REVIEW

Existing Data Sources in Clinical Epidemiology: The Taiwan National Health Insurance Laboratory Databases

Po-Chang Lee[1]
Feng-Yu Kao[1]
Fu-Wen Liang [ID] [2]
Yi-Chan Lee[3]
Sheng-Tun Li[4]
Tsung-Hsueh Lu [ID] [3]

[1]National Health Insurance Administration, Ministry of Health & Welfare, Taipei, Taiwan; [2]Department of Public Health, Kaohsiung Medical University, Kaohsiung, Taiwan; [3]Department of Public Health, College of Medicine, National Cheng Kung University, Tainan, Taiwan; [4]Department of Industrial and Information Management, College of Management, National Cheng Kung University, Tainan, Taiwan

Abstract: This paper provides an introduction to laboratory databases established by Taiwan National Health Insurance Administration (NHIA) since 2015 and released for research since June 2017. The National Health Insurance (NHI) is a government-run single-payer program introduced in 1995 that now covers more than 99% of 23 million Taiwanese citizens. To prevent duplication of medication prescriptions and laboratory test and examination prescriptions, contracted health care providers are required to upload the results of laboratory tests and reports of examinations to the NHIA. The cumulative number of laboratory test results was 5.64 billion from January 2015 to the end of August 2020 for 602 types of test. There are 35 variables for each laboratory test result stored in the databases that can be used for research. However, different hospitals might use different format in reporting the results. The researchers therefore have to develop algorithms to include and exclude incompatible records and to determine whether the results are positive or negative (normal or abnormal). The NHIA suggests that researchers release their source codes of algorithms so that other researchers can modify the codes to improve inter-study comparability. Through the unique personal identification number, the laboratory data can be linked to NHI inpatient and outpatient claims data for further value-added analyses. Non-Taiwanese researchers can collaborate with Taiwan researchers to access the NHI laboratory databases.
Keywords: biomarkers, database, data resource, laboratory information systems, National Health Insurance claims data, Taiwan

Introduction

Taiwan's National Health Insurance (NHI) is a government-run single-payer program introduced in 1995 that now covers more than 99% of 23 million Taiwanese citizens.[1–3] NHI claims data, both outpatient and inpatient claims data are released for research purposes.[4–6] The variables available for research are listed in supplementary Tables 1 and 2. At the end of 2020, more than 6800 papers had been published in PubMed-covered journals that employed Taiwan NHI claims data.[7] In December 2014, the Taiwan NHI Administration (NHIA) initiated a program that requires contracted health care providers to upload the results of laboratory tests and reports of examinations (chest radiography, magnetic resonance imaging, and cardiac catheterization) in addition to claims for reimbursements for these tests and examinations. The NHIA has compiled the uploaded results stored in the NHI databases and released them for research purposes since June 2017. In this review, we first explain why the NHIA initiated

Correspondence: Tsung-Hsueh Lu
Department of Public Health, College of Medicine, National Cheng Kung University, No. 1, Dah Hsueh Road, East District, Tainan, 701, Taiwan
Tel +886-6-2353535 ext. 5567
Fax +886-6-2359033
Email robertlu@mail.ncku.edu.tw

the program collecting the results of laboratory tests. We then describe the content of the laboratory databases and caveats that must be accounted for when analyzing the data. Lastly, we suggest how to add value to the databases.

Why Were the Databases Established?

Several forces drove the NHIA to begin collecting the results of laboratory tests and reports of examinations. The first was the implementation of pay-for-performance (P4P) programs for five diseases—diabetes mellitus, tuberculosis, breast cancer, cervical cancer, and asthma—in 2001 to encourage health care providers to perform more patient monitoring and follow-up care.[8] In the first period of the program's evolution, financial incentives were based on process indicators (eg, whether a hemoglobin A1C or lipid test was prescribed). In the second period, the NHIA began paying extra bonuses based on treatment outcome measures (eg, the percentage of patients with HbA1C level higher than 9.5% or low-density lipoprotein level higher than 130 mg/dL).[9,10] Health care providers participating in the P4P programs must upload the results of laboratory tests to the NHIA. The P4P programs initiated the early information technology (IT) infrastructure system for uploading the results of laboratory tests.

The second driving force was the desire to curb duplication of medication prescriptions, laboratory tests, and examination prescriptions in the present health care delivery system. Taiwan's NHI service has no gatekeeper system; therefore, patients have complete freedom of choice among providers when they seek care and can easily access sub-specialists in tertiary medical centers.[2,3] One of the adverse consequences of this free choice was the aforementioned duplications. To reduce the duplication of medication prescriptions, the NHIA launched the nationwide PharmaCloud platform in August 2013 on which physicians and pharmacists could obtain information for the preceding 3 months of a patient's medication prescribed in other medical settings.[11–13] The PharmaCloud platform has been extended into the MediCloud System; health care providers, after obtaining the informed consent of a patient, can query the health information exchange system to view the results of laboratory tests and reports of examinations performed in previous medical encounters to prevent duplicative testing and examinations.[14,15] The NHIA has also provided the personal health record

"MyHealth Bank" service for patients, who can view the results of laboratory tests and reports of examinations through their mobile phone and a desktop app.[16,17] The contracted hospitals, clinics, and laboratory centers are thus required to upload the results of laboratory tests and reports of examinations to the NHIA.[18] These programs scaled up the IT infrastructure system for uploading the results of laboratory tests.

The third driving force was demand from researchers. One of the often-mentioned limitations in using NHI claims data for research was the lack of information on disease severity.[6] For example, the International Classification of Diseases Tenth Revision (ICD-10) code E78.5 could be employed to identify people with hyperlipidemia, but it was impossible to differentiate between patients with hyperlipidemia who had an extremely high versus a moderately high cholesterol level; this reduced the explanatory power of using hyperlipidemia as an explanatory variable to predict the outcomes of interest. One of the solutions to this limitation was to obtain laboratory test result information that could indicate the disease severity. The NHIA has thus compiled the uploaded results of laboratory tests in databases released for research since June 2017. However, examination reports are not released for research because report content contains some personal identification information.

What Do the Databases Contain?
Infrastructure

There were 25 medical centers, 82 regional hospitals, 366 district hospitals, and 10,496 primary care clinics submitted 32, 44, 32, and 259 million outpatient claims, respectively, to the NHIA for reimbursement of the health care expenditures in 2019. All contracted health care providers submitted these electronic claims data to oracle database in the NHIA through the virtual private network (VPN). The NHIA provided some financial supports for increasing the bandwidth of VPN for uploading the results of laboratory tests and reports of examinations.

Almost all hospitals have laboratory centers and most of the laboratory tests were performed (sampled) in the hospitals and the results were automatically transmitted to hospital electronic medical record system. Some primary care clinics with large amount of patients have equipment for essential laboratory tests and performed (sampled) the tests in the clinics. On the other hand, most primary care clinics without laboratory equipment have contracted

laboratory centers to perform (sample) the tests. Some contracted laboratory centers would transmit the results of laboratory tests to the clinics electronically and some contracted laboratory centers just gave the results to the patients in paper form.

All hospitals contracted with the NHIA have to participate in the accreditation run by a third party 'Joint Commission of Taiwan' (https://www.jct.org.tw/mp-2.html) because the reimbursement payment would be different for different level of accreditation. Quality of laboratory tests and examinations is one important item in the hospital accreditation. Some of the hospitals in Taiwan further pursuit the International Organization for Standardization (ISO 15,189) certification,

Coverage

The coverage rate of results uploaded among claimed increased from 51% in 2015 to 78% in 2019 (Figure 1). The number of claims for laboratory tests in 2019 was 125 million from medical centers, 136 million from regional hospitals, 64 million from district hospitals, and 49 million from primary care clinics. However, the number of results uploaded was 113, 118, 53, and 8 million, respectively, with coverage rate of 90%, 87%, 82%, and 16%, respectively. The main reason for the low coverage rate in primary care clinics was that many clinics prescribed only small amount of laboratory tests which might not cover the uploading expense requested by the IT vendors.

Initially, the results of only 72 commonly prescribed laboratory tests had to be uploaded in December 2014. This number increased to 216 in August 2019 and then 641 in January 2020. The cumulative number of laboratory test results was 5.64 billion from January 2015 to the end of August 2020. Figure 2 illustrates the 20 most frequently
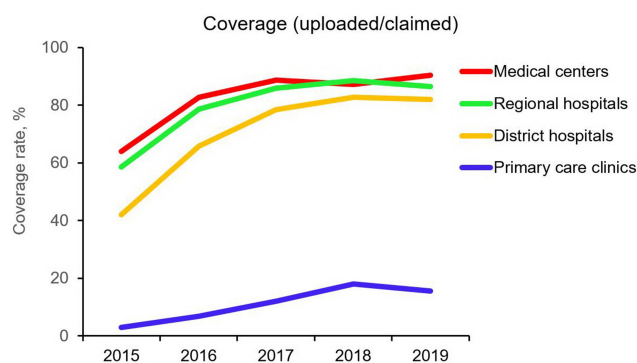

Figure 1 The coverage of laboratory tests results uploaded among claimed to the National Health Insurance by level of clinical setting in Taiwan, 2015 to 2019.

performed laboratory tests in Taiwan. The overall pattern is similar to that in other countries; nevertheless, the exact ranking of the particular laboratory tests might be different due to the differences in disease prevalence and prescription behaviors across countries.[19–24] The five most commonly performed tests in Taiwan were whole blood count (including the hemoglobin and white blood count), urine routine, white blood cell differential count, blood sugar, and blood creatinine. When considering the number of persons, the leading five tests were blood creatinine, glutamate pyruvate transaminase or alanine transaminase (GPT/ALT), blood sugar, glutamic-oxaloacetic transaminase or aspartate aminotransferase (GOT/AST), and potassium.

Variables

There are 35 variables available for research in the laboratory database. The name and description of the 35 variables are listed in Table 1. V1 to V10 are variables as key to link to outpatient and inpatient claims data. V11 to V16 are various date information with regard to the medical encounters (visits). V17 to V29 are information related to laboratory tests. It is noteworthy to mention that there are three dates related to the laboratory test. V19 is the date of uploading the result of the test, V20 is the date of prescribing the test, V21 is the date of the test been performed (sampled), and V23 is the sequence number of repeated tests in the same day.

We used the hepatitis B virus surface antigen (HBsAg) test and GOT/AST liver function test as examples to illustrate that the format of content recorded in the same variable might be varied across different hospitals (Table 2). Some hospitals used text form (No. 1–5 in Table 2) and some hospitals used numeric form (No. 6–10 in Table 2) to present the results under the same variable (ASSAY_VALUE in Table 2). Similarly, different hospitals used different units (S/N, S/CO, IU/mL, or COI) under the same variable (UNIT_DATA in Table 2). All of these formats are acceptable.

Quality

Several quality issues should be considered when using the laboratory databases. First, some hospitals might use laboratory codes that were not designated by the NHI. Second, the name of laboratory was not compatible with the NHI laboratory code. For example, in case No. 8 in Table 2, the name of laboratory test is GOT (AST); yet, the laboratory code is 14032C for HBsAg test.
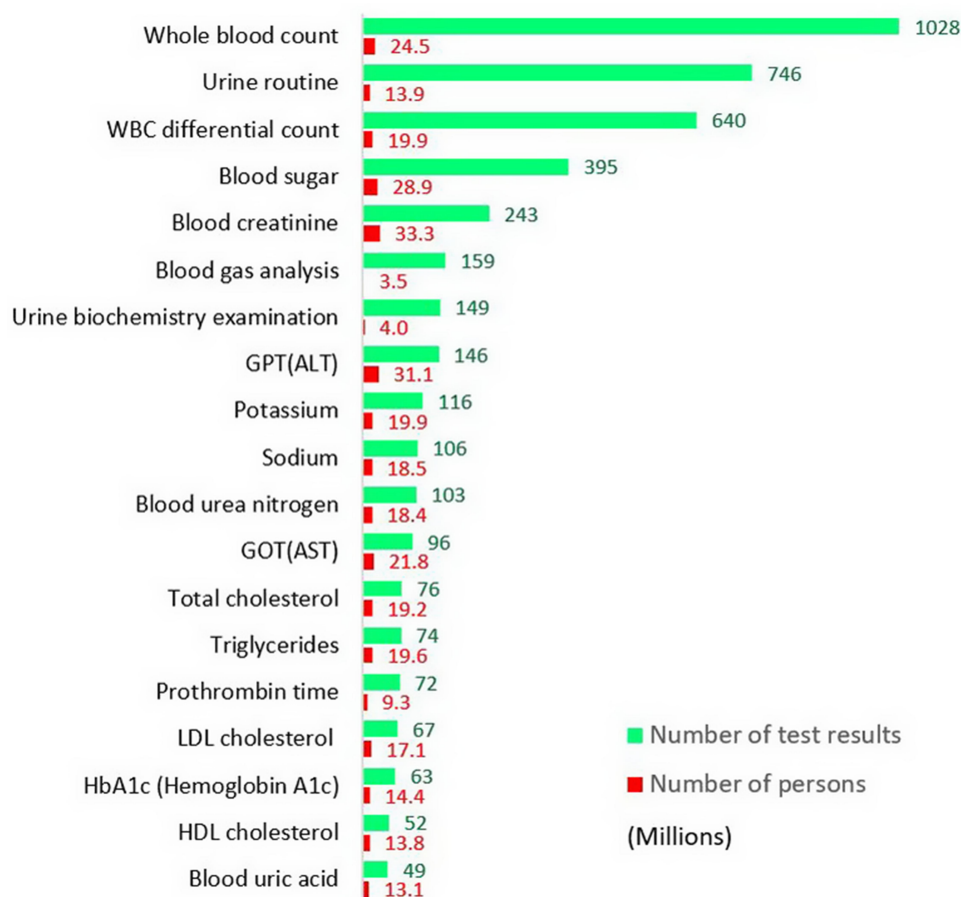
Figure 2 Cumulative number of laboratory test results and persons (in millions) for the 20 most frequent laboratory tests performed in Taiwan, from 2015 to the end of August 2020.
Abbreviations: GOT (AST), glutamic-oxaloacetic transaminase or aspartate aminotransferase; GPT (ALT), glutamate pyruvate transaminase (alanine transaminase); HDL, high-density lipoprotein; LDL, low-density lipoprotein; WBC, white blood cell.

According to the study of Sheu et al, of the 869,974 records with laboratory code 14032C (HBsAg test), 7.8% had incompatible name of laboratory tests in 2015. The proportion decreased to 3.5% (42,713/ 1,210,019) in 2019 after feedbacks from the NHIA to the hospitals uploaded the incompatible name of laboratory tests.[25] The main reason of the incompatibility issue was that many hospitals use a package of test (such as liver function test or renal function test) which included several laboratory tests, which might be uploaded repeatedly.

Similar incompatible problems occurred between different variables. For example, case No. 4 and case No. 5 in Table 2 had the same unit (IU/mL); but the cutting point was <0.05 and <0.2, respectively. In Case No. 6 in Table 2, the unit recorded in unit variable is S/N; nevertheless, the unit recorded in reference value is COI. The percentage of

this kind of incompatible problem was 12.6% in 2015 and decreased to 7.5% in 2019.[25]

How Can Value Be Added to the Databases?
Tackling the Incompatible Issues
Despite some of the quality issues mentioned in the previous section, several strategies could be applied to tackle these incompatible issues. First, we can cross-check the multiple results of the same test for the same patient to get robust results using the personal identification number and sequential number listed in Table 1. Second, for some records with incompatible laboratory test code, such as Case No. 8 in Table 2 which might be due to repeatedly uploading, the contents of GOT (AST) results are still valid and could be used for analysis.

Table 1 The Parameter Variables Available in the Taiwan National Health Insurance (NHI) Laboratory Information Databases

No.	Variable	Length	Description
V1	CASE_REPORT_TYPE	1	1. Laboratory test; 2. Image; 3. Pathology
V2	HOSP_ID_PKNO	10	Identification number of clinical setting
V3	HOSP_DATA_TYPE	2	Type of clinical setting
V4	FEE_YM	6	Year and month of payment
V5	APPL_TYPE	1	Type of claims application
V6	APPL_DATE	9	Year and month of claims application
V7	CASE_TYPE	2	Type of special payment program
V8	SEQ_NO	10	Sequential number set by the clinical setting
V9	ID	10	Identification number of beneficiary
V10	BIRTHDAY	8	Birth date of beneficiary
V11	FUNC_DATE	9	Date of beginning the clinical encounter (visit)
V12	CURE_E_DATE	9	Date of ending the clinical encounter (visit)
V13	IN_DATE	9	Date of hospitalization
V14	OUT_DATE	9	Date of discharge
V15	APPL_S_DATE	9	Date of beginning of the claims application
V16	APPL_E_DATE	9	Date of ending of the claims application
V17	ORDER_SEQ_NO	10	Sequential number of prescription of the test or examination
V18	ORDER_CODE	15	Specific code given by NHI for each test or examination
V19	ASSAY_UPLOAD_DATE	9	Date of uploading the results or reports
V20	RECIPE_DATE	9	Date of prescribing the test or examination
V21	REAL_INSPECT_DATE	9	Date of performing (sampling) the test or examination
V22	INSPECT_MODE	400	Method of performing (sampling) the test or examination
V23	CASE_SEQ_NO	10	Sequence number of repeated tests in the same day
V24	ASSAY_ITEM_NAME	200	Name of the test or examination for different purposes
V25	ASSAY_METHOD	200	Method of the test such as qualitative or quantitative
V26	ASSAY_VALUE	60	Result of the test
V27	UNIT_DATA	100	Unit of the result
V28	CONSULT_VALUE	100	Reference value of the result
V29	WEB_RECV_SEQ	20	Sequential number of reception the uploaded data by NHI
V30	BRANCH_ID	1	NHI branch
V31	TXT_MARK	1	Note on the changes
V32	TXT_DATE	9	Date of changes
V33	DTL_PKNO	19	Sequential number set by NHI
V34	OHP_SEV_TYPE	1	Type of encounter (outpatient, inpatient, or pharmaceutical)
V35	FUNC_SEQ_NO	15	Sequential number of clinical encounter (visit)

Transparency of Algorithms

Researchers using the laboratory databases should develop algorithms to include or exclude some incompatible records and to determine whether a result is positive or negative (normal or abnormal). Different researchers may employ differing criteria when designing such algorithms, and would affect the comparability between studies. The NHIA thus recommends that researchers using the laboratory databases to release the source codes of their algorithms in an open platform to alleviate the hurdle of data incompatibility.

Linking Laboratory Data with Claims Data

To better interpret the results of laboratory tests, the laboratory database should be linked to NHI inpatient and outpatient claims data to provide the contextual information (diagnosis, medical encounters and demographic information of the patient). One of the strengths of the use of laboratory database is to valid the International Classification of Disease (ICD) coding for case definition and measurement of exposure variables, confounding variables, or outcome variables. The second strength is to

Table 2 Examples of Contents Recorded in the Five Variables in Laboratory Test Database

No.	ORDER_CODE	ASSAY_ITEM_NAME	ASSAY_METHOD	ASSAY_VALUE	UNIT_DATA	CONSULT_VALUE
1	14032C	HbsAg (screening)	Roche E-411	Pos. in Chinese	S/N	<2.0
2	14032C	HBsAg-EIA	Instrument name	Pos.	cut off index	≥1 (+)
3	14032C	HBsAg Quantitative	Sample in Chinese	Negative	S/CO	Negative<1.00
4	14032C	HBsAg (Quantitation)	Unknown	Positive (16.51)	IU/mL	<0.05
5	14032C	HBsAg	Blood	Reactive	IU/mL	<2.0~
6	14032C	HBsAg in Chinese	Name in Chinese	0.515	S/N	<1.0 COI (-)
7	14032C	HBsAg (S/CO)	None	0.443	COI	(-)<1.0
8	14032C	GOT (AST)	Blood	0.33	IU/mL	<2.0~
9	09025C	HBsAg	Blood	0.87	COI	Nonreactive<1.0
10	09025C	GOT (AST)	Photometry	13	U/L	13–39

Abbreviations: EIA, enzyme immunoassay; GOT (AST), glutamic-oxaloacetic transaminase or aspartate aminotransferase; HBsAg, hepatitis B virus surface antigen.

indicate the severity of disease. The third strength, through the use of various information on dates we could better monitoring the trajectory of some important biomarkers.

The researchers wish to use laboratory databases and the claims data in the Applied Health Research Data Integration Service from NHIA should first submit their research proposal to the Institute Review Boards (IRB) qualified by the Taiwan Ministry of Health and Welfare. After getting the approval from IRB, the researchers can apply the purchasing of the variables their research needed. The fee for each variable per year is 250 NT dollars (the current exchange rate between NT dollar and US dollar is 28:1). The utilization fee in the NHI Analysis Center is 1600 NT dollars per day. Non-Taiwanese researchers have to collaborate with Taiwanese researchers to access the NHI laboratory results and the claims databases.

Funding

Disclosure

The authors report no conflicts of interest in this work.

References

1. Chiang TL. Taiwan's 1995 health care reform. *Health Policy*. 1997;39 (3):225–239. doi:10.1016/S0168-8510(96)00877-9
2. Cheng TM. Taiwan's new national health insurance program: genesis and experience so far. *Health Aff*. 2003;22(3):61–76. doi:10.1377/hlthaff.22.3.61
3. Cheng TM. Reflections on the 20th anniversary of Taiwan's single-payer National Health Insurance System. *Health Aff*. 2015;34:502–510. doi:10.1377/hlthaff.2014.1332
4. Hsing AW, Ioannidis JPA. Nationwide population science: lessons from the Taiwan National Health Insurance Research Database. *JAMA Intern Med*. 2015;175(9):1527–1529. doi:10.1001/jamainternmed.2015.3540
5. Lin LY, Warren-Gash C, Smeeth L, Chen PC. Data resource profile: the National Health Insurance Research Database (NHIRD). *Epidemiol Health*. 2018;40:e2018062. doi:10.4178/epih.e2018062
6. Hsieh CY, Su CC, Shao SC, et al. Taiwan's national health insurance research database: past and future. *Clin Epidemiol*. 2019;11:349–358. doi:10.2147/CLEP.S196293
7. Taiwan NHI papers search. Accessed at January 10, 2021. https://visualizinghealthdata.idv.tw/?route=article/thesis.
8. Lee TT, Cheng SH, Chen CC, Lai MS. A pay-for-performance program for diabetes care in Taiwan: a preliminary assessment. *Am J Manag Care*. 2010;16(1):65.
9. Chen TT, Chung KP, Lin IC, Lai MS. The unintended consequence of diabetes mellitus pay-for-performance (P4P) program in Taiwan: are patients with more comorbidities or more severe conditions likely to be excluded from the P4P program? *Health Serv Res*. 2011;46 (1p1):47–60. doi:10.1111/j.1475-6773.2010.01182.x
10. Hsieh HM, Gu SM, Shin SJ, Kao HY, Lin YC, Chiu HC. Cost-effectiveness of a diabetes pay-for-performance program in diabetes patients with multiple chronic conditions. *PLoS One*. 2015;10: e0133163. doi:10.1371/journal.pone.0133163
11. Huang SK, Wang PJ, Tseng WF, et al. NHI-PharmaCloud in Taiwan —a preliminary evaluation using the RE-AIM framework and lessons learned. *Int J Med Inform*. 2015;84(10):817–825. doi:10.1016/j.ijmedinf.2015.06.001
12. Tseng YT, Chang EH, Kuo LN, et al. Preliminary physician and pharmacist survey of the National Health Insurance PharmaCloud system in Taiwan. *Comput Methods Programs Biomed*. 2017;149:69–77. doi:10.1016/j.cmpb.2017.07.004
13. Liu MC, Lee CC. (2018). An investigation of pharmacists' acceptance of NHI-PharmaCloud in Taiwan. *J Med Syst*. 2018;42(11):213. doi:10.1007/s10916-018-1017-3
14. Liao CY, Wu MF, Poon SK, et al. Improving medication safety by cloud technology: progression and value-added applications in Taiwan. *Int J Med Inform*. 2019;126:65–71. doi:10.1016/j.ijmedinf.2019.03.012
15. Chiang H, Chang C. Introduction to and application analysis of Taiwan's NHI-MediCloud System. *J Serv Sci Res*. 2019;11:93–115. doi:10.1007/s12927-019-0005-6
16. Rau HH, Wu YS, Chu CM, et al. Importance-performance analysis of personal health records in Taiwan: a web-based survey. *J Med Internet Res*. 2017;19(4):e131. doi:10.2196/jmir.7065
17. Hsu MH. Chapter 11: health Information Technology. In: Chiang TL, SH C, editors. *Health Care Policy in East Asia: A World Scientific Reference Volume 4: Health Care System Reform and Policy Research in Taiwan*. Singapore: World Scientific Publishing Co. 2020:197–205.

18. Taiwan National Health Insurance Administration. Guidelines for uploading the results of laboratory tests and reports of examinations. Accessed at https://www.nhi.gov.tw/Content_List.aspx?n=264416706E2EF4DA&topn=D39E2B72B0BDFA15. Aug 30, 2019.

19. Pretorius C. Utilisation of pathology procedures in the South African private pathology sector between 2003 and 2005. *S Afr Med J*. 2007;97:51–57.

20. Makuwaza L, Musarurwa C, Goma Z. Evaluation of cost per test of clinical biochemistry tests at Parirenyatwa Central Hospital Laboratory, Harare, Zimbabwe. *Cen Afr J Med*. 2009;55:59–63.

21. Grann AF, Erichsen R, Nielsen AG, Froslev T, Thomsen RW. Existing data sources for clinical epidemiology: the clinical laboratory information system (LABKA) research database at Aarhus University, Denmark. *Clin Epidemiol*. 2011;3:133–138. doi:10.2147/CLEP.S17901

22. Horton S, Sullivan R, Flaningan J, et al. The top 25 laboratory tests by volume and revenue in five different countries. *Lancet*. 2018;391:1953–1964. doi:10.1016/S0140-6736(18)30460-4

23. Horton S, Fleming KA, Kuti MA, Looi LM, Pai SA, Lawler M. Delivering modern, high quality, affordable pathology and laboratory medicine to low-income and middle-income countries: a call to action. *Am J Clin Pathol*. 2019;151:446–451. doi:10.1093/ajcp/aqy165

24. Arendt JFH, Hansen AT, Ladefoged SA, Sørensen HT, Pedersen L, Adelborg K. Existing data sources in clinical epidemiology: laboratory information system databases in Denmark. *Clin Epidemiol*. 2020;12:469–475. doi:10.2147/CLEP.S245060

25. Sheu MJ, Ling FW, Li ST, Li CY, Lu TH. Validity of ICD-10-CM codes used to identify patients with chronic hepatitis B and C virus infection in administrative claims data from the Taiwan National Health Insurance outpatient claims dataset. *Clin Epidemiol*. 2020;12:185–192. doi:10.2147/CLEP.S236823