Original Research

Development and Implementation of a Dynamically Updated Big Data Intelligence Platform Using Electronic Medical Records for Secondary Hypertension

Nanfang Li[1],*,[†]●, Qing Zhu[1],[†], Yujie Dang[1], Yin Zhou[2], Xintian Cai[1], Mulalibieke Heizhati[1], Delian Zhang[1], Xiaoguang Yao[1], Qin Luo[1], Junli Hu[1], Guoliang Wang[1], Yingchun Wang[1], Menghui Wang[1], Jing Hong[1]

[1]Hypertension Center of People's Hospital of Xinjiang Uygur Autonomous Region; Xinjiang Hypertension Institute; National Health Committee Key Laboratory of Hypertension Clinical Research; Key Laboratory of Xinjiang Uygur Autonomous Region "Hypertension Research Laboratory"; Xinjiang Clinical Medical Research Center for Hypertension (Cardio-Cerebrovascular) Diseases, 830001 Urumqi, Xinjiang, China
[2]Medical Department, Yidu Cloud (Beijing) Technology Co., Ltd., 100191 Beijing, China
*Correspondence: lnanfang2016@sina.com (Nanfang Li)
[†]These authors contributed equally.

Abstract

Background: The accurate identification and diagnosis of secondary hypertension is critical, especially while cardiovascular heart disease continues to be the leading cause of death. To develop a big data intelligence platform for secondary hypertension using electronic medical records to contribute to future basic and clinical research. **Methods**: Using hospital data, the platform, named Hypertension DATAbase at Urumchi (UHDATA), included patients diagnosed with hypertension at the People's Hospital of Xinjiang Uygur Autonomous Region since December 2004. The electronic data acquisition system, the database synchronization technology, and data warehouse technology (extract–transform–load, ETL) for the scientific research big data platform were used to synchronize and extract the data from each business system in the hospital. Standard data elements were established for the platform, including demographic and medical information. To facilitate the research, the database was also linked to the sample database system, which includes blood samples, urine specimens, and tissue specimens. **Results**: From December 17, 2004, to August 31, 2022, a total of 295,297 hypertensive patients were added to the platform, with 53.76% being males, with a mean age of 59 years, and 14% with secondary hypertension. However, 75,802 patients visited the Hypertension Center at our hospital, with 43% (32,595 patients) being successfully diagnosed with secondary hypertension. The database contains 1458 elements, with an average fill rate of 90%. The database can continuously include the data for new hypertensive patients and add new data for existing hypertensive patients, including post-discharge follow-up information, and the database updates every 2 weeks. Presently, some studies that are based on the platform have been published. **Conclusions**: Using computer information technology, we developed and implemented a big database of dynamically updating electronic medical records for patients with hypertension, which is helpful in promoting future research on secondary hypertension.

Keywords: secondary hypertension; database; electronic medical records

1. Introduction

Hypertension is a pressing public health concern worldwide and a major modifiable risk factor for cardio-cerebrovascular disease and renal failure [1]. Hypertension can be divided into primary hypertension with unknown causes and secondary hypertension with clear causes [2]. Previously, primary hypertension was believed to be the most common form of the disease, with secondary hypertension only accounting for 5–10% of cases [3–6]. However, with a comprehensive understanding of the etiology of hypertension and the improvements in clinical diagnostic techniques, the proportion of secondary hypertension has been found to significantly exceed current expectations [4,7,8]. According to existing research and clinical observations, secondary hypertension often leads to more serious target organ damage than primary hypertension [9–13]. Therefore, early detection and treatment of secondary hypertension are of crucial clinical significance.

However, the etiology of secondary hypertension is complex, involving cardiovascular diseases, endocrine diseases, kidney diseases, sleep, mental illness, and other disciplines (**Supplementary Table 1**). Thus, it is necessary to pay attention to the standardized procedures during screening to avoid the occurrence of missed diagnoses and misdiagnoses. It is not advisable to blindly screen everyone for secondary hypertension since it will result in a waste of medical resources and cause an economic burden on both the patients and society. Therefore, it requires the involvement of specialized departments, teams, and clinicians [2,14].

Since 1997, the Hypertension Center at the People's Hospital of Xinjiang Uygur Autonomous has gradually established a platform for secondary hypertension screening, diagnosis, and treatment (**Supplementary Figs. 1,2**). A detailed description of our center is provided in the **Supplementary Materials**. Since its establishment, the Hypertension Center has provided diagnostic and treatment services for each patient with hypertension from various regions of Xinjiang. Among all patients who visit our center, some do so to clarify their hypertension etiology, while others are refractory hypertension patients who have been referred by primary hospitals. For patients who visit our hypertension center, we first judged whether they were clearly hypertensive and measured their blood pressure level according to their home blood pressure, office blood pressure, or ambulatory blood pressure results. Next, patients were given corresponding examinations to evaluate their cardiovascular risk factors and the damage to their target organs (Fig. 1). In addition, doctors with 5 years or more working experience judge whether there are screening clues for secondary hypertension according to any symptoms and signs demonstrated by the patients and the results of routine examination (**Supplementary Table 2**). Patients with positive clues will be screened for secondary hypertension (Fig. 1). All our test items were performed in accordance with the standards of international guidelines [15], such as polysomonography (PSG) test preparation and interpretation standards for obstructive sleep apnea syndrome (OSAS) [2,7,8], and the requirements of renin aldosterone determination on drugs and disease status during primary aldosteronism (PA) screening [2,8,16–18]. The procedures of screening, diagnosis, and treatment for common secondary hypertension are shown in **Supplementary Figs. 3,4,5,6**.

At present, it is one of the main problems that we need to solve to determine the main influencing factors of secondary hypertension on target organs and cardiovascular diseases. In addition, there are still many unknown problems to be solved in the different types of secondary hypertension, such as the different degrees of target organ damage, and different treatment methods. It is also urgent to optimize the process of PA diagnosis and lateral typing. Therefore, we analyzed the data for every secondary hypertension case to solve these various clinical problems. However, in previous clinical research, medical data were dispersed across several business information systems within medical institutions, which resulted in data structure diversity, unstable quality, and inconsistent data standards. Detailed data collation is highly dependent on the manual operation of doctors. However, the cumbersome manual entry process leads to high error rates, difficulties in organizing the data, and the arrangement of complex clinical events, such as long research feasibility assessments. Consequently, this array of issues can hinder the effective integration and utilization of medical data [19]. In 2019, a project was launched in our hospital to develop a big intelligent database for hypertension by integrating current electronic medical records to provide high-quality basic and clinical research on hypertension, especially secondary hypertension.
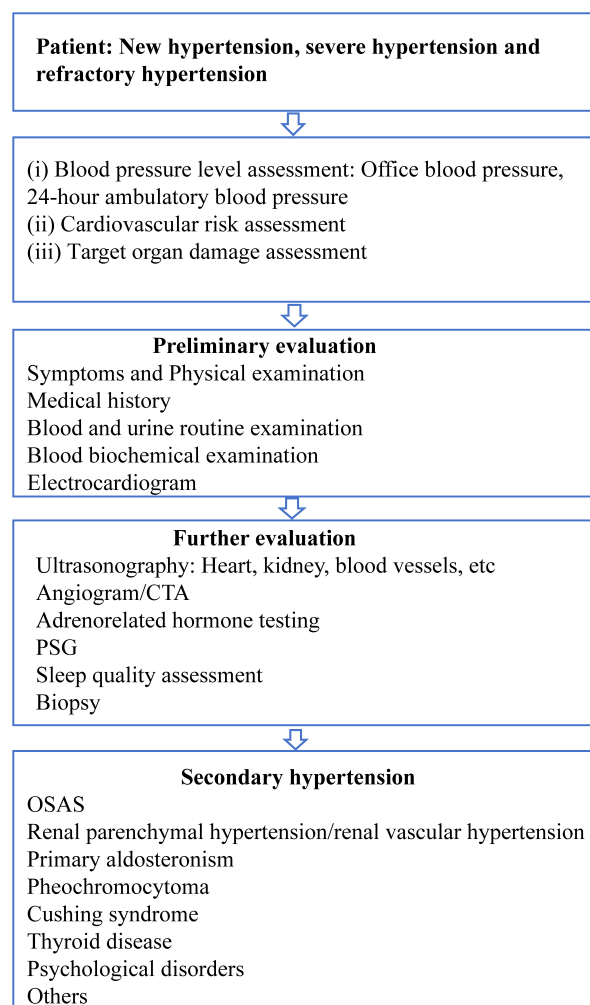

Fig. 1. Flow chart for diagnosis and treatment of hypertensive patients. CTA, computed tomography angiography; PSG, polysomonography; OSAS, obstructive sleep apnea syndrome.

2. Methods

2.1 Data Source and Procedure of Hypertension-Specific Database

Data were obtained from electronic medical records for hypertensive patients at the People's Hospital of the Xinjiang Uygur Autonomous Region in China.

The project included four procedures: (a) full-scale, multi-source, and heterogeneous data extraction and integration in hospitals; (b) data structurization and normalization; (c) development of hypertension-specific standard data elements; (d) construction of functional modules on the platform. In terms of integrating multi-source and het-

erogeneous data, the database synchronization technology of the scientific research big data platform and data warehouse technology (extract–transform–load, ETL) is used to synchronize and extract the data from various business systems in the hospital, to realize the collection, aggregation, and cleaning of multi-source and heterogeneous data for multiple information systems in the hospital (Fig. 2) [20–22]. Concurrently, it supports the import of information islands and data from external sources to ensure that eligible cases are automatically and continuously entered into the database and a full, continuous, complete, and reusable data asset is formed. In the process of data integration, it is worth noting that the corresponding relationship between the hypertension-specific disease data set, business activity, and the data source needs to be systematically elucidated. Conversely, it ensures that the hypertension-specific disease scientific research database comprehensively covers all information systems in the hospital and ensures comprehensive and effective data collection. Additionally, the traceability of individual data (business systems, business activities) can be ensured (**Supplementary Fig. 7**). Cardiologists, professional doctors with deputy senior titles or above from our hypertension center, informatics technicians, and software engineers from a technology company (Yidu Cloud Company) were all involved in this project. This database construction scheme was approved by the Ethics Committee for our hospital in 2019.

2.2 Establishment of Hypertension-Specific Standard Data Elements

To define all the variables in the platform, standard data elements were established, including 26 data element modules with a total of 1458 common and hypertension-specific variables (**Supplementary Fig. 8** and **Supplementary Table 3**). Hypertension-specific data elements were constructed with reference to international standards and local databases and terminology standards, as shown in **Supplementary Table 4**.

2.3 Data Extraction and Integration

Patients with hypertension who had visited our hospital since 2004 (our hospital began to use electronic medical records in 2004) were included in the platform, including patients with primary hypertension and secondary hypertension. Inclusion criteria: Visit type = hospitalization, and the diagnosis name during the visit contains "hypertension/HT" or an international classification of diseases (ICD)10 code containing "I10"–"I15". For each patient, data were extracted from 13 EHR systems [20] and a follow-up system (**Supplementary Figs. 2,7**). Additionally, it was also connected to the physical examination information and biological sample bank system.

Data extractions were performed using ETL according to the predefined standard data elements [20]. In this process, we used the unique identification number provided by the Enterprise Patient Master Index (EPMI) to consolidate each record for the same patient from 16 systems into one record.

2.4 Data Structurization and Normalization

We used natural language processing (NLP) and multiple machine learning models for data structurization and normalization. Detailed processes of NLP are presented in previous research [20]. At present, a total of 1458 variables have been produced: 920 first-level variables, 440 second-level variables, and 98 third-level variables. By constantly modifying and constantly optimizing, the precision and recall rate for most variables can reach more than 95%.

2.5 Functional Modules on the Platform

To facilitate the research, eight functional modules were designed for the platform, including case retrieval and extraction, project creation, statistical analysis, follow-up, etc.

2.6 Patient Privacy and Platform Security

To protect the privacy and data security of each patient, we desensitized their personal privacy information and established different permission sets specific to each patient's need [19,23].

3. Results

3.1 Overview of the Platform

From December 17, 2004, to August 31, 2022, a total of 295,297 hypertensive patients were added to the platform, with 53.76% being males, with a mean age of 59 years, and 14% with secondary hypertension (Fig. 3). However, 75,802 patients visited the Hypertension Center at our hospital, with 43% (32,595 patients) being successfully diagnosed with secondary hypertension. For the hypertension classification, 158,663 people (53.73%) were in grade 3, 109,998 people (37.25%) were in grade 2, and 26,665 people (9.03%) were in grade 1. From 2010 to 2022, the annual number of patients with hypertension gradually increased, and the number of patients affected during the epidemic decreased in 2020 (Fig. 3).

3.2 Examples of Advanced Search

To build a condition tree, advanced retrieval uses multi-dimensional inclusion and exclusion criteria to retrieve the diagnosis and treatment information for each patient. For example, if the age at treatment is ≥ 18, and the diagnosis name is hypertension, there are 254,672 eligible people. Then, a total number of hospitalizations of $\geq$ 2,108,109 people also meet the conditions (**Supplementary Fig. 9**).

3.3 eCRF

An electronic case report form (eCRF) can be created on the platform for prospective clinical trials and cohort
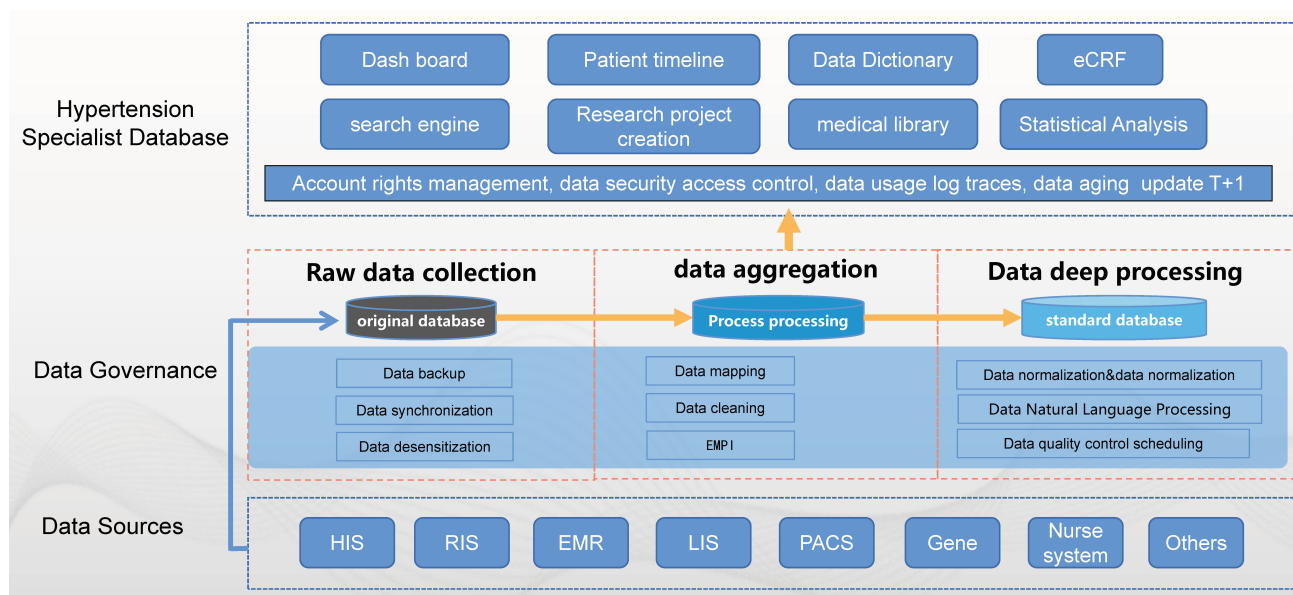
Fig. 2. Framework of database construction. HIS, hospital information system; EMR, electronic medical record; RIS, radiology information system; LIS, laboratory information system; PACS, picture archiving and communication system; eCRF, electronic case report form; EMPI, Enterprise Master Patient Index.

studies. Moreover, through follow-ups and an extension of the hospital's service function, the prognosis of the disease can be identified to a certain extent, thereby providing a more comprehensive understanding of the disease.

3.4 Platform Implementation Status

The platform was launched in December 2019 and has been adopted by several clinicians and graduate students. At present, more than 30 retrospective studies and 2 prospective cohort studies are either being performed or have been conducted based on the platform. Six studies have been published, involving studies on aldosterone and the brain and macrovascular complications [24–29].

4. Discussion

A research disease database is an important tool for clinicians to collect and organize data and conduct scientific research. At present, the method of collecting and sorting data is still mainly manual, which provides some problems, such as heavy workload, low efficiency, high error rate, and difficulty sharing and utilizing collected and sorted data. Following the construction of the diversified scientific research disease database for clinical departments under the big data environment, a number of scientific research databases have also been produced based on electronic medical record system [30–34], which is the result of the continuous development and intersection of computer technology, information technology, and medicine. Moreover, CRF visual personalized customization and template management, data integration and integration, batch import, automation and auxiliary control filling, and multi-way data

security management, have been implemented to assist the department in performing clinical research to the highest quality and efficiency.

Hypertension affects nearly 1.3 billion people worldwide, and the burden of cardiovascular and cerebrovascular diseases caused by hypertension remains first in the global disease burden list. In particular, damage to the target organs by secondary hypertension is more serious than in primary hypertension, meaning the mechanisms need to be explored and addressed through more clinical or basic research. Thus, our hospital has built a big data intelligence platform for hypertension research (hypertension-specialist database), named Hypertension DATAbase at Urumchi (UHDATA), to promote research on hypertension (including secondary hypertension), using the electronic data acquisition system, the database synchronization technology, and data warehouse technology (extract–transform–load, ETL) on the scientific research big data platform. Presently, data on nearly 300,000 hypertensive patients are available on the platform. To the best of our knowledge, this is currently the largest hypertension-specialized database and the largest database for secondary hypertension.

At present, most databases based on electronic medical record systems are specialized disease databases from a single center or hospital. Although it does not contain information on all patients of multiple centers or even an entire province, it contains more comprehensive and detailed data than found in electronic health insurance databases and commercial insurance databases, especially in real-world studies, which may need to control for many confounding factors to ensure the reliability of the study. In addition, the electronic medical record database contains more imag-
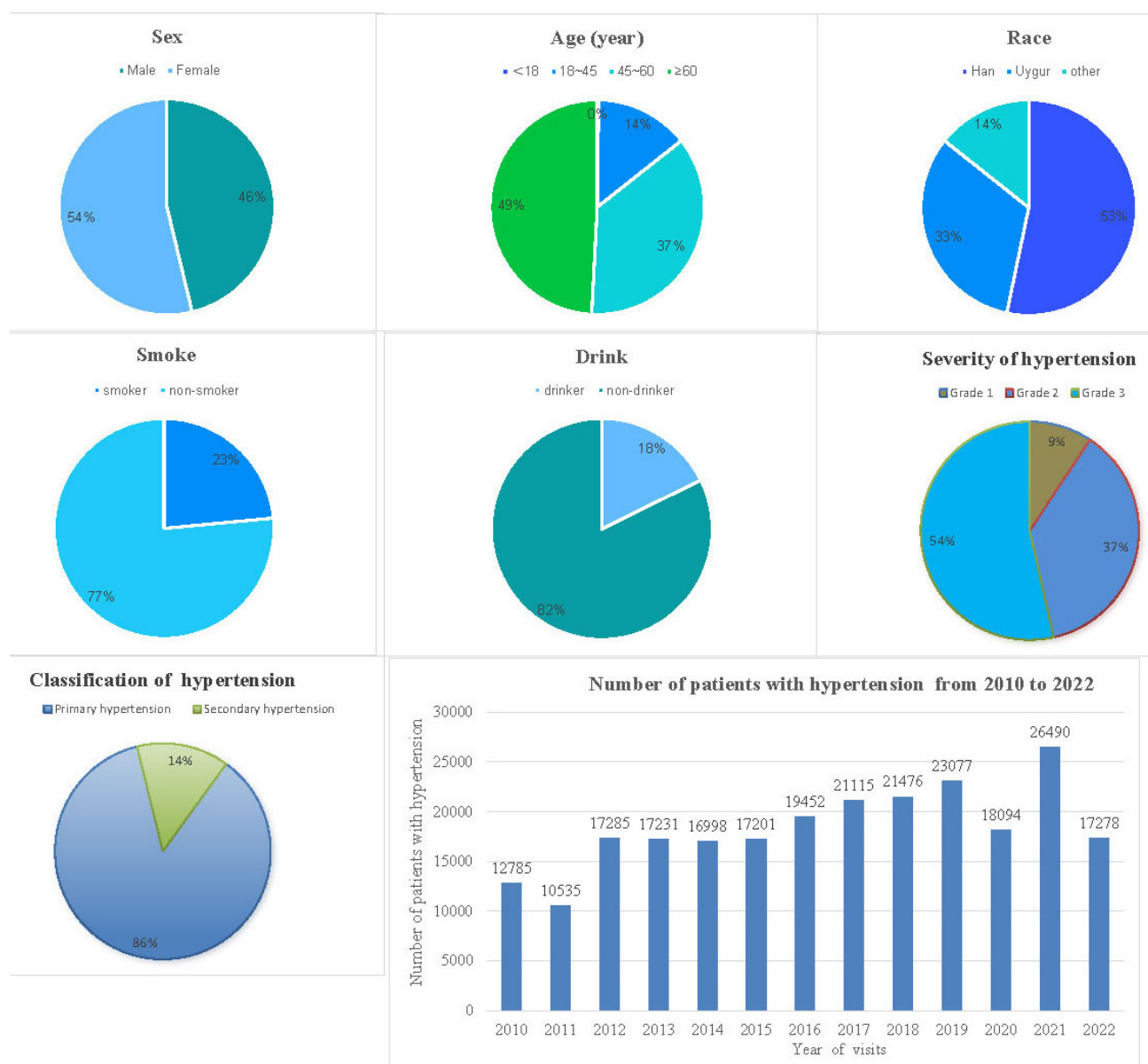
Fig. 3. The basic characteristics of hypertensive patients on the platform.

ing data and laboratory indexes, which are very helpful for diagnosing and treating secondary hypertension and any associated complications [35,36]. This will be very helpful in developing robots that can perform clinical diagnosis and treatment in the future.

Our big data platform has several limitations. First, the medical data archived within the database were not originally intended for secondary analysis. Thus, some missing values and inconsistencies may occur due to technical errors, system integration, and data preprocessing, which need continuous improvement. Second, there are currently no target standards for obtaining data elements or any definitions for those standard elements. Third, using such databases in either retrospective studies or real-world research would not replace the need to perform randomized controlled trials. However, they could serve as an important

tool to supplement the contributions of trials for evidence-based medicine. Finally, it currently only exists as a local database from a single institution. Thus, to address the problem of data sharing, the data from multiple hospitals should be shared and implemented in the future.

5. Conclusions

We developed and implemented a big database of electronic medical records for hypertension that dynamically updates, and offers an important perspective on the future study of secondary hypertension.

Availability of Data and Materials

The datasets for this study are available from the corresponding author on reasonable request.

Author Contributions

Ethics Approval and Consent to Participate

Acknowledgment

Funding

Conflict of Interest

Supplementary Material

References

[1] Diao X, Huo Y, Yan Z, Wang H, Yuan J, Wang Y, *et al*. An Application of Machine Learning to Etiological Diagnosis of Secondary Hypertension: Retrospective Study Using Electronic Medical Records. JMIR Medical Informatics. 2021; 9: e19739.

[2] Nanfang Li. Secondary Hypertension Screening, Diagnosis and Treatment. Springer Nature: Singapore Pte Ltd. 2020.

[3] Rossi GP, Seccia TM, Pessina AC. Secondary hypertension: the ways of management. Current Vascular Pharmacology. 2010; 8: 753–768.

[4] Charles L, Triscott J, Dobbs B. Secondary Hypertension: Discovering the Underlying Cause. American Family Physician. 2017; 96: 453–461.

[5] Hirsch JS, Hong S. The Demystification of Secondary Hypertension: Diagnostic Strategies and Treatment Algorithms. Current Treatment Options in Cardiovascular Medicine. 2019; 21: 90.

[6] Morimoto R, Omata K, Ito S, Satoh F. Progress in the Management of Primary Aldosteronism. American Journal of Hypertension. 2018; 31: 522–531.

[7] Li NF, Wang L, Zhou KM, Wang XL, Zu FY, Zhang DL, *et al*. Analysis of etiology of the patients with hypertension from the People's Hospital of Xinjiang Uygur Autonomous Region. Zhonghua Xin Xue Guan Bing Za Zhi. 2007; 35: 865–868.

[8] Wang L, Li N, Yao X, Chang G, Zhang D, Heizhati M, *et al*. Detection of Secondary Causes and Coexisting Diseases in Hypertensive Patients: OSA and PA Are the Common Causes Associated with Hypertension. BioMed Research International. 2017; 2017: 8295010.

[9] Rimoldi SF, Scherrer U, Messerli FH. Secondary arterial hypertension: when, who, and how to screen? European Heart Journal. 2014; 35: 1245–1254.

[10] Huang X, Yu S, Xiao H, Pei L, Chen Y, Chen W, *et al*. Comparison of Clinical Features between Primary Aldosteronism and Essential Hypertension in Chinese Patients: A Case-Control Study. International Journal of Endocrinology. 2021; 2021: 6685469.

[11] Kobayashi H, Abe M, Nakamura Y, Takahashi K, Fujita M, Takeda Y, *et al*. Association Between Acute Fall in Estimated Glomerular Filtration Rate After Treatment for Primary Aldosteronism and Long-Term Decline in Renal Function. Hypertension (Dallas, Tex.: 1979). 2019; 74: 630–638.

[12] Onohara T, Takagi T, Yoshida K, Iizuka J, Okumi M, Kondo T, *et al*. Assessment of postoperative renal function after adrenalectomy in patients with primary aldosteronism. International Journal of Urology: Official Journal of the Japanese Urological Association. 2019; 26: 229–233.

[13] Lopez-Vargas PA, Tong A, Sureshkumar P, Johnson DW, Craig JC. Prevention, detection and management of early chronic kidney disease: a systematic review of clinical practice guidelines. Nephrology (Carlton, Vic.). 2013; 18: 592–604.

[14] Hirsch JS, Hong S. The Demystification of Secondary Hypertension: Diagnostic Strategies and Treatment Algorithms. Current Treatment Options in Cardiovascular Medicine. 2019; 21: 90.

[15] Williams B, Mancia G, Spiering W, Agabiti Rosei E, Azizi M, Burnier M, *et al*. 2018 ESC/ESH Guidelines for the management of arterial hypertension. European Heart Journal. 2018; 39: 3021–3104.

[16] Funder JW, Carey RM, Mantero F, Murad MH, Reincke M, Shibata H, *et al*. The Management of Primary Aldosteronism: Case Detection, Diagnosis, and Treatment: An Endocrine Society Clinical Practice Guideline. The Journal of Clinical Endocrinology & Metabolism. 2016; 101: 1889–1916.

[17] Luo Q, Li N, Wang M, Yao X, Heizhati M, Zhang D, *et al*. Mild primary aldosteronism (PA) followed by overt PA are possibly the most common forms of low renin hypertension: a single-center retrospective study. Journal of Human Hypertension. 2020; 34: 633–640.

[18] Adilijiang M, Luo Q, Wang M, Zhang D, Yao X, Wang G, *et al*. Minor Change of Plasma Renin Activity during the Saline Infusion Test Provide an Auxiliary Diagnostic Value for Primary Aldosteronism. International Journal of Endocrinology. 2021; 2021: 5757305.

[19] Li B, Li J, Jiang Y, Lan X. Experience and reflection from China's Xiangya medical big data project. Journal of Biomedical Informatics. 2019; 93: 103149.

[20] Lin L, Liang W, Li CF, Huang XD, Lv JW, Peng H, *et al*. Devel-

opment and implementation of a dynamically updated big data intelligence platform from electronic health records for nasopharyngeal carcinoma research. The British Journal of Radiology. 2019; 92: 20190255.

[21] Fang Z, Fan X, Chen G. A study on specialist or special disease clinics based on big data. Frontiers of Medicine. 2014; 8: 376–381.

[22] Denney MJ, Long DM, Armistead MG, Anderson JL, Conway BN. Validating the extract, transform, load process used to populate a large clinical research database. International Journal of Medical Informatics. 2016; 94: 271–274.

[23] Nakaya N, Goto K, Saito-Nakaya K, Inagaki M, Otani T, Akechi T, *et al*. The lung cancer database project at the National Cancer Center, Japan: study design, corresponding rate and profiles of cohort. Japanese Journal of Clinical Oncology. 2006; 36: 280–284.

[24] Yuan Y, Li N, Liu Y, Zhu Q, Heizhati M, Zhang W, *et al*. Positive Association Between Plasma Aldosterone Concentration and White Matter Lesions in Patients With Hypertension. Frontiers in Endocrinology. 2021; 12: 753074.

[25] Yuan Y, Li N, Liu Y, Wang M, Heizhati M, Zhu Q, *et al*. Plasma aldosterone concentration is associated with white matter lesions in patients with primary aldosteronism. Endocrine. 2022; 75: 889–898.

[26] Zhu Q, Heizhati M, Lin M, Wang M, Yao X, Gan L, *et al*. Higher Plasma Aldosterone Concentrations Are Associated With Elevated Risk of Aortic Dissection and Aneurysm: a Case-Control Study. Hypertension (Dallas, Tex.: 1979). 2022; 79: 736–746.

[27] Li N, Wen W, Cai X, Zhu Q, Hu J, Heizhati M, *et al*. The Use of Aspirin Increases the Risk of Major Adverse Cardiac and Cerebrovascular Events in Hypertensive Patients with Obstructive Sleep Apnea for the Primary Prevention of Cardiovascular Disease: A Real-World Cohort Study. Journal of Clinical Medicine. 2022; 11: 7066.

[28] Cai X, Li N. Association between Use of Spironolactone and Risk of Stroke in Hypertensive Patients: A Cohort Study. Pharmaceuticals (Basel, Switzerland). 2022; 16: 57.

[29] Wang H, Heizhati M, Li N, Gan L, Lin M, Yang W, *et al*. Association of objective and subjective parameters of obstructive sleep apnea with plasma aldosterone concentration in 2066 hypertensive and 25,368 general population. Frontiers in Endocrinology. 2023; 13: 1016804.

[30] Tannen RL, Weiner MG, Xie D. Use of primary care electronic medical record database in drug efficacy research on cardiovascular outcomes: comparison of database and randomised controlled trial findings. BMJ (Clinical Research Ed.). 2009; 338: b81.

[31] Broberg CS, Mitchell J, Rehel S, Grant A, Gianola A, Beninato P, *et al*. Electronic medical record integration with a database for adult congenital heart disease: Early experience and progress in automating multicenter data collection. International Journal of Cardiology. 2015; 196: 178–182.

[32] Jin S, Chen L, Chen K, Hu C, Hu S, Zhang Z. Establishment of a Chinese critical care database from electronic healthcare records in a tertiary care medical center. Scientific Data. 2023; 10: 49.

[33] Shao SC, Chan YY, Kao Yang YH, Lin SJ, Hung MJ, Chien RN, *et al*. The Chang Gung Research Database-A multi-institutional electronic medical records database for real-world epidemiological studies in Taiwan. Pharmacoepidemiology and Drug Safety. 2019; 28: 593–600.

[34] Liu M, Luo J, Li L, Pan X, Tan S, Ji W, *et al*. Design and development of a disease-specific clinical database system to increase the availability of hospital data in China. Health Information Science and Systems. 2023; 11: 11.

[35] Avvisato R, Forzano I, Varzideh F, Mone P, Santulli G. A machine learning model identifies a functional connectome signature that predicts blood pressure levels: imaging insights from a large population of 35 882 patients. Cardiovascular Research. 2023; 119: 1458–1460.

[36] Eisenhofer G, Durán C, Cannistraci CV, Peitzsch M, Williams TA, Riester A, *et al*. Use of Steroid Profiling Combined With Machine Learning for Identification and Subtype Classification in Primary Aldosteronism. JAMA Network Open. 2020; 3: e2016209.