Methods and Applications

The Incidence of Lymphoma in Beijing: Comparing Results from MIS-CASS (2019) and Beijing Cancer Registry (2017)

Hongrui Tian[1,&]; Weiping Liu[2,&]; Yuqin Song[2]; Lan Mi[2]; Zhen Liu[1]; Mengfei Liu[1]; Chuanhai Guo[1]; Fangfang Liu[1]; Ying Liu[1]; Yaqi Pan[1]; Jun Zhu[2,#]; Zhonghu He[1,#]; Yang Ke[1]

ABSTRACT

Introduction: This study aimed to evaluate the performance of the Medical-Insurance-System-based Cancer Surveillance System (MIS-CASS) in estimating cancer incidence by comparing the results with the Beijing Cancer Registry (BCR), which is one of the highest-quality population-based cancer registries in China.

Methods: Using lymphoma as an example, we extracted relevant claims data from the administrative systems of medical insurance in Beijing (2012–2020) and estimated the most current lymphoma incidence in Beijing (2019) using a standard data processing procedure. The absolute number of new cases, crude incidence rate, and age-standardized incidence rate of lymphoma were compared with the latest data reported by the BCR (2017).

Results: Both lymphoma incidence rates and age distribution of new cases estimated based on MIS-CASS were similar to the BCR data (crude incidence rate: 9.8/100,000 *vs.* 10.6/100,000). However, because MIS-CASS included more designated hospitals and covered a larger local stationary population irrespective of household registration (*hukou*), the absolute number of incident lymphoma cases identified by MIS-CASS was 39.1% higher than that reported by the BCR (2,002 *vs.* 1,439).

Conclusions: The MIS-CASS approach reflected the actual cancer burden in a more complete and timely manner as compared with the current BCR, providing new insights for improving cancer surveillance strategies in China.

INTRODUCTION

Continuous cancer surveillance is fundamental for cancer prevention and control as well as for allocation of medical resources. Population-based cancer registration is recognized as the gold standard for reporting cancer incidence in a defined population (*1*). In China, the establishment of cancer registration dates back to the late 1950s, and currently, cancer registries cover approximately 43% of the population in the country (*2*). In the past 60 years, cancer registration in China has played a crucial role in the measurement of the cancer burden, allocation of medical resources, and promotion of cancer research. However, due to the vast territory, huge population, and limited resources of China, it is still difficult to establish a real nationwide population-based cancer registry (PBCR), and the quality of available cancer data showed significant variation across geographic areas (*2*). In addition, since the collection and ascertainment of incident cancer cases is labour intensive and time consuming under the current surveillance approach, the delay in reporting annual national cancer incidence statistics is often as long as 3 to 5 years (*3*).

The basic medical insurance system, including the Urban Employee Basic Medical Insurance (UEBMI), the New Rural Co-operative Medical Care Scheme (NCMS), and the Urban Residents Basic Medical Insurance (URBMI), is administrated by local governments, and in recent years has covered 95% of the population in China (*4*). For each discharged patient, diagnosis-related information is automatically submitted to the medical insurance system in real time once reimbursement is complete. We have previously conducted 2 studies to evaluate the accuracy of NCMS claims data in capturing cancer cases in a rural area of China, observing high sensitivity and specificity of the approach (*5–6*). Based on these results, we established the Medical-Insurance-System-based Cancer Surveillance System (MIS-CASS) and reported the latest cancer incidence in 2 regions (over 6 million residents altogether) not covered by PBCRs (*7*). The study demonstrated that the MIS-CASS efficiently and accurately reflected local cancer burden with a lag time of only 6 months. But the real-world performance of MIS-CASS in cancer incidence surveillance still needs to be further explored and validated based on evidence

from direct comparison with high-quality PBCR data.

The Beijing Cancer Registry (BCR) was established in 1976 and has covered all the 6 urban and 10 rural districts in Beijing since 1998 (8). This registry is one of the highest-quality PBCRs in China (9). Generally, registrars of all the BCR-designated hospitals (168 in the year of 2017) collect and submit data of new cancer cases through a special health information system (10). The entire household registered population in Beijing (13 million in 2017) are used for calculating cancer incidence rates (8).

Using lymphoma as an example neoplasm, this study aimed to estimate the latest incidence of lymphoma in Beijing based on medical insurance claims data and to evaluate the performance of MIS-CASS by comparing with the BCR.

METHODS

Data Collection and Data Processing

The medical insurance system of Beijing covers individuals who continuously worked or resided in the administrative area of Beijing, irrespective of place of household registration. In this study, we obtained data of the insured population in 2019 (covering about 20 million permanent residents and local employees) from the medical insurance administrative systems and extracted all the lymphoma-related medical claims records from January 1, 2012 to December 31, 2020, using both the International Classification of Diseases (tenth revision) (ICD-10) code (C81–85, C88, C90, C96) and text-based diagnosis (using a dictionary of lymphoma diagnosis constructed by reviewing the Chinese version of the ICD-10 and extracting all keywords indicating lymphoma cases) (7). The period between 2012 and 2018 was set as the 7-year time window for identifying prevalent cases in 2019, and the year 2020 was also included to avoid missing 2019 incident cases due to reimbursement delay (7). The standard procedure for identifying incident cancer cases (lymphoma) in 2019 has been described previously (Figure 1) (7). In brief, among the extracted claims records, those from primary or unrated healthcare facilities were excluded due to insufficient diagnostic qualifications (consistent with the criteria adopted in BCR). Lymphoma diagnoses were then determined using both ICD-10 codes and text-based diagnoses. When either was missing or inconsistency occurred, manual evaluation was performed through group discussion among 3 trained investigators

specialized in lymphoma and/or epidemiology. For patients with multiple records, only the first one with definite lymphoma diagnosis was retained for each patient. Finally, the new records from January 1, 2019 to December 31, 2019 were kept, and the corresponding patients were defined as incident lymphoma cases in this study. The corresponding admission date of each lymphoma case was defined as its incidence date, which was in accordance with the BCR (10).

Out of regard for confidentiality, individual-level identifiable personal information (name, identity number, telephone number, home address, and so on) were masked, and a unique encrypted identification code was generated for each case in order to eliminate duplicate records.

Statistical Analysis

In this study, incidence rate generated from MIS-CASS referred to the number of new lymphoma cases occurring in a given year (2019) in the insured population in Beijing [see the following formula (11)].

$$\text{Incidence rate} = \frac{\text{Number of new lymphoma cases}}{\text{Insured population at risk}} \text{ in 2019}$$

Crude incidence rate, age-specific incidence rate, and age-standardized incidence rate (ASR) were calculated by gender and subtype. Lymphoma was generally categorized into Hodgkin lymphoma (C81), non-Hodgkin lymphoma (C82–85, C96), and multiple myeloma (C88, C90) (consistent with the method used for the BCR and other PBCRs in China), and was further subdivided when describing morphologic distribution. The China standard population in 2000 (ASR China) and the Segi standard population (ASR World) were applied for calculation of ASRs. We then compared MIS-CASS estimates of 2019 with the latest BCR-reported lymphoma incidence data of 2017 which was published in April, 2021 (12). This study was approved by the Institutional Review Board of the Peking University School of Oncology, China (No. 2019KT44).

RESULTS

The lymphoma incidence in Beijing estimated by MIS-CASS was close to that reported by BCR (crude incidence rate: 9.8/100,000 *vs.* 10.6/100,000) (Table 1), and no significant difference in morphologic distribution was detected between MIS-CASS and BCR (Hodgkin lymphoma: 5.1% *vs.* 4.6%, non-
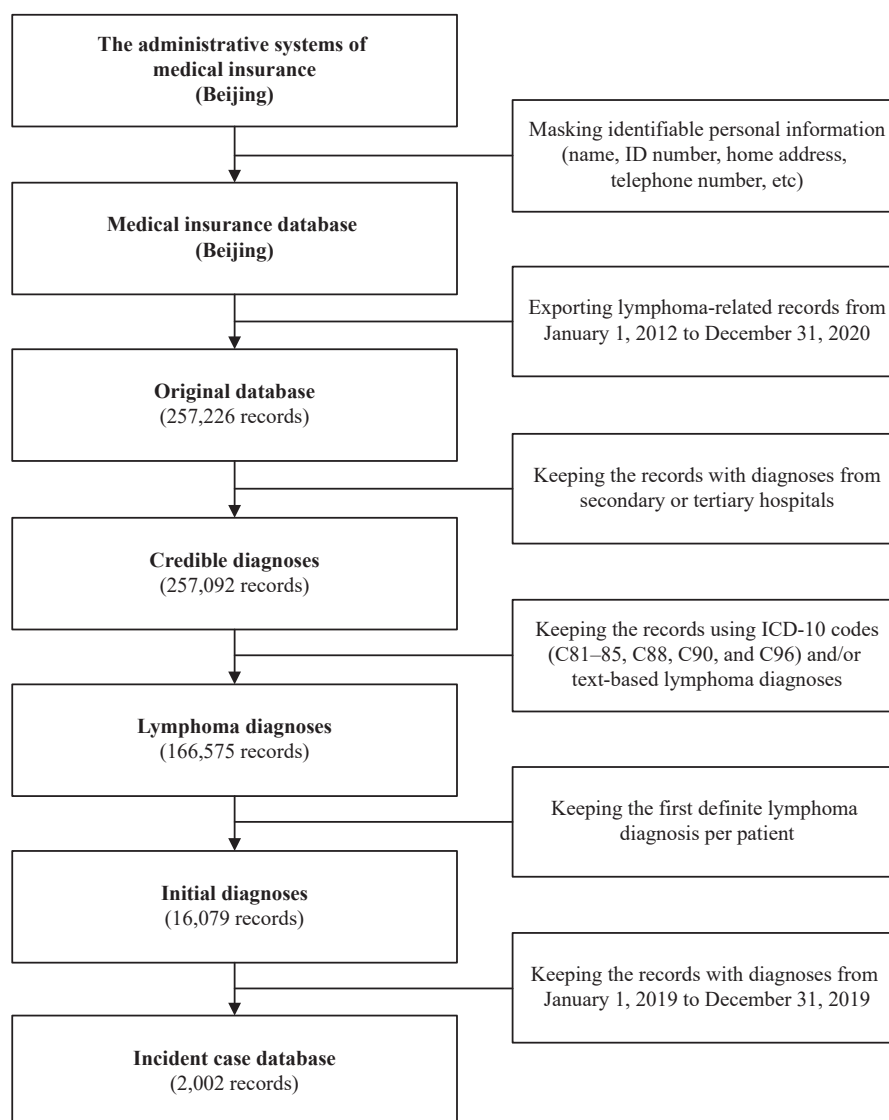
FIGURE 1. The procedure for capturing incident lymphoma cases in 2019 from the administrative systems of medical insurance in Beijing (MIS-CASS).
Abbreviations: ICD-10=International Classification of Diseases (tenth revision); MIS-CASS=Medical-Insurance-System-based Cancer Surveillance System.

Hodgkin lymphoma: 70.6% *vs.* 69.8%, multiple myeloma: 24.3% *vs.* 25.6%, *P*=0.559) (Figure 2) (*12*).

However, the absolute number of new lymphoma cases identified by MIS-CASS was 39.1% higher than that reported by the BCR (2,002 *vs.* 1,439) (still 34.1% higher when only claims data from BCR-designated hospitals were included for analysis) (Table 1). Further analysis revealed that, among the 110 hospitals where the new lymphoma cases were diagnosed in 2019 according to MIS-CASS, only 90 were BCR-designated hospitals (*12*), and the remaining 20 hospitals contributed 22.2% of the excess patients (125/563). In addition, due to differences in target population between MIS-CASS

(insured population) and BCR (household registered population), the population covered by MIS-CASS was 6.5 million larger than that for BCR in the age group of 20–49 years (Supplementary Figure S1, available in http://weekly.chinacdc.cn/).

Regarding age distribution, the incidence rate of lymphoma was relatively low in patients aged below 50 years and peaked at 75–84 years (Figure 3). The MIS-CASS-estimated number of new lymphoma cases peaked at age 60–69 years for both males and females, which was similar with the age distribution reported by the BCR, but was notably higher for the majority of the age groups, especially for men of 30–69 years and for women of 20–69 years (Figure 4).

TABLE 1. Comparison of the latest lymphoma incidence estimates in Beijing between MIS-CASS (2019) and BCR (2017).

Surveillance approach	Subtype	ICD-10 of lymphoma	Gender	No. of cases	Crude incidence rate (1/10^5)	ASR China (1/10^5)	ASR World (1/10^5)
MIS-CASS	All	C81–85, C88, C90, C96	Both	2,002	9.8	6.8	6.6
			Male	1,090	10.3	7.3	7.3
			Female	912	9.3	6.3	5.9
	Hodgkin lymphoma	C81	Both	103	0.5	0.4	0.4
			Male	48	0.5	0.4	0.3
			Female	55	0.6	0.5	0.4
	Non-Hodgkin lymphoma	C82–85, C96	Both	1,413	6.9	4.8	4.7
			Male	754	7.1	5.1	5.1
			Female	659	6.7	4.5	4.3
	Multiple myeloma	C88, C90	Both	486	2.4	1.5	1.5
			Male	288	2.7	1.8	1.8
			Female	198	2.0	1.3	1.2
BCR*	All	C81–85, C88, C90, C96	Both	1,439	10.6	5.9	5.6
			Male	814	12.0	6.8	6.6
			Female	625	9.2	4.9	4.7

Abbreviations: MIS-CASS=Medical-Insurance-System-based Cancer Surveillance System; BCR=Beijing Cancer Registry; ICD-10=International Classification of Diseases (tenth revision); ASR China=age-standardized incidence rate by China standard population in 2000; ASR World=age-standardized incidence rate by world standard population (Segi population).
* The latest lymphoma incidence data of BCR (year of 2017) were derived from *Beijing Cancer Registry Annual Report 2020* (*12*).

DISCUSSION

We have previously demonstrated that MIS-CASS could efficiently reflect cancer burden with a shorter reporting delay than the current cancer registries in China (6 months *vs*. 3 years) (*5–7*). In addition to timeliness, accuracy and completeness were also crucial for cancer surveillance (*13–15*). In this head-to-head comparative study conducted in an area covered by one of the highest-quality PBCRs (BCR) in China, we evaluated the performance of the MIS-CASS approach in estimating the most current cancer incidence (2019), using lymphoma as an example neoplasm.

As shown in this study, the incidence rate of lymphoma estimated by MIS-CASS was close to that reported by the BCR, and these 2 approaches were also consistent in terms of morphologic distribution as well as in the pattern of lymphoma burden across age groups. However, the absolute number of new lymphoma cases estimated by MIS-CASS was notably higher than that of the BCR. A more in-depth analysis of the numerator and denominator for calculation of

cancer incidence reveals the following 2 crucial findings:

First, MIS-CASS covers more designated hospitals than the BCR. For BCR, the information on cancer cases were merely collected from BCR-designated hospitals (e.g., 168 hospitals in 2017). Basic medical insurance covers a broader range of hospitals in Beijing specialized in cancer diagnosis and treatment, including a series of top-level military hospitals in China which are not covered by the BCR. Of the newly occurred lymphoma cases missed by the BCR, as many as 72% were diagnosed in these military hospitals. Absence of cancer diagnoses in these health facilities leads to underestimation of the absolute number of incident lymphoma cases in Beijing.

Secondly, all cancer cases reported by the BCR were restricted to the registered population, i.e., those with a local household registration (*hukou*), which is a fundamental social management system in China initiated in the 1950s. Hence, new cancer cases that occur in residents in Beijing but registered in other
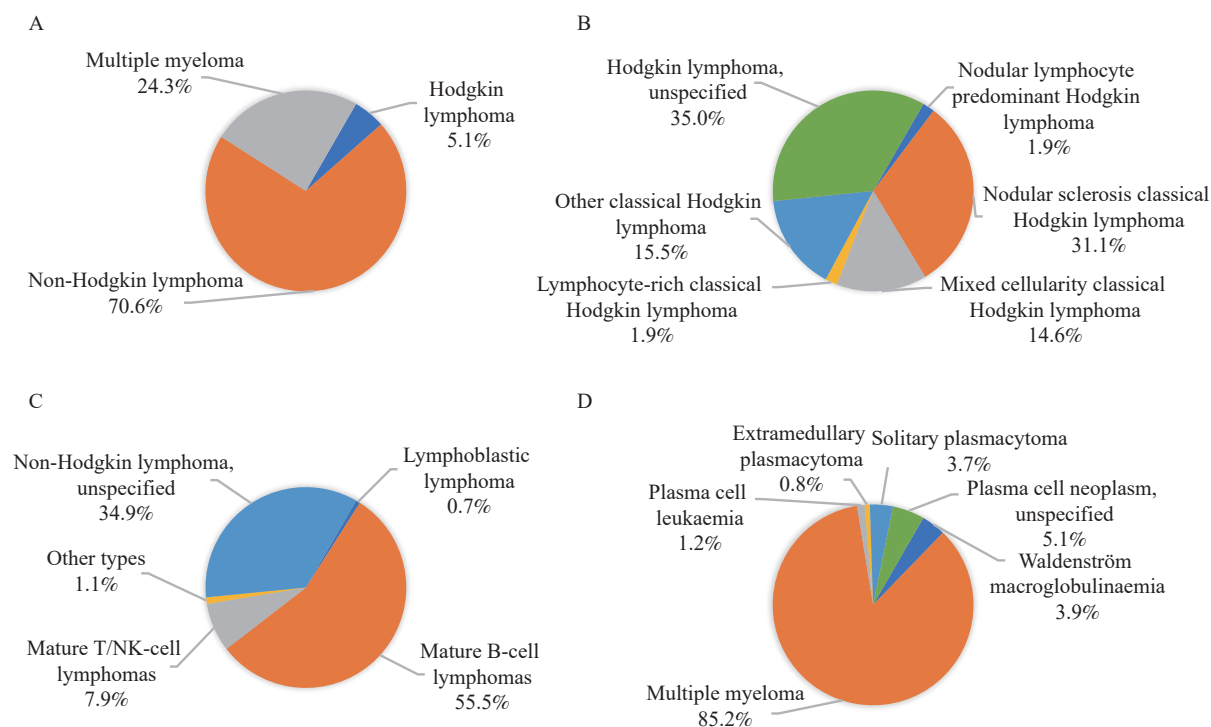
FIGURE 2. Morphologic distribution of incident lymphoma cases in Beijing, 2019 (MIS-CASS). (A) Lymphoma (C81–85, C88, C90, C96); (B) Hodgkin lymphoma (C81); (C) Non-Hodgkin lymphoma (C82–85, C96); (D) Multiple myeloma (C88, C90).
Abbreviation: MIS-CASS=Medical-Insurance-System-based Cancer Surveillance System.

areas would not be counted by BCR (8). However, Beijing is characterized by high population mobility due to its central role in economic and cultural development, and many individuals from other provinces work and reside in Beijing for long periods of time. They are not registered in the Beijing household registration system, but are insured by the local medical insurance system. As a result, the population covered by MIS-CASS represents the permanent population in Beijing to a greater extent than that covered in BCR. This study revealed that the current BCR failed to capture millions of these individuals, especially for young and middle-aged groups (20–49 years) in local area (Supplementary Figure S1). More importantly, as the cancer patients in this group are generally diagnosed and treated in Beijing, they would also be missed by any cancer registry in places where they registered their household according to the current regulation of PBCRs in China. The fact that PBCRs in China are based on the household registration system would further lead to underestimation of both the number of newly detected cancer cases and the size of the target population simultaneously, that is, the numerator and

denominator of the cancer incidence rate. It is worth noting that the degree of underestimation may be further magnified for the high-incidence cancer types in the middle-age population such as breast cancer, cervical cancer, and thyroid cancer (7).

Therefore, the MIS-CASS approach could provide a more complete picture of cancer incidence and actual consumption of corresponding medical resources. The unique advantage of MIS-CASS for cancer incidence surveillance is due to the fact that basic medical insurance systems in China are administrated and operated by central and local governments and that the submitted claims data are highly standardized with the establishment of the National Healthcare Security Administration and integration of major medical insurance systems (covering 95% of national population). Despite that, efforts should also be continuously made to further improve the quality of the original claims data through training and supervising doctors entering the diagnosis in real-world clinical practice and the staff responsible for submission of claims records, especially those in lower grade hospitals. The approach to identify cancer cases using medical insurance claims data has been explored in several countries in the past decades; however, few
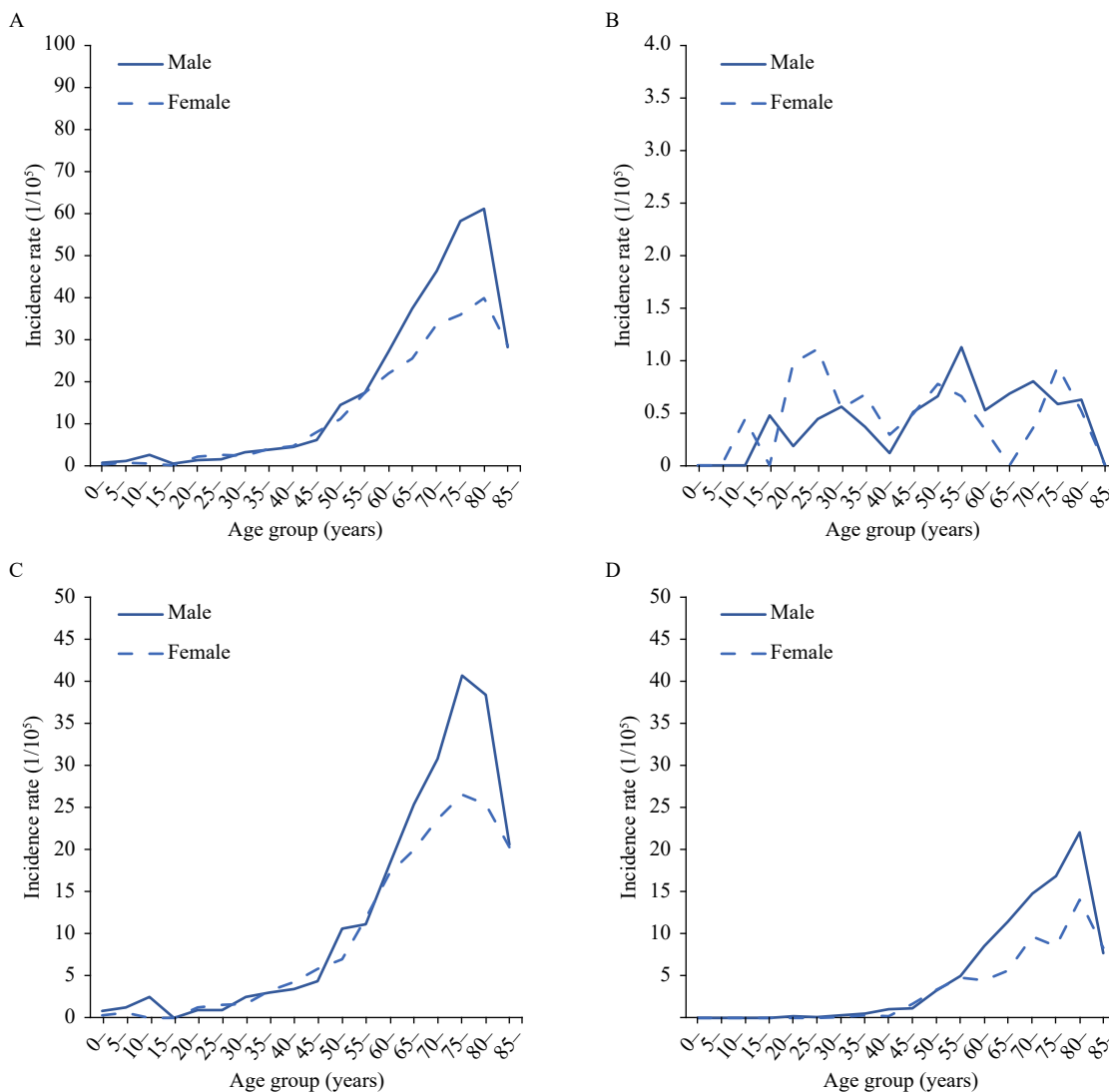
FIGURE 3. Age-specific incidence rates of lymphoma by gender and subtype in Beijing, 2019 (MIS-CASS). (A) Lymphoma (C81–85, C88, C90, C96); (B) Hodgkin lymphoma (C81); (C) Non-Hodgkin lymphoma (C82–85, C96); (D) Multiple myeloma (C88, C90).
Abbreviation: MIS-CASS=Medical-Insurance-System-based Cancer Surveillance System.

countries have established the national or subnational cancer incidence surveillance system based on this data source, partly due to relatively low coverage of health insurance, popularity of private insurance, decentralization of data management, etc. In this regard, this study provided a new insight into the improvement of cancer incidence surveillance strategies for countries or regions with uniform and high-quality health insurance system covering the whole population.

This study was subject to at least two limitations. First, this was still a single-center study focusing only on lymphoma, and similar comparative studies including more cancer types are warranted in other regions in China that are similarly covered by high-

quality PBCRs. Second, the inclusion of more diagnosis-related information (such as ICD-O-3 codes and grading) into the claims database is deemed necessary to further enhance the accuracy of cancer diagnosis ascertainment and to enrich the application of MIS-CASS.

In conclusion, the present head-to-head comparative study provides direct evidence that MIS-CASS can estimate cancer incidence in a complete and timely manner. Compared with the current PBCRs in China that rely upon incomplete designated hospitals and a strict household registration system, the MIS-CASS has more comprehensive coverage of incident cancer cases and local population, and thus better reflects the actual cancer burden and consumption of medical
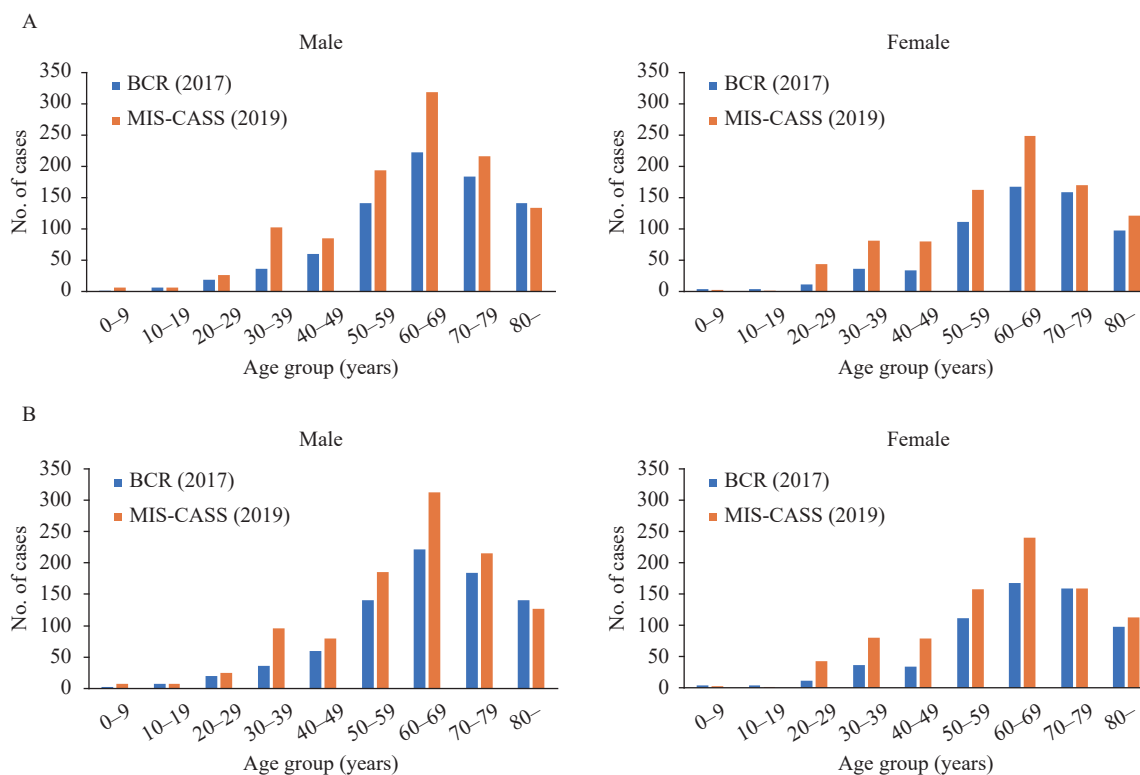
FIGURE 4. Comparison of the absolute number of new lymphoma cases by age group in Beijing between MIS-CASS (2019) and BCR (2017). (A) Including all designated medical insurance hospitals for MIS-CASS. (B) Including only BCR-designated hospitals for MIS-CASS.

Note: The latest lymphoma incidence data of BCR (year of 2017) were derived from *Beijing Cancer Registry Annual Report 2020* (*12*).

Abbreviations: MIS-CASS=Medical-Insurance-System-based Cancer Surveillance System; BCR=Beijing Cancer Registry.

resources in a given area. With the advent of the era of big data and the increasing availability of standardized electronic health-related records, MIS-CASS should have more importance attached in reforming the nationwide cancer incidence surveillance. We advocate sharing claims data by medical insurance administration departments with disease control departments, so that this new approach for cancer incidence surveillance can be efficiently implemented in broader areas in China, especially where PBCRs have not been established so far.

Conflicts of interest: No conflicts of interest.

Corresponding authors: Zhonghu He, zhonghuhe@foxmail.com; Jun Zhu, zhu-jun2017@outlook.com.

1 Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education/Beijing), Laboratory of Genetics, Peking University Cancer Hospital & Institute, Beijing, China; 2 Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education/Beijing), Department of Lymphoma, Peking University Cancer Hospital & Institute, Beijing, China.
& Joint first authors.

REFERENCES

1. Bray F, Znaor A, Cueva P, Korir A, Swaminathan R, Ullrich A, et al. Planning and developing population-based cancer registration in low- and middle-income settings. Lyon: International Agency for Research on Cancer; 2014. IARC Technical Publication No. 43. https://www.who.int/immunization/hpv/iarc_technical_report_no43.pdf.
2. Wei WQ, Zeng HM, Zheng RS, Zhang SW, An L, Chen R, et al. Cancer registration in China and its role in cancer prevention and control. Lancet Oncol 2020;21(7):e342 – 9. http://dx.doi.org/10.1016/s1470-2045(20)30073-5.
3. He J, Wei WQ. 2019 China cancer registry annual report. Beijing:

People's Medical Publishing House. 2021. https://www.dushu.com/book/13807902/. (In Chinese).

4. Barber SL, Yao L. Health insurance systems in China: a briefing note. Geneva: World Health Organization; 2010. World Health Report (2010) Background Paper, No. 37. https://www.who.int/healthsystems/topics/financing/healthreport/37ChinaB_YFINAL.pdf. [2020-10-23].

5. Shi C, Liu MF, Liu Z, Guo CH, Li FL, Xu RP, et al. Using health insurance reimbursement data to identify incident cancer cases. J Clin Epidemiol 2019;114:141 – 9. http://dx.doi.org/10.1016/j.jclinepi.2019.06.009.

6. Tian HR, Xu RP, Li FL, Guo CH, Zhang LX, Liu Z, et al. 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. http://dx.doi.org/10.21147/j.issn.1000-9604.2019.04.13.

7. Tian HR, Yang W, Hu YJ, Liu Z, Chen L, Lei L, et al. Estimating cancer incidence based on claims data from medical insurance systems in two areas lacking cancer registries in China. EClinicalMedicine 2020;20:100312. http://dx.doi.org/10.1016/j.eclinm.2020.100312.

8. Yang L, Yuan YN, Sun TT, Li HC, Wang N. Population-based cancer incidence analysis in Beijing, 2008-2012. Chin J Cancer Res 2015;27(1):13 – 21. http://dx.doi.org/10.3978/j.issn.1000-9604.2015.01.07.

9. Bray F, Colombet M, Mery L, Piñeros M, Znaor A, Zanetti R, et al. Cancer Incidence in Five Continents, Vol. XI. Lyon: International Agency for Research on Cancer; 2021. IARC Scientific Publications, No. 166. https://publications.iarc.fr/Databases/Iarc-Cancerbases/Cancer-Incidence-In-Five-Continents-Vol.-XI-2017.

10. Liu S, Yang L, Yuan YN, Li HC, Tian J, Lu SJ, et al. Cancer incidence in Beijing, 2014. Chin J Cancer Res 2018;30(1):13 – 20. http://dx.doi.org/10.21147/j.issn.1000-9604.2018.01.02.

11. Jensen OM, Parkin DM, MacLennan R, Muir CS, Skeet RG. Cancer registration: principles and methods. Lyon: International Agency for Research on Cancer; 1991. IARC Scientific Publication No. 95. https://publications.iarc.fr/Book-And-Report-Series/Iarc-Scientific-Publications/Cancer-Registration-Principles-And-Methods-1991.

12. Ji JF. Beijing cancer registry annual report 2020. Beijing: Peking University Medical Press. 2021. (In Chinese).

13. Bray F, Parkin DM. Evaluation of data quality in the cancer registry: principles and methods. Part I: comparability, validity and timeliness. Eur J Cancer 2009;45(5):747 – 55. http://dx.doi.org/10.1016/j.ejca.2008.11.032.

14. Parkin DM, Bray F. Evaluation of data quality in the cancer registry: principles and methods Part II. Completeness. Eur J Cancer 2009;45(5):756 – 64. http://dx.doi.org/10.1016/j.ejca.2008.11.033.

15. National Cancer Center. Chinese guideline for cancer registration. Beijing: People's Medical Publishing House. 2016. (In Chinese).

SUPPLEMENTARY MATERIAL
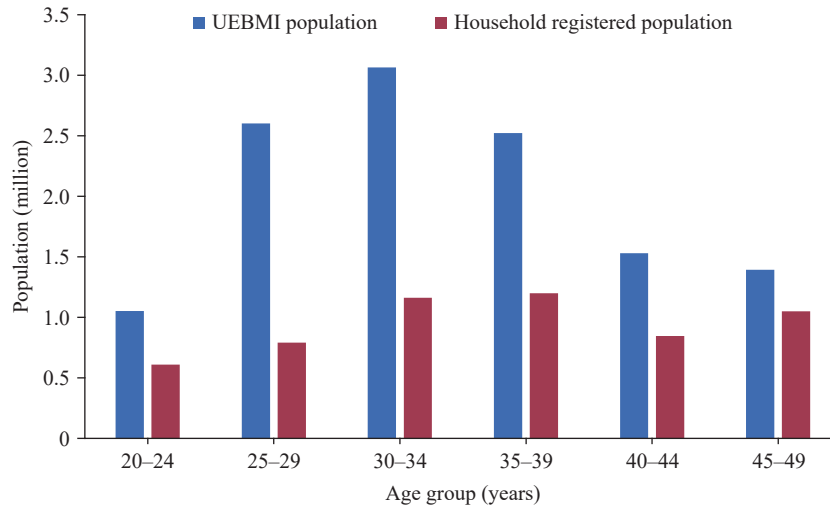
SUPPLEMENTARY FIGURE S1. The age distribution of UEBMI population and household registered population (the young and the middle-aged, 20–49 years) in Beijing, 2019.
Abbreviation: UEBMI=Urban Employee Basic Medical Insurance.