacol 2017;8:883.
14. Babiarz KS, Miller G, Yi H, et al. New evidence on the impact of China's New Rural Cooperative Medical Scheme and its implications for rural primary healthcare: multivariate difference-in-difference analysis. BMJ 2010;341:c5617.
15. Wang J, Zhou HW, Lei YX, et al. Financial protection under the new rural cooperative medical schemes in China. Med Care 2012;50:700-4.
16. Zhou M, Liu S, Kate Bundorf M, et al. Mortality in rural China declined as health insurance coverage increased, but no evidence the two are linked. Health Aff (Millwood) 2017;36:1672-8.
17. Shi C, Liu M, Liu Z, et al. Using health insurance reimbursement data to identified incident cancer cases. J Clin Epidemiol 2019;114:141-9.

18. Li F, Li X, Guo C, et al. Estimation of cost for endoscopic screening for esophageal cancer in a high-risk population in rural China: Results from a population-level randomized controlled trial. Pharmacoeconomics 2019;37:819-27.
19. Wang S. Henan Statistical Yearbook, 2017 (in Chinese). Beijing: China Statistics Press, 2017.
20. Hua County Bureau of Statistics. Statistical bulletin of national economy and social development of Hua County in 2017 (in Chinese). 2018. Available online: http://www.ha.stats.gov.cn/sitesources/hntj/page_pc/tjfw/tjgb/szgxgb/article677fc6a201e54965b8ece848a48f125c.html
21. National Bureau of Statistics of China. Statistical Communiqué of the People's Republic of China on the 2017 National Economic and Social Development. 2018. Available online: http://www.stats.gov.cn/english/PressRelease/201802/t20180228_1585666.html
22. Li X, Cai H, Wang C, et al. Economic burden of gastrointestinal cancer under the protection of the New Rural Cooperative Medical Scheme in a region of rural China with high incidence of oesophageal cancer: cross-sectional survey. Trop Med Int Health 2016;21:907-16.
23. World Health Organization. International Statistical Classification of Diseases and Related Health Problems 10th Revision. 2016. Available online: https://icd.who.int/browse10/2016/en
24. Brown LD, Cai TT, DasGupta A. Interval estimation for a binomial proportion. Statistical Science 2001;16:101-33.
25. Henan Cancer Center. Henan Cancer Registry Annual Report (in Chinese). Zhengzhou: Henan Science and Technology Press, 2018.
26. Pinsky PF, Yu K, Black A, et al. Active follow-up versus passive linkage with cancer registries for case ascertainment in a cohort. Cancer Epidemiol 2016;45:26-31.
27. Chen W, Zheng R, Zhang S, et al. Cancer incidence and mortality in China in 2013: an analysis based on urbanization level. Chin J Cancer Res 2017;29:1-10.
28. Chen W, Sun K, Zheng R, et al. Cancer incidence and mortality in China, 2014. Chin J Cancer Res 2018;30:1-12.
29. Bray F, Znaor A, Cueva P, et al. Planning and developing population-based cancer registration in low-and middle-income settings. Lyon: International Agency for Research on Cancer, 2014.
30. Parlett LE, Beachler DC, Lanes S, et al. Validation of an Algorithm for Claims-based Incidence of Prostate Cancer. Epidemiology 2019;30:466-71.
31. Wu JW, Azoulay L, Huang A, et al. Identification of incident pancreatic cancer in Ontario administrative health data: A validation study. Pharmacoepidemiol Drug Saf 2018.
32. Sato I, Yagata H, Ohashi Y. The accuracy of Japanese claims data in identifying breast cancer cases. Biol Pharm Bull 2015;38:53-7.
33. Sylla BS, Wild CP. A million africans a year dying from cancer by 2030: what can cancer research and control offer to the continent? Int J Cancer 2012;130:245-50.
34. Cooper GS, Yuan Z, Stange KC, et al. The utility of Medicare claims data for measuring cancer stage. Med Care 1999;37:706-11.

Cite this article as: Tian H, Xu R, Li F, Guo C, Zhang L, Liu Z, Liu M, Pan Y, He Z, Ke Y. Identification of cancer patients using claims data from health insurance systems: A real-world comparative study. Chin J Cancer Res 2019;31(4):699-706. doi: 10.21147/j.issn.1000-9604.2019.04.13

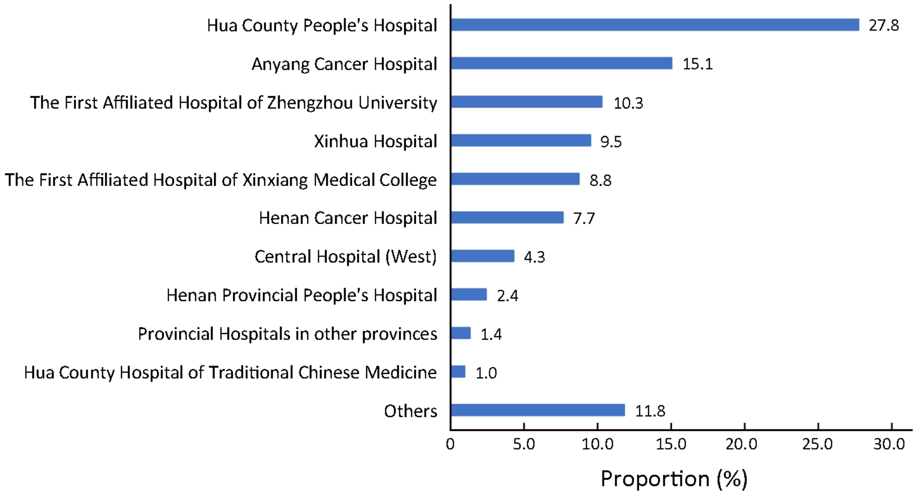

Figure S1 Proportion (%) of New Rural Cooperative Medical Scheme (NCMS) claims in 10 hospitals where cancer patients in Hua County in 2017 were mostly diagnosed or treated.