BMJ Health & Care Informatics

Development of automated HIV case reporting system using national electronic medical record in Thailand

Thitipong Yingyong,[1] Suchunya Aungkulanon ,[2] Wasun Saithong,[3] Supiya Jantaramanee,[1] Kanjanakorn Phokhasawad,[2] Ian Fellows,[4] Thananda Naiwatanakul,[2] Jariya Mobnarin,[3] Narong Charoen,[3] Paiboon Waikayee,[3] Sanny Chen Northbrook[2]

To cite: Yingyong T, Aungkulanon S, Saithong W, *et al*. Development of automated HIV case reporting system using national electronic medical record in Thailand. *BMJ Health Care Inform* 2022;**29**:e100601. doi:10.1136/bmjhci-2022-100601

► Additional supplemental material is published online only. To view, please visit the journal online (http://dx.doi.org/10.1136/bmjhci-2022-100601).

Received 22 April 2022
Accepted 15 August 2022

[1]Division of Epidemiology, Department of Disease Control, Ministry of Public Health, Nonthaburi, Thailand
[2]Division of Global HIV and TB, US Centers for Disease Control and Prevention, Nonthaburi, Thailand
[3]Information Communication Technology Center, Permanent Secretary Office, Ministry of Public Health, Nonthaburi, Thailand
[4]Division of Global HIV and TB, US Centers for Disease Control and Prevention, Atlanta, Georgia, USA

Correspondence to
Dr Suchunya Aungkulanon; ooz2@cdc.gov

ABSTRACT

Background An electronic medical record (EMR) has the potential to improve completeness and reporting of notifiable diseases. We developed and assessed the validity of an HIV case detection algorithm and deployed the final algorithm in a national automated HIV case reporting system in Thailand.

Methods The HIV case detection algorithms leveraged a combination of standard laboratory codes, prescriptions and International Classification of Diseases, 10th Revision diagnostic codes to identify potential cases. The initial algorithm was applied to the national EMR from 2014 to June 2020 to identify HIV-infected subjects to build the national HIV case reporting system (Epidemiological Intelligence Information System (EIIS)). A subset of potential positives identified by the initial algorithm were then validated and reviewed by infectious disease specialists. This review identified that a proportion of the false positives were due to pre-exposure prophylaxis/postexposure prophylaxis (PrEP/PEP) antiretrovirals, and so the algorithm was refined into a 'Final Algorithm' to address this.

Results Positive predictive value of identifying HIV cases was 90% overall for the initial algorithm. Individuals misclassified as HIV-positive were HIV-negative patients with incorrect diagnostic codes, prescription records for PrEP, PEP and hepatitis B treatment. Additional revision to the algorithm included triple drug regimen to avoid further misclassification. The final HIV case detection algorithm was applied to national EMR between 2014 and 2020 with 449 088 HIV-infected subjects identified from 1496 hospitals. EIIS was designed by applying the final algorithm to automated extract HIV cases from the national EMR, analysing them and then transmitting the results to the Ministry of Public Health.

Conclusions EMR data can complement traditional provider-based and laboratory-based disease reports. An automated algorithm incorporating laboratory, diagnosis codes and prescriptions have the potential to improve completeness and timeliness of HIV reporting, leading to the implementation of a national HIV case reporting system.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Traditional passive surveillance is burdensome to clinicians, and it is often incomplete and delayed as it may lack information needed for public health purposes.
⇒ Electronic medical record (EMR) is capable of accurately reporting and monitoring notifiable diseases.

WHAT THIS STUDY ADDS

⇒ This study developed HIV case detection algorithms using the national EMR database which validated using National AIDS Programme data and reviewed by infectious disease specialists.
⇒ Our novel algorithm to identify HIV infection in the EMR system contributed to the development of an automated HIV case reporting system in Thailand.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ HIV case detection algorithms have the potential to improve completeness and timeliness of HIV reporting system compared to the traditional system.
⇒ Much effort should be concentrated on improving data quality of national EMR data.

BACKGROUND

Electronic medical record (EMR) has the potential to improve completeness and reporting of notifiable diseases beyond traditional clinician-initiated and laboratory-based disease reporting systems.[1] Traditional passive surveillance is burdensome to clinicians, and it is often incomplete and delayed as it may lack information needed for public health purposes (eg, patient signs and symptoms, prescribed treatments and pregnancy status).[2 3] EMR, however, contains this information and stores it in a form that can be used for electronic analysis and reporting. Consequently, EMR-based reporting has the potential to provide active notifiable disease surveillance that is more timely, complete and clinically detailed that enables longitudinal disease reporting and analysis. With the advent and adoption of EMR, researchers are now able to rapidly identify potential disease

cases for clinical studies. Disease detection algorithms are needed to search across billing data, laboratory data and clinical documentation to perform case detection. These disease detection algorithms can be conceived in a manner that has high sensitivity and specificity for identifying individual's true disease status using methods borrowed from routine clinical care.[4]

The first case of HIV/AIDS in Thailand was officially reported in July 1984.[5] Shortly after, HIV/AIDS was declared a highly infectious disease requiring mandatory notification in 1985. In the past, physicians were required to report all patients with HIV infection, including asymptomatic cases in Report 506/1 to the provincial health office, who forwarded the information to the Ministry of Public Health (MoPH). This passive surveillance relies on physicians to report new cases of HIV infection or AIDS directly to the MoPH. Data from passive surveillance are often slow to accrue and incomplete and may not support a timely and well-aimed public health purpose. To address the problem MoPH stopped mandating asymptomatic HIV case reporting using Report 506/1 in 2014 and classified HIV as a notifiable condition and mandated healthcare providers and laboratories to report HIV cases to provincial health officials in the Communicable Disease Act BE 2558 (2015).[6]

HIV infection is a disease that lends itself to an algorithm-based case detection, given the reliance on laboratory-based testing. Previous studies have attempted to identify HIV-infected patients in selected population or selected hospitals in developed countries.[7–10] However, there is no validated procedure for using data from medical records to identify diagnosed HIV-infected patients for national HIV/AIDS surveillance purposes. In order to maximise use of EMR for automated HIV/AIDS reporting system, we developed and assessed the validity an HIV case detection algorithm and estimated the positive predictive value (PPV) of the algorithm to detect new diagnosed HIV cases and develop a national automated HIV case reporting system in Thailand.

EMR database

Since 2007, Thailand MoPH has established a national electronic health system to house the central and provincial health data centres (HDC) for management of EMR. All levels of health facilities under the MoPH are required to upload and transfer disaggregated, individualised data to a central database using cloud technology at least once a month. In 2021, the MoPH HDC platform received EMR data (ie, patient demographics, vital signs, test orders, test results, prescriptions, diagnostic codes and healthcare provider details) from 947 MoPH public hospitals, 55 non-MoPH public hospitals and 9760 subdistrict health promotion hospitals. Thai healthcare is dominated by public health facilities accounting for 79% of hospital beds[11] and with 21% of total beds in private hospitals. Specialised HIV treatment and care services are mainly provided in public health facilities under the management of universal coverage (UC) to ensure equitable access.

The standard patient-level data collected at each health facility includes demographics and health services.[12] Data are subsequently managed and summarised for key performance indicators. The visualised report is accessible for health management and disease control on the web-based HDC dashboard (http://hdcservice.moph.go.th)

METHODS

Initial algorithm development

HIV infection was identified by applying the surveillance case definition of HIV/AIDS from the 2015 national guidelines for reporting notifiable communicable disease in Thailand.[13] In order to maximise the utility and performance characteristics of the algorithm, six separate conditions using complementary approaches to HIV case detection were created. A panel of physicians, including experts in the diagnosis and treatment of HIV infection, provided recommendations for the development of these conditions. The HIV case detection algorithm combined national laboratory testing codes and results, Thai Medicine Terminology prescriptions[14] and International Classification of Diseases, 10th Revision (ICD-10) diagnostic codes. A list of laboratory tests, ICD-10 and antiretroviral (ARV) medication can be found in online supplemental appendix 1.

We identified six conditions under which a classification of HIV positive would be a reasonable conclusion and ordered them based on the panel's likelihood of false positivity. The initial algorithm classified an individual as positive if any one of the six conditions (A, B1, B2, C1, C2 and D1) were met. Condition A is met if laboratory test (anti-HIV, HIVDNA PCR or viral load) identified the individual as HIV infected. For HIV antibody testing, only a 'positive' result was defined as positive, whereas 'negative' and 'inconclusive' results were defined as negative. The presence of detectable viral load was defined as HIV-positive regardless of test result. Conditions B1, B2, C1 and C2 were based on clinical evidence and designed to detect those patients not identified by condition A due to incomplete HIV testing results in EMR. As a result of confidentiality concerns with sharing HIV laboratory results, the laboratory information system not linked to the hospital information system (HIS), and in other cases formatting issues when transferring data from hospital HIS to MoPH HDC prevent HIV laboratory results from being represented in the EMR. Condition B1 was considered met if an individual had three or more HIV-related ICD10 events as well as ARV drug use for at least three visits. Condition B2 was considered satisfied if there were one or two HIV-related ICD10 events and a CD4 test result of <200. C1 is a weaker condition that is satisfied with the presence of any HIV-related ICD10 event in the patient history. C2 and D1 are satisfied based solely on ARV use with C1 triggered when ARVs are reported on

Table 1 Summary of criteria included in the initial HIV case detection algorithm

Surveillance criteria	Condition	Laboratory	Diagnosis	Pharmacy
Laboratory-confirmed evidence	A	Presence of positive HIV antibody or HIV DNA PCR (in children <12 months) or result of detectable viral load		
Presumptive clinical evidence	B1		Presence of at least three HIV-related ICD-10 events	Report of antiretroviral drugs for at least three visits
	B2	Result of CD4 test result <200 cell count	Presence of one or two HIV-related ICD-10 event	
Probable clinical evidence	C1		Presence of at least one HIV-related ICD-10 event	
	C2			Report of antiretroviral drugs for at least two clinic visits, dated more than 60 days apart
Possible clinical evidence	D1			Report of antiretroviral drugs for two clinic visits, dated less than or equal to 60 days apart

ICD-10, International Classification of Diseases, 10th Revision.

two visits >60 days apart and D1 is triggered if they are ≤60 days apart. The initial algorithm distinguishes four levels of confidence in a determination of a positive, ranging from laboratory confirmed to possible clinical evidence. Our algorithm criteria are summarised in table 1.

Assessing validity

We assessed the validity the initial algorithm by applying it to the MoPH EMR from 31 hospitals that were interested in participating from 3 high HIV burden provinces in 2017. HIV positive patients from 31 hospitals were extracted from 460,575 HIV positives cases in the MoPH HDC platform to assess algorithm performance. To verify that subjects were HIV-infected, we used national AIDS programme (NAP) data for 'true HIV-positive' and cross-matched with individuals identified through the HIV case detection algorithm. The NAP is National Health Security Office's electronic database used for recording clinical and laboratory services for HIV monitoring and reimbursement.[15] Identification of an HIV-infected patient by the NAP occurs at the hospital, where local healthcare providers register all HIV patients into the NAP for HIV test, CD4, and viral load test reimbursement for all Thai citizens regardless of health insurance scheme. This helps ensure that patients identified by the NAP are truly infected with HIV. Charts of patients who were not identified by NAP were further reviewed by infectious disease specialists from that hospital. The PPV of identifying HIV/AIDS cases was used to measure accuracy of the algorithm. We also reviewed the distribution of subjects by number of diagnostic codes and confirmatory evidence of HIV infection using HIV viral load and ARV therapy results.

Final algorithm development

The expert reviews for the false positives found in the validation step were then analysed to determine whether any refinement to the algorithm conditions could be made. Frequencies were calculated for the expert determined reason for the misclassification and each reason was assessed to investigate whether there were additional constraints to the conditions that could be added to reduce the false positive likelihood. The resulting algorithm was labelled the final algorithm and was applied to the MoPH EMR from 2014 to 2020 to identify HIV-infected subjects for HIV case reporting system.

Automated HIV case reporting system development

The final algorithm was used to build the new HIV case reporting system called Epidemiological Intelligence Information System (EIIS). Data were integrated, stored, managed and analysed in the MoPH HDC cloud-based warehouse. EIIS Data warehouse security measures were strictly enforced to ensure data integrity at all levels. The system limited access to different level of authorised users. The data warehouse was set up to be read-only by default to prevent any threatening from being executed on the data. National identification number was encrypted applying the customised Hash Algorithm. Specifically, only the hashed version of national ID was stored in a database, which was decrypted to only authorised users.

RESULTS

Application of the initial HIV case detection algorithm

We applied the initial algorithm to the MoPH EMR data. All patients who were seen at hospitals under the MoPH HDC platform between 2014 and 2020 were included in this study. The HIV case detection algorithm identified a total of 4 60 575 cases among all patients receiving HIV services in 947 MoPH hospitals, 2243 subdistrict health promotion hospitals and 55 non-MoPH public hospitals. Twelve per cent of the cases were reported to have died. Table 2 shows unduplicated counts of patient who met the case detection criteria for each condition. Individuals

Table 2 Demographic characteristics of identified HIV cases by condition, 2014–2020

	A		B1		B2		C1		C2		D1	
	N	%	N	%	N	%	N	%	N	%	N	%
Total individuals	11 487		337 453		838		91 256		6543		12 998	
Sex												
Male	6572	57.2	186 215	55.2	574	68.5	54 780	60	3270	50	3658	28.1
Female	4915	42.8	151 238	44.8	264	31.5	36 475	40	3273	50	9340	71.9
Health insurance												
UC	7881	68.6	237 534	70.4	509	60.7	51 762	56.7	3048	46.6	6046	46.5
CSMBS	447	3.9	11 228	3.3	29	3.5	5133	5.6	1200	18.3	1922	14.8
SSS	2010	17.5	58 418	17.3	200	23.9	14 157	15.5	1445	22.1	2344	18
Migrant	271	2.4	10 975	3.3	22	2.6	6451	7.1	167	2.6	843	6.5
Self-paid	878	7.6	19 298	5.6	78	9.3	13 753	14.8	683	10.2	1843	13.9
Age												
0–15 years	136	1.2	3360	1	5	0.6	6044	6.6	231	3.5	4269	32.8
16–20 years	360	3.1	7702	2.3	22	2.6	4102	4.5	134	2.1	1221	9.4
21–25 years	765	6.7	18 340	5.4	99	11.8	8877	9.7	873	13.3	2617	20.1
26–30 years	889	7.7	23 776	7.1	101	12.1	9901	10.9	888	13.6	1663	12.8
31–35 years	1047	9.1	29 300	8.7	120	14.3	10 675	11.7	526	8	1000	7.7
36–40 years	1453	12.7	42 831	12.7	125	14.9	11 575	12.7	520	8	750	5.8
41–45 years	2013	17.5	61 388	18.2	131	15.6	10 995	12.1	556	8.5	537	4.1
46–50 years	1662	14.5	52 204	15.5	82	9.8	7598	8.3	483	7.4	283	2.2
>50 years	3162	27.5	98 552	29.2	153	18.3	21 489	23.6	2332	35.6	658	5.1
Nationality												
Thai	11 151	97.1	327 273	97	803	95.8	83 559	91.6	6450	98.6	12 687	97.6
Cambodia	14	0.1	916	0.3	1	0.1	879	1	12	0.2	37	0.3
Laos	21	0.2	1132	0.3	2	0.2	896	1	10	0.2	20	0.2
Myanmar	168	1.5	4858	1.4	15	1.8	3964	4.3	35	0.5	106	0.8
Others	133	1.2	3274	1	17	2	1958	2.2	36	0.6	148	1.1
Year of registration in EIIS												
2014	0	0	129 267	38.3	0	0	11 887	13	987	15.1	837	6.4
2015	0	0	70 436	20.9	0	0	13 267	14.5	979	15	1228	9.5
2016	0	0	41 094	12.2	0	0	10 798	11.8	1163	17.8	1586	12.2
2017	177	1.5	30 704	9.1	8	1	13 140	14.4	1262	19.3	2085	16
2018	2766	24.1	30 923	9.2	157	18.7	14 070	15.4	1045	16	2487	19.1
2019	4785	41.7	21 148	6.3	330	39.4	13 818	15.1	749	11.5	2736	21.1
2020	3759	32.7	13 881	4.1	343	40.9	14 276	15.6	358	5.5	2039	15.7

A1: Laboratory-confirmed evidence: HIV-positive test result or VL result; B1: Presumptive clinical evidence: at least 3 HIV-related ICD-10 events and report of ARV; B2: Presumptive clinical evidence: CD4 <200 and one or two HIV-related ICD-10 event; C1: Probable clinical evidence: at least one HIV-related ICD-10 event; C2: Probable clinical evidence: ARV use dated more than 60 days apart; D1: Possible clinical evidence: ARV use dated less than 60 days apart.
ARV, antiretroviral; CSMBS, Civil Servants' Medical Benefit Scheme; EIIS, Epidemiological Intelligence Information System; ICD-10, International Classification of Diseases, 10th Revision; SSS, social security scheme; UC, universal coverage.

were assigned to the highest evidence level for which they qualified. Seventy percent of cases were identified by both diagnosis and prescription (condition B1). Four percent of identified cases had strong evidence of HIV laboratory results (condition A) because HIV laboratory codes were included in the standard 43 files later on in 2017.

Accuracy of the initial algorithm

We identified 26 138 individual patients who met the HIV-infected case detection algorithms at 31 participating hospitals. After matching individual data with NAP, we found 18 647 records (71%) registered in NAP and an additional 7491 patients not identified by NAP. Patient charts of 7491 (100%) were reviewed by infectious disease specialists from their respective hospitals. Of these 7491, 4924 (66%) were correctly classified as HIV positive by our case-detection algorithms, 2120 (28%) were misclassified as HIV-positive (false positive) and 447 (6%) were not found in the hospital database. Positive cases who were not registered in NAP were mainly non-Thai citizens, covered under non-UC health insurance schemes or self-paid patients. PPV of identifying HIV cases based on the algorithm was (18 647+4924)/26 138=90% overall and 98% (16741+3481)/20701 among individuals receiving services in the past year.

Generating a final algorithm

Of the 2120 false positives identified by the review, the experts noted incorrect ICD10 code as the reason in 71% of cases where a reason was given (see table 3). This indicates that condition C1 might reduce the false positivity rate considerably; however, this would not be desirable as it also accounts for a large proportion of the individuals the algorithm classifies as positive (see table 2). As such, we felt that the cost in terms of the introduction of false negatives would be too high.

Postexposure prophylaxis (PEP) and pre-exposure prophylaxis (PrEP) accounted for 28% of false positives where a reason was given and Hepatitis B for an additional 1%. Recognising that (1) some medications used to treat HIV infection are also used for hepatitis B; (2) monotherapy of ARV drugs is not considered an effective HIV medication unless prescribed in combination with an additional drug and (3) dual therapy with emtricitabine and tenofovir are prescribed for PrEP and PEP, we further refined the algorithm to include triple drug regimen for HIV medication to avoid misclassification.

Table 3 Reasons given for false positivity by specialist review

	Count	%
PEP	415	26
PrEP	27	2
Hepatitis B	20	1
Incorrect ICD-10	1152	71
No information	506	

ICD-10, International Classification of Diseases, 10th Revision; PEP, postexposure prophylaxis; PrEP, pre-exposure prophylaxis.

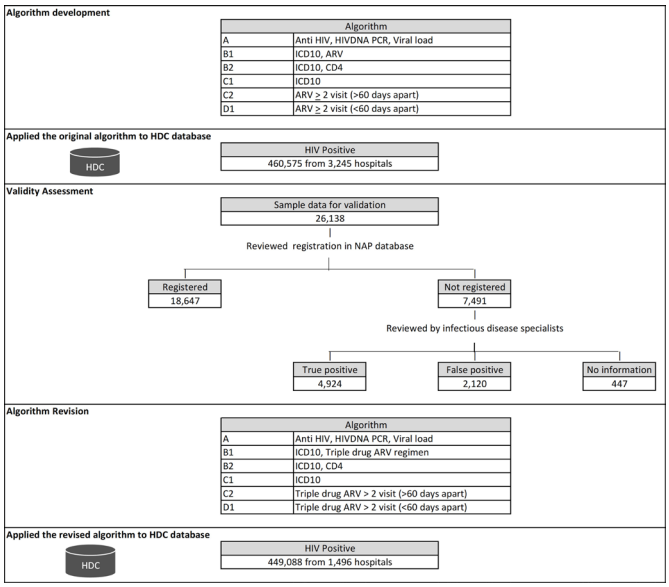

Figure 1 Development and validity assessment of the HIV case detection algorithms. ARV, antiretroviral; HDC, health data centres; ICD-10, International Classification of Diseases, 10th Revision; NAP, National AIDS Programme.

Figure 1 summarises development and validity assessment of the HIV case detection algorithms. The final algorithm replaces conditions B1, C2 and D1 by B1F, C2F and D1F, where the only difference is that the conditions for the final algorithm require triple drug ARV therapy to be triggered. Applying it to the MoPH EMR data found that the final algorithm classified 449 088 individuals as HIV positive, vs 460 575 for the initial algorithm. The similarity of volume of classified cases indicates that the restriction of the conditions to triple therapy was not overly restrictive.

The final algorithm experienced a large reduction in the number of hospitals, going from 3245 in the initial algorithm to 1496 in the final. Around 70% of hospitals are subdistrict health promoting hospitals providing primary care and health promotion services, but not ART. Triple drug regimen requirement reduced number of cases from these hospitals. Thus it is unsurprising that most cases in the EMR system should come from a more limited number of hospitals, whereas PrEP and PEP regimens would be present in a wider array of facilities.

DISCUSSION

To our knowledge, this is the first description of an algorithm using EMR data to identify patients with HIV positive status in Thailand. Our findings are consistent with the NAP. In June 2021, the final algorithm as implemented in the EIIS identified a total of 517 503 cases compared with 535 286 HIV infected individuals registered in NAP.[16] This provides confidence in the robustness of the algorithm to develop a case report system. Approximately 96% of cases were detected by the algorithm using clinical evidence and only 4% were detected by laboratory information.

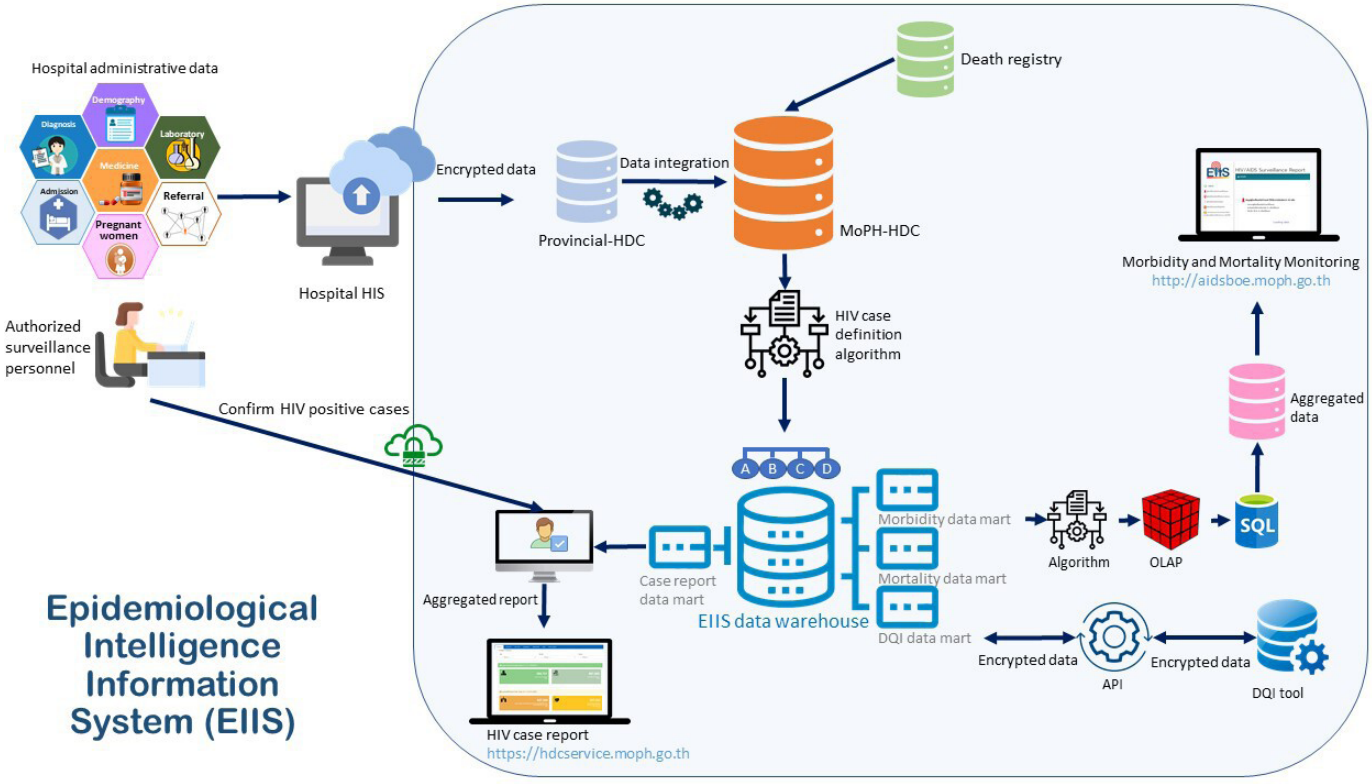

Figure 2 Diagram flow of the EIIS system. API, application programming interface; DQI, data quality improvement; HDC, health data centres; HIS, hospital information system; MoPH, Ministry of Public Health; OLAP, online analytical processing; SQL, structured query language.

The widespread adoption of EMRs for clinical documentation in Thailand has led to unprecedented opportunities for national communicable disease notification and monitoring system. In the past, reporting of notifiable diseases took a considerable amount of time and was performed by aggregating large amount of paper-based information from distant service delivery points up to the programme management level. This process was also prone to errors, loss of information, under-report and duplication when data from all service providers were combined.[17] The time taken in compiling information hindered the ability to respond to HIV situations in a timely manner. Previous studies have revealed considerable under-reporting in the national reporting system, limiting its ability to quantify prevalence or provide reliable estimates of future trends.[18 19]

Our novel algorithmic approach to the identification of HIV infection in the EMR system contributed to the development of an HIV case reporting system in Thailand. In October 2018, Thailand MoPH developed the EIIS to report HIV infection and monitor HIV/AIDS situation at national and subnational levels in Thailand. Figure 2 illustrates the diagram flow of the EIIS system. EMR included demographics, lab, prescription and diagnosis transfer from the HIS to provincial HDC and subsequently to MoPH HDC at least once a month. HDC also received death registry from the Ministry of Interior. EIIS applied the final algorithm to select patients who met the HIV case detection algorithm and transferred these cases to the EIIS data warehouse and case report DataMart for authorised surveillance staff at each health facility to review and confirm HIV infection status via a secured internet Web browser. According to the Disease Control Law and the MoPH regulation on data security, restricted access to the dataset was applied for only authorised HIV/AIDS disease control staff. Accessing the EIIS dataset required an online registration through the EIIS website and approval from the Department of Disease Control. In addition, EIIS data extraction was modified and further used for other purposes such as HIV morbidity and mortality monitoring system, and data quality improvement programme to track loss to follow-up patients.

Three years after implementation from 1 October 2018 to 30 June 2021, final algorithm as implemented in the EIIS detected and stored 518 684 cases in the EIIS data warehouse. Of those, 238 067 or 46% of cases were eventually reviewed by authorised healthcare workers at health facilities. Of those reviewed, 0.5% (1240 cases) were misclassified as HIV-infected when they were actually HIV-negative. EIIS can improve the reporting system through early notification and provision of detailed contact information; however, it may be limited in capturing information critical for risk factors and exposures. These data elements are variably documented by clinicians and typically only recorded as free-text rather than structured data. Natural-language processing techniques may improve capacity to report these risks in the future.

There are important limitations to what can be expected of EMR-based reporting systems. The ability of an HIV case detection algorithm relies on quality and completeness

of data. Ongoing efforts to improve EMR data quality are necessary. This system may not be able to capture patients receiving services from private hospitals. However, majority of HIV patients received HIV treatment in public facilities under UC where they submit EMR to MoPH HDC, accounting for 79% of all inpatient beds nationwide. The algorithms detect cases by searching for laboratory tests, prescriptions, and diagnosis codes that in combination are suggestive of HIV infection. The algorithms, therefore, need to be updated when new tests or new drugs are introduced, and new coding systems are implemented (eg, ICD-10). Periodic evaluation and system-wide update to continually calibrate algorithms will ensure efficacy of the HIV reporting system.[20] Only 4% of cases were detected by the laboratory-confirmed evidence and 96% of cases were detected by the algorithm using diagnosis and prescription criteria. The HIV case reporting may be delayed as cases who do not meet laboratory criteria will need to wait for HIV diagnosis and treatment data to determine their HIV status. The time and effort required to complete this process vary widely depending on the quality of records of reportable cases and the availability of clinical staff to review cases flagged by the EIIS system for accuracy. Additional resources and efforts are needed to strengthen the system.

CONCLUSIONS

This national automated case reporting system initiative developed in this study is a model for how EMR can automatically identify HIV-infected subjects. Algorithms incorporating laboratory, standard diagnosis codes and medication prescriptions have the potential to improve completeness and timeliness of HIV reporting system compared with the traditional system. Much effort should be concentrated on improving data quality of national EMR data.

Acknowledgements We would like to acknowledge Dr Tanarak Plipat, Dr Polawat Witoolkollachit, and Dr Anun Kanoksil, (Thailand MOPH) for policy and guidance support in the EIIS development, Dr Achara Teeraratkul (ex-CDC Thailand colleague), and Dr Piti Ongmongkolkul (Mahidol University) for providing advice on system design and applying big data analytics in algorithm development. We also would like to express our thanks to health care workers and public health staff of 54 public health hospitals and Provincial Health Offices in Chiang Mai, Chiang Rai and Nakhon Ratchasima provinces who participated to the development phase of EIIS study.

Contributors TY, TN and SCN were in charge of overall direction and planning; KP and SJ carried out data collection; SA, TN, IF, and SCN contributed to data analysis, interpretation of the data, and writing of menuscript; all authors read and approved the final manuscript; TY is guarantor.

Funding This project has been supported in part by the US President Emergency Plan for AIDS Relief (PEPFAR) through the Centers for Disease Control and Prevention (CDC) under the terms of 5 U2GGH001923.

Competing interests None declared.

Patient consent for publication Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data may be obtained from a third party and are not publicly available.

ORCID iD

Suchunya Aungkulanon http://orcid.org/0000-0003-4831-7528

REFERENCES

1 Lazarus R, Klompas M, Campion FX, *et al*. Electronic support for public health: validated case finding and reporting for notifiable diseases using electronic medical data. *J Am Med Inform Assoc* 2009;16:18–24.
2 Jajosky RA, Groseclose SL. Evaluation of reporting timeliness of public health surveillance systems for infectious diseases. *BMC Public Health* 2004;4:29.
3 Thaewnongiew K, Promthet S, Nilvarangkul K, *et al*. The surveillance system in health centers in northeastern Thailand. *Jpn J Infect Dis* 2009;62:444–9.
4 Afzal Z, Engelkes M, Verhamme KMC, *et al*. Automatic generation of case-detection algorithms to identify children with asthma from large electronic health record databases. *Pharmacoepidemiol Drug Saf* 2013;22:826–33.
5 Punpanich W, Ungchusak K, Detels R. Thailand's response to the HIV epidemic: yesterday, today, and tomorrow. *AIDS Educ Prev* 2004;16:119–36.
6 Thai Department of Disease Control. *Communicable disease act, B.E. 2558*. Ministry of Public Health, 2015.
7 Paul DW, Neely NB, Clement M, *et al*. Development and validation of an electronic medical record (EMR)-based computed phenotype of HIV-1 infection. *J Am Med Inform Assoc* 2018;25:150–7.
8 Fasciano NJ, Cherlow AL, Turner BJ. Profile of Medicare beneficiaries with AIDS: application of an AIDS casefinding algorithm. *Health Care Financ Rev* 1998;19:19–38.
9 Goetz MB, Hoang T, Kan VL, *et al*. Development and validation of an algorithm to identify patients newly diagnosed with HIV infection from electronic health records. *AIDS Res Hum Retroviruses* 2014;30:626–33.
10 Fultz SL, Skanderson M, Mole LA, *et al*. Development and verification of a "virtual" cohort using the National VA Health Information System. *Med Care* 2006;44:S25–30.
11 Healthcare resource statistics [Internet], 2021. Available: http://phdb.moph.go.th/
12 MoPH. *Standard operational guideline for data collection and transfereing of data standard format of MoPH, version 2.4*. Nonthaburi: Bureau of Policy and Strategy, MoPH, 2021.
13 Bureau of Epidemiology. Department of Disease Control. Ministry of Public Health. *Guidelines for reporing of notifiable communicable disease under the communicable disease control act B.E. 2558*. Nonthaburi, Thailand, 2015.
14 Thai Health Information Standard Development Center. Thai medicine terminology, 2021. Available: https://www.this.or.th/index.php
15 United Nations Development Programme. *The journey of universal access to antiretroviral treatment in Thailand*. Bangkok, Thailand, 2017.
16 National Health Security Office. NAP Web Report, 2021. Available: http://napdl.nhso.go.th/NAPWebReport/main_rep.jsp
17 Bayer R, Fairchild A. *The role of name-based notification in public health and HIV surveillance*. Geneva, Switzerland: UNAIDS, 2000.
18 Kanlayanaphotporn J, Brady MA, Chantate P, *et al*. Pneumonia surveillance in Thailand: current practice and future needs. *Southeast Asian J Trop Med Public Health* 2004;35:711–6.
19 Wichmann O, Yoon I-K, Vong S, *et al*. Dengue in Thailand and Cambodia: an assessment of the degree of underrecognized disease burden based on reported cases. *PLoS Negl Trop Dis* 2011;5:e996.
20 Gilliam D, Powell J, Haugh E, *et al*. Addressing software security and mitigations in the life cycle. *28th Annual NASA Goddard Software Engineering Workshop*; Proceedings; 3-4 Dec, 2003.

BMJ Health & Care Informatics

High quality, safe healthcare = technology + people + systems thinking

David Greenfield,[1] Usman Iqbal ,[2,3] Yu-Chuan (Jack) Li [4,5,6]

To cite: Greenfield D, Iqbal U, Li Y-CJ. High quality, safe healthcare = technology + people + systems thinking. *BMJ Health Care Inform* 2022;**29**:e100673. doi:10.1136/bmjhci-2022-100673

Received 02 September 2022
Accepted 07 September 2022

[1]School of Population Health, Faculty of Medicine and Health, University of New South Wales (UNSW), Sydney, New South Wales, Australia
[2]Global Health and Health Security Department, College of Public Health, Taipei Medical University, Taipei, Taiwan
[3]Health ICT, Department of Health, Hobart, Tasmania, Australia
[4]Graduate Institute of Biomedical Informatics, College of Medical Science & Technology, Taipei Medical University, Taipei, Taiwan
[5]Dermatology Department, Wan-Fang Hospital, Taipei, Taiwan
[6]International Association of Medical Informatics (IMIA), Geneva, Switzerland

Correspondence to
Professor Yu-Chuan (Jack) Li; jack@tmu.edu.tw

The BMJ Health & Care Informatics presented two editors' choice papers examining two different, but related papers, focused on health professional's perspectives on if and how technology can improve care processes and delivery. The empirical study of Bowden *et al*[1] explored clinicians' perceptions of digital access to patients' past medical history (PMH) as a basis for justifying significant investment into shared electronic health records (SEHR). Bates *et al*[2] work reported a roundtable expert discussion on the challenges and future direction in smart medication management.

Bowden *et al* surveyed clinicians from the front line, those in emergency departments and providing urgent care. In these time critical environments, clinicians reported that access to PMH is imperative to be able provide a response that accounts for heath status, current treatment regime and other health data related to the immediate presentation. Clinician's valued and wanted to obtain information from a trusted SEHR; there is a high level of technology acceptance. Five major suggested improvements were identified: increasing the number of patient records available; standardisation of information presentation; increased system reliability; expanded access to information and validation by authoritative/trusted sources. Two policy implications were identified: the need to focus on higher levels of patient participation; and, to ensure patient record curation and stewardship increasing the breadth and depth of information and processes.

Bates *et al* work records an expert discussion on the challenges and future direction in smart medication management. The key focus of the discussion was to reconsider the critical question: how can the original goal of improved healthcare quality and medication safety through electronic medical records be achieved? The challenges identified relate to established individual behaviours and beliefs, defined care delivery systems, and inflexible service requirements. They suggest that improvements are to be found through addressing simultaneously four interrelated issues: digital and information technology systems; safe prescribing; communication and education of both clinicians and patients; and medication adherence.

Individually, and together, Bowden *et al* and Bates *et al* highlight the need for a whole of systems approach that encompasses all healthcare providers to develop, implement, evaluate and improve technology to enhance care processes and delivery. They bring to attention, once again, that end-user involvement, including the pressing need for increased patient involvement, will likely raise the uptake and success of technology driven improvements.[3 4] Each work promotes renewed recognition that addressing usability and human factors are critical to building safe and effective health systems and care delivery processes.[5–7]

Twitter Yu-Chuan (Jack) Li @jaak88

Contributors Initial conception design and drafting of the manuscript: DG, UI and Y-CL. Drafting the manuscript: DG and UI. Critical revision of the paper: DG, UI and Y-CL.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient consent for publication Not applicable.

Provenance and peer review Commissioned; internally peer reviewed.

ORCID iDs
Usman Iqbal http://orcid.org/0000-0002-0614-123X
Yu-Chuan (Jack) Li http://orcid.org/0000-0001-6497-4232

REFERENCES

1 Bowden TC, Lyell D, Coiera E. Emergency department and urgent care clinician perspectives on digital access to past medical histories. *BMJ Health Care Inform* 2022;29:e100567.
2 Bates DW, Cheng H-Y, Cheung NT, *et al*. 'Improving smart medication management': an online expert discussion. *BMJ Health Care Inform* 2022;29:e100540.
3 Jacobsohn GC, Leaf M, Liao F, *et al*. Collaborative design and implementation of a clinical decision support system for automated fall-risk identification and referrals in emergency departments. *Healthc* 2022;10:100598.
4 Poncette A-S, Mosch LK, Stablo L, *et al*. A remote patient-monitoring system for intensive care medicine: mixed methods Human-Centered design and usability evaluation. *JMIR Hum Factors* 2022;9:e30655.
5 Carayon P, Hoonakker P. Human factors and usability for health information technology: old and new challenges. *Yearb Med Inform* 2019;28:071–7.
6 Di Pelino S, Lamarche L, Carr T, *et al*. Lessons learned through two phases of developing and implementing a technology supporting integrated care: case study. *JMIR Form Res* 2022;6:e34899.
7 Turner P, Kushniruk A, Nohr C. Are we there yet? Human factors knowledge and health information technology - the challenges of implementation and impact. *Yearb Med Inform* 2017;26:84–91.

BMJ Health & Care Informatics

Early detection of autism spectrum disorder in young children with machine learning using medical claims data

Yu-Hsin Chen ,[1] Qiushi Chen ,[1] Lan Kong ,[2] Guodong Liu [2,3,4,5]

To cite: Chen Y-H, Chen Q, Kong L, *et al*. Early detection of autism spectrum disorder in young children with machine learning using medical claims data. *BMJ Health Care Inform* 2022;**29**:e100544. doi:10.1136/bmjhci-2022-100544

► Additional supplemental material is published online only. To view, please visit the journal online (http://dx.doi.org/10.1136/bmjhci-2022-100544).

Received 06 January 2022
Accepted 19 August 2022

For numbered affiliations see end of article.

Correspondence to
Dr Qiushi Chen;
q.chen@psu.edu

ABSTRACT

Objectives Early diagnosis and intervention are keys for improving long-term outcomes of children with autism spectrum disorder (ASD). However, existing screening tools have shown insufficient accuracy. Our objective is to predict the risk of ASD in young children between 18 months and 30 months based on their medical histories using real-world health claims data.

Methods Using the MarketScan Health Claims Database 2005–2016, we identified 12 743 children with ASD and a random sample of 25 833 children without ASD as our study cohort. We developed logistic regression (LR) with least absolute shrinkage and selection operator and random forest (RF) models for predicting ASD diagnosis at ages of 18–30 months, using demographics, medical diagnoses and healthcare service procedures extracted from individual's medical claims during early years postbirth as predictor variables.

Results For predicting ASD diagnosis at age of 24 months, the LR and RF models achieved the area under the receiver operating characteristic curve (AUROC) of 0.758 and 0.775, respectively. Prediction accuracy further increased with age. With predictor variables separated by outpatient and inpatient visits, the RF model for prediction at age of 24 months achieved an AUROC of 0.834, with 96.4% specificity and 20.5% positive predictive value at 40% sensitivity, representing a promising improvement over the existing screening tool in practice.

Conclusions Our study demonstrates the feasibility of using machine learning models and health claims data to identify children with ASD at a very young age. It is deemed a promising approach for monitoring ASD risk in the general children population and early detection of high-risk children for targeted screening.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Growing evidence has shown that existing autism spectrum disorder (ASD)-specific screening tools (eg, Modified Checklist for Autism in Toddlers) may not yield sufficient accuracy for early detection of children with ASD in clinical practice.

⇒ Previous clinical and health service research has identified clinical risk factors associated with ASD, but the clinical factors from an individual's prior medical history have not been used comprehensively to assess the risk of ASD in young children.

WHAT THIS STUDY ADDS

⇒ This study demonstrated the feasibility of predicting ASD diagnosis with promising accuracy based on an individual's medical record from health claims data using machine learning models.

⇒ Our prediction models were clinically interpretable, which systematically identified key predictors in line with known risk factors and symptoms among ASD children in the literature.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ This study may serve as a basis for integrating predictive modelling into the health information system and the clinical workflow to enhance the current ASD screening practice.

INTRODUCTION

Autism spectrum disorder (ASD) is a developmental disorder that involves persistent challenges in social interaction, speech and nonverbal communication, and restricted and repetitive behaviours.[1] In the USA, the prevalence of ASD has increased substantially in the past two decades, with an estimate of every 1 in 44 children to be identified with ASD by age 8 in 2016.[2] Although there exist evidence-based interventions which improve core symptoms in children with ASD, many children with ASD still experience long-term challenges with daily life, education and employment.[3]

Early diagnosis is the key to early intervention for improving the long-term outcomes of children with ASD. However, despite the growing evidence shows that accurate and stable diagnoses can be made by 2 years,[4] in real-world settings, the median age of ASD diagnosis is 50 months.[2] To improve early diagnosis, the American Academy of Pediatrics (AAP) has recommended universal screening among all children at 18-month and 24-month well-child visits in the primary care settings using the Modified Checklist for Autism in Toddlers (M-CHAT),[5] a questionnaire that assesses children's behaviour for toddlers.[6] However, growing evidence has shown that using M-CHAT alone may

not yield sufficient accuracy in detecting ASD cases, with a sensitivity below 40% and a positive predictive value (PPV) under 20%.[7 8]

In addition to ASD-specific behavioural questionnaires, general clinical and healthcare records may also contain meaningful signals to differentiate the ASD risks among very young children. Studies have found that children with ASD are oftentimes accompanied by certain symptoms and medical issues such as gastrointestinal problems,[9] infections[10 11] and feeding problems.[12] This implies that past diagnosis and healthcare encounter information, commonly available from health insurance claims or Electronic Healthcare Record (EHR), could potentially be used for ASD risk prediction. In fact, medical claims and EHR data have been widely used in the health informatics literature for identifying disease-specific early phenotypes even before the hallmark symptoms start to manifest, such as for chronic diseases like heart failures,[13] diabetes[14] and Alzheimer's disease.[15] In the context of ASD, health record data has been used to identify the ASD subtypes[16 17] and to predict the suicidal risk in adolescents with ASD[18]; however, its use for predicting ASD diagnosis in young children has remained limited. To fill this gap, the objective of this study is to examine the feasibility of using large-scale real-world medical claims data to develop a prediction model for ASD diagnosis in young children, which can be used to support effective ASD screening strategies and facilitate early detection.

METHODS

Data source

We used the deidentified individual-level longitudinal healthcare claims data from the IBM MarketScan Commercial Claims and Encounters Database from 2005 to 2016. This database includes over 273 million unique individuals for both privately and publicly insured people in the USA.[19] The claims data include baseline demographics (eg, sex, birth year, postal region), service providers, insurance plans, medical diagnoses (in international Classification of Diseases (ICD)-9/10 codes) and procedures (in Healthcare Common Procedure Coding System (HCPCS) and Current Procedural Terminology-4 codes) at each encounter of healthcare services.

Study population

We constructed an initial cohort consisting of young children with and without ASD (figure 1). The inclusion criteria of the ASD cohort are as follows: (1) having at least 2 outpatient or 1 inpatient ASD diagnosis encounters (299 for ICD-9 and F84 for ICD-10) throughout the existing records[20 21]; and (2) having continuous enrolment from 4 months to 30 months to ensure the completeness of health records from the claims data that can be used for diagnosis prediction at up to 30 months (online supplemental figure S1). To create the non-ASD cohort, we first identified individuals without any ASD diagnosis throughout their health records, then downsampled 5% of the population to obtain a computationally manageable yet sufficiently large subset of samples. To ensure patients had adequate follow-up time to receive confirmed ASD diagnosis in the database, we restricted our selection of non-ASD patients by requiring a full enrolment period from 4 months to 60 months (online supplemental table S1).

Predictor variables for ASD diagnosis

We examined all diagnosis and procedure codes of a child's medical encounters available from as early as within 4 months after birth up to the age for prediction of ASD. We applied the Clinical Classifications Software (CCS),[22] a commonly used tool in health informatics research, to aggregate the large number of distinct diagnosis and procedure codes into clinically meaningful groups (figure 1). The single-level CCS maps the ICD-9/10 and HCPCS codes to a substantially smaller yet

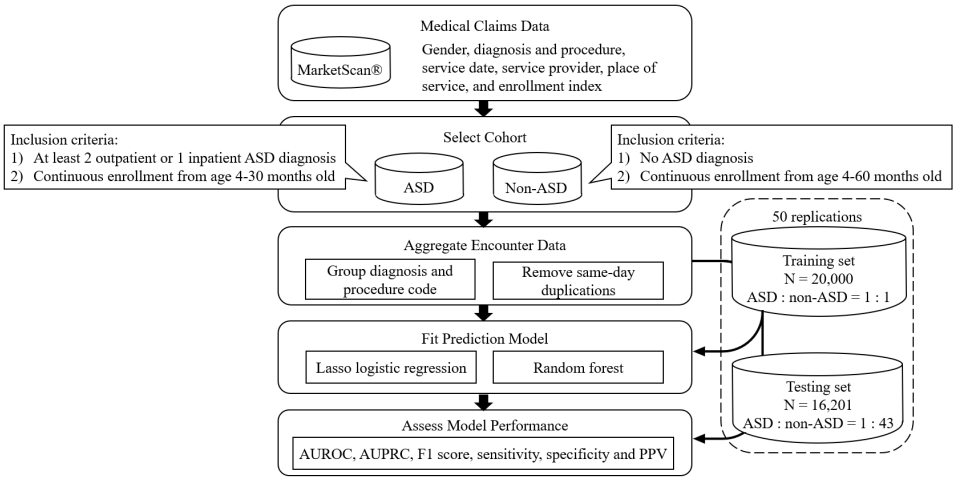

Figure 1 Overview of study design for the predictive analysis. ASD, autism spectrum disorder; AUROC, area under receiver operating characteristic curve; AUPRC, area under precision-recall curve; LASSO, least absolute shrinkage and selection operator; PPV, positive predictive value.

practical set that includes 285 diagnosis and 231 procedure categories.[22] We further removed the same-day duplications of CCS codes after the mapping by counting at most one encounter of a specific CCS category for each person on each day.

To predict the ASD diagnosis at the age of 24 months in our base case model, in line with the age when a diagnosis can possibly be made by an experienced professional,[4] we defined the predictor variables as the total number of encounters for each CCS category up to the age for prediction of 24 months. We also included sex and the encounters of emergency department visits, which are well-known clinically relevant factors associated with the autism population.[23] Variables that were present in <1% of both ASD and non-ASD cohorts were excluded.[24] A total of 170 input predictor variables were included for prediction at the age of 24 months. Having considered that the course of clinical events may be following a different pattern after an encounter with ASD diagnosis, we excluded any children who had at least one encounter with ASD diagnosis code prior to the age for prediction in our analysis.

Prediction model development and validation

We employed two machine learning methods, logistic regression (LR) and random forest (RF), which have been widely used for developing risk prediction models in various clinical settings. LR assumes that the independent variables are linearly related to the log odds and that the effects of multiple variables are additive, whereas RF is particularly suitable for exploiting nonlinear interactive effects in high-dimensional data. For the LR model, we also applied the least absolute shrinkage and selection operator (LASSO) as a feature selection technique to enforce the coefficients of weak predictors to be zero. The RF model was limited to up to 100 decision trees in the base case setting (other choices of the maximum number of trees were tested in sensitivity analysis).

To train our model, we sampled 10 000 ASD and 10 000 non-ASD subjects (N=20 000) from the initial cohort to build a large balanced training sample for maximising the discriminatory power learnt by the prediction model. To evaluate the model prediction performance, we created an independent imbalanced testing set (N=16 201) comprised of ASD and non-ASD patients from the remaining cohort that were mutually exclusive from the training set. The testing set resembled the real-world estimates for ASD prevalence of 2.3% (ie, 1 in every 44) in the general population.[2]

We measured the prediction performance with sensitivity (also known as true positive rate or recall), specificity (or true negative rate) and PPV (or precision)[25] at various selected risk thresholds. The model's overall discrimination ability was measured using the area under the receiver operating characteristic curve (AUROC). We also calculated the area under the precision-recall curve (AUPRC) where the precision-recall curve represents the relationship between PPV and sensitivity, and F1 score is defined as the harmonic mean of PPV and sensitivity, which are suited for evaluating the prediction performance for the imbalanced testing sample.[26 27] To assess the stability and the uncertainty of prediction performance, we repeated the training and testing set sampling, model training, testing and performance evaluation with 50 independent replications. The 95% CIs of all performance measures were reported.

Predicting ASD diagnosis at different ages

In addition to the base case prediction model where the risk of ASD diagnosis was assessed based on clinical information up to 24 months, we compared the accuracy of ASD prediction with varying lengths of available medical history at (1) a younger age, 18 months, considering that the universal ASD screening is recommended for children at both 18 months and 24 months[5]; and (2) an older age, 30 months, which is still a critical time point for monitoring the developmental delays and consideration of early intervention.[28] We followed the same approach in the base case to exclude predictor variables of low frequency (resulting in 150 and 180 predictor variables in total for prediction at 18 and 30 months, respectively) and the children with ASD diagnosis prior to the age for prediction.

Identifying key predictor variables

We further explored how many and which key predictive variables had the most impact on the prediction performance using the Gini importance index from the RF model. We added variables incrementally following the order of Gini Index (ie, starting with the most important variable) and evaluated how the prediction accuracy changed as more variables were included. Selected key predictive variables were then compared with those identified by alternative strategies using (1) the absolute value of coefficients from the LASSO LR model and (2) the prevalence of each variable in the identified ASD cohort.

Separating inpatient and outpatient visits

Considering that the underlying severity of the symptoms could potentially differ by inpatient hospitalisations and outpatient visits,[29] we split the number of encounters for each diagnosis and procedure by inpatient and outpatient visit separately and augmented the prediction model with more detailed encounter variables. We compared the prediction performance of the models using the augmented variables with our base case models.

Sensitivity analysis

We performed sensitivity analysis on several modelling assumptions to assess the robustness of our prediction models. Specifically, we strengthened the inclusion criteria for non-ASD subjects by requiring one additional year of enrollment, that is, increased from 4–60 months to 4–72 months. Furthermore, we assessed the potential loss of information due to excluding variables with <1% prevalence, to verify that such a variable prescreening

procedure would not miss out on rare but crucial predictive information.

RESULTS

Predicting ASD diagnosis at age of 24 months

We identified the study cohort consisting of 12 743 ASD subjects and 25 833 non-ASD subjects (more details in online supplemental table S1). When predicting the ASD diagnosis at the age of 24 months in independent testing samples, the LR and RF models achieved the AUROC of 0.758 (95% CI 0.755 to 0.762) and 0.775 (95% CI 0.771 to 0.779), respectively (table 1, figure 2). Compared with the LR model, RF model also showed a higher AUPRC (LR 0.101 (95% CI 0.098 to 0.104); RF 0.143 (95% CI 0.138 to 0.148)) and F1 score (LR 0.193 (95% CI 0.188 to 0.197); RF: 0.246 (95% CI 0.240 to 0.251)). The limit of up to 100 trees in the RF model was deemed sufficient to achieve stable performance. Further increasing the model complexity did not translate to an improvement in prediction accuracy (online supplemental table S2).

Predicting ASD diagnosis at different ages

Comparing the prediction models at the ages of 18, 24 and 30 months, we found that the prediction performance increased substantially with the age. Specifically for the RF model, the AUROC increased from 0.717 (0.714–0.721) at age of 18 months to 0.832 (0.828–0.835) at 30 months (table 1). Similarly, the AUPRC increased from 0.067 (0.065–0.069) to 0.234 (0.227–0.240) (figure 3), and F1 score increased from 0.130 (0.125–0.134) to 0.326 (0.322–0.331) from age of 18–30 months. The LR model, although with a lower prediction accuracy compared with the RF model in general, also showed a consistently increasing prediction performance as the age increased.

Identifying key predictive variables

As the RF model included more variables following the importance order by the Gini index, it showed higher AUROC (online supplemental figure S2). For prediction at age of 24 and 30 months, 30–40 most important variables were sufficient to achieve stable prediction performance with AUROC, whereas for an earlier age of 18 months, the top 50 important variables contributed to most of the prediction performance, while including additional variables could continue to marginally improve the prediction performance. We closely examined the 50 most important variables of the RF model (ranked by Gini index) and the LR model (ranked by the median absolute value of the coefficient) for prediction at age of 24 months (online supplemental figure S3). The identified important variables included sex, developmental and nervous system disorders, psychological and psychiatric services, respiratory system infections and symptoms, gastrointestinal-related diagnosis, ear and eye infections, perinatal conditions, and ED visits, which have also been seen as separate risk factors associated with ASD cases in the clinical literature. The key predictors of the RF model

Table 1 Performance of LASSO logistic regression and random forest models in prediction of autism spectrum disorder

Settings and prediction model	AUROC (95% CI)	AUPRC (95% CI)	F1 (95% CI)	Sensitivity target, % (95% CI)	Specificity, % (95% CI)	PPV, % (95% CI)
At age of 24 months (base case)						
LASSO logistic regression	0.758 (0.755 to 0.762)	0.101 (0.098 to 0.104)	0.193 (0.188 to 0.197)	40*	90.0 (89.7 to 90.4)	8.7 (8.4 to 9.0)
				50	83.7 (83.2 to 84.2)	6.7 (6.5 to 6.9)
				70	66.1 (65.4 to 66.8)	4.6 (4.5 to 47)
Random forest	0.775 (0.771 to 0.779)	0.143 (0.138 to 0.148)	0.246 (0.240 to 0.251)	40	93.0 (92.6 to 93.5)	12.1 (11.4 to 12.8)
				50	87.3 (86.7 to 87.9)	8.4 (8.1 to 8.8)
				70	69.6 (68.7 to 70.4)	5.1 (4.9 to 5.2)
At age of 18 months (younger)						
LASSO logistic regression	0.720 (0.716 to 0.723)	0.066 (0.064 to 0.068)	0.128 (0.124 to 0.132)	50	78.4 (77.9 to 78.9)	5.1 (5.0 to 5.2)
Random forest	0.717 (0.714 to 0.721)	0.067 (0.065 to 0.069)	0.130 (0.125 to 0.134)	50	78.8 (78.3 to 79.4)	5.1 (5.0 to 5.2)
At age of 30 months (older)						
LASSO logistic regression	0.800 (0.797 to 0.803)	0.148 (0.143 to 0.153)	0.255 (0.249 to 0.261)	50	90.4 (90.0 to 90.8)	11 (10.6 to 11.5)
Random forest	0.832 (0.828 to 0.835)	0.234 (0.227 to 0.240)	0.326 (0.322 to 0.331)	50	95.6 (95.3 to 95.8)	21.0 (20.2 to 21.8)

*The sensitivity threshold of 40% was selected to be comparable with the estimated sensitivity of 33%–39% for the existing autism-specific screening tools from real-world clinical settings.[7 8]
AUPRC, area under precision-recall curve; AUROC, area under receiver operator characteristic curve; LASSO, least absolute shrinkage and selection operator; PPV, positive predictive value.

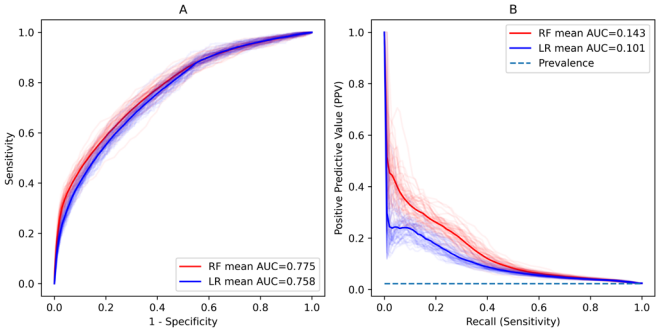

Figure 2 Receiver operating characteristic curves (A) and precision-recall (PR) curves (B) for prediction of autism spectrum disorder (ASD) diagnosis at age of 24 months. The prevalence stands for the baseline 2.27% (ie, 1 in 44) ASD prevalence in the general population. AUC, area under curve; LR, logistic regression; RF, random forest.

were also highly consistent with high prevalence variables, sharing 47 out of 50 most common variables in the ASD cohort (online supplemental figure S4).

Prediction using separated inpatient and outpatient data

Separating inpatient and outpatient encounters further increased the AUROC for prediction at the age of 24 months to 0.766 (95% CI 0.762 to 0.769) in the LR model and 0.834 (95% CI 0.831 to 0.837) in the RF model. At the target sensitivity of 40%, the RF model achieved a higher specificity of 96.4% (95% CI 96.2% to 96.5%) with a PPV of 20.5% (95% CI 19.8% to 21.1%), outperforming the existing screening tool M-CHAT/F (with a sensitivity of 38.8%, specificity of 94.9% and PPV of 14.6%). We found that using claims data separated by inpatient and outpatient visits improved the prediction performance consistently across all ages (figure 4).

Robustness check and sensitivity analysis

With a more stringent inclusion criterion for non-ASD subjects by requiring a longer full enrollment period up to 72 months (vs 60 months in our base case), we found that the prediction performance had modest improvement (online supplemental table S3). It could be partially attributed to the fact that with longer years to ascertain

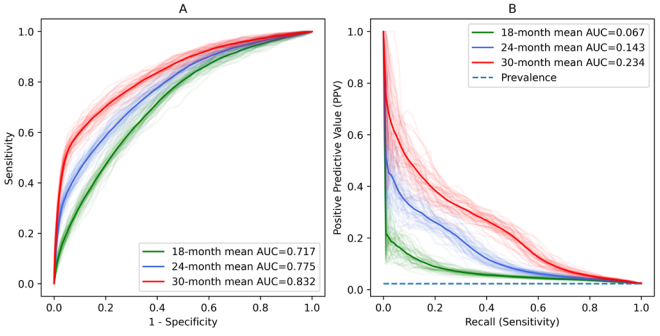

Figure 3 Receiver operating characteristic curves (A) and precision-recall curves (B) for prediction of autism spectrum disorder at ages of 18, 24 and 30 months, respectively, by the random forest model. AUC, area under curve.

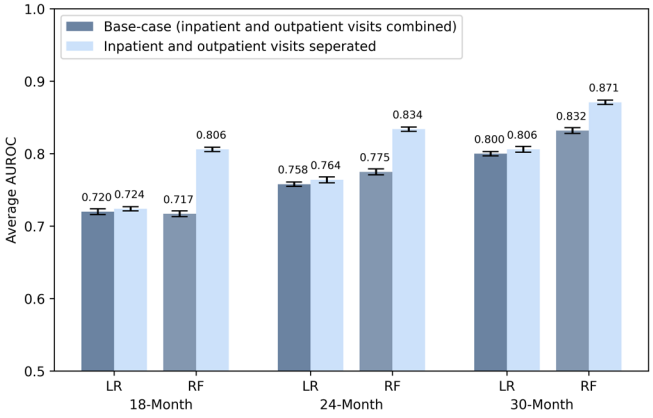

Figure 4 Comparison of area under the receiver operating characteristic curve (AUROC) with combined versus separated inpatient and outpatient encounters by LASSO logistic regression (LR) and random forest (RF) models, at the age of 18, 24 and 30 months, respectively. Error bars in the figure represent the 95% CIs based on results from 50 replications of independent runs. LASSO, least absolute shrinkage and selection operator.

the non-ASD cohort, children would be less likely to be misclassified. We also verified that including the low-prevalence variables would not result in substantial differences but only marginal changes of AUROC within 0.01 across all model specifications.

DISCUSSION

Early identification is vital for children with ASD to ensure their access to timely intervention and to optimise long-term outcomes. In this study, we demonstrated the feasibility of predicting ASD diagnosis at early ages using health claims data and machine learning models. We found that LASSO LR and RF models achieved an overall AUROC above 0.75 when predicting ASD diagnosis at age of 24 months. Our results also showed that prediction performance increased with age at the time of prediction. This is reasonable because more clinical information accumulated over a longer follow-up period since birth may contain more distinctive patterns to effectively differentiate children with ASD. The prediction models developed in our study are clinically interpretable. Key predictors, such as sex (male), developmental delays, gastrointestinal disorders, respiratory system infections and otitis media have shown strong predictive values for ASD diagnosis, which are in line with previous clinical studies that have shown these symptoms being associated with ASD children. Finally, our study showed that separating inpatient and outpatient claims as predictors could further improve the prediction accuracy.

In our study, both LASSO LR and RF models showed promising accuracy in predicting ASD diagnosis based on an individual's medical claims data. This robust finding implies that there may exist distinct patterns in health conditions and health service needs among young children with ASD, well before the onset of most hallmark ASD behavioural symptoms. Such predictive signals can

be easily extracted from the electronic health records or medical claims administrative data, and used for the early identification of ASD cases. We also observed differences in the performance between the two models. The RF model outperformed the LASSO LR model in general, likely because, with its tree-based model structure, the RF model is better at capturing complex interactive effects among the predictor variables to distinguish between the ASD and non-ASD cases, whereas the LR model synthesises the effects of multiple variables additively. The advantage of the RF model became more salient when input variables were separated by inpatient and outpatient claims into a more granular level.

Our study has made an important contribution to applying health informatics in the field of ASD. Although there exists a plethora of literature identifying individual risk factors of ASD, using large healthcare service data and machine learning models to systematically predict ASD diagnosis has remained much less explored. Unlike existing clinical informatics studies that focused on detecting ASD subtypes,[16 17] we aim to detect ASD cases among the general children population, that is, the early detection. This could be particularly challenging due to the low prevalence of ASD in the general population (ie, a highly imbalanced dataset), and the scarcity of information available at such a young age. Nevertheless, our model showed promising prediction performance. The RF model with separated inpatient and outpatient encounters achieved a specificity of 96.4% at a sensitivity of 40% for the ASD prediction at the age of 24 months, outperforming the accuracy of the existing ASD-specific screening tool (sensitivity: 38.8%; specificity: 94.9%) from a clinical observational study.[7] It is worth noting that under a similar ASD prevalence (2.2%), our model showed a higher PPV (20.5% vs 14.6%).

Our prediction model for ASD diagnosis could lead to a significant impact on the screening strategies for ASD in young children. Although the AAP guidelines recommend universal screening in all children, it has been debated that, without the perfect screening tool, universal screening may result in overburdened diagnostic services in the healthcare system as these clinical resources are in extremely short supply.[30] Our prediction models have demonstrated promising improvement over the existing ASD screening tool by using clinical information, which could potentially serve as a 'triaging tool' for identifying high-risk patients for diagnostic evaluation. Moreover, the models only based on health claims data makes it practically feasible to integrate into an EHR system or insurance claims database. It could further enable an automatic screening tool, which can continuously monitor an individual's risk as new diagnosis and procedure information emerges, and send reminders to patients or providers for a timely clinical assessment if necessary. On the other hand, it is possible that some diagnosis and procedure information appear after a concern that the child had autism has already existed, such as following a positive screening event, which could alter the course of subsequent clinical events. As such, our prediction model is not designed to direct the screening decisions, but rather a tool to enhance the screening accuracy. If more detailed electronic health record data were available, the proposed risk prediction model could be further extended by incorporating screening results with clinical information, or by differentiating the clinical information before versus after the screening events, to further improve the accuracy of identifying high-risk ASD cases for further diagnostic evaluation.

Our study has several limitations. First, diagnosis of ASD established only based on existing diagnosis codes from claims data could be inaccurate and unreliable sometimes in practice. We followed a validated approach in ASD health service research literature to identify the ASD cohort in our study.[31] Second, the absence of ASD diagnosis codes in one's health record may not necessarily indicate an individual not having ASD, especially for children born in later years, due to limited follow-up time prior to the cut-off date in the database. Thus, we required full enrollment up to 60 months without ASD diagnoses to identify the non-ASD cohort, and verified the robustness of our base case results in a sensitivity analysis requiring full enrolment up to 72 months. Third, as autistic children are likely to have a wide range of comorbid conditions with various frequencies, for individuals who do not present comorbid conditions from the past healthcare encounter data, our model may provide limited value. Our risk prediction model can be further augmented by additional information other than information from the health claims database, such as ASD/developmental screening results and behaviour-related information from a more comprehensive EHR dataset in future studies. Lastly, the diagnosis and procedure codes in insurance claims data may be subject to variabilities and irregularities. Instead of the original detailed clinical codes, we used aggregated CCS categories for diagnoses and procedures for more robust clinical measures.

CONCLUSIONS

Using real-world health claims data and machine learning methods, we developed a prediction model that can successfully predict ASD diagnosis for children under 30 months with promising prediction accuracy. Our model also identified the important predictors for the diagnosis prediction, which showed meaningful clinical relevance and intuition. Our predictive modelling approach could potentially be generalised to broader clinical settings for predicting the diseases that may show early signals from past healthcare service encounters in claims or EHR data. Future studies could explore the prediction of ASD diagnosis dynamically over time as new healthcare encounter occurs, and investigate how validated risk prediction models could be integrated and used to inform ASD screening strategies.

Author affiliations
[1]The Harold and Inge Marcus Department of Industrial and Manufacturing Engineering, The Pennsylvania State University, University Park, Pennsylvania, USA
[2]Department of Public Health Sciences, The Pennsylvania State University College of Medicine, Hershey, Pennsylvania, USA

[3]Department of Psychiatry and Behavioral Health, The Pennsylvania State University College of Medicine, Hershey, Pennsylvania, USA
[4]Department of Pediatrics, The Pennsylvania State University College of Medicine, Hershey, Pennsylvania, USA
[5]The Center for Applied Studies in Health Economics (CASHE), The Pennsylvania State University College of Medicine, Hershey, Pennsylvania, USA

Contributors GL, QC and Y-HC conceived of the presented idea, and developed it with support from GL and QC. Y-HC cleaned and preprocessed the data, developed prediction models, and performed model evaluations. All authors interpreted the model results. Y-HC and QC drafted the manuscript, which was critically revised by all authors. QC is the guarantor of the project.

Funding This work has been supported by Penn State Social Science Research Institute Level 1 Seed Grant (QC, GL), Penn State College of Engineering Multidisciplinary Research Seed Grant (Y-HC, QC, GL) and NIH R21 grant: 1 R21 MH119480-01A1 (GL, LK).

Competing interests None declared.

Patient consent for publication Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data may be obtained from a third party and are not publicly available.

ORCID iDs
Yu-Hsin Chen http://orcid.org/0000-0002-3678-7517
Qiushi Chen http://orcid.org/0000-0003-4031-2669
Lan Kong http://orcid.org/0000-0001-6098-9445
Guodong Liu http://orcid.org/0000-0001-8683-0803

REFERENCES

1 American Psychological Association. *Diagnostic and statistical manual of mental disorders (DSM-5®)*. American Psychiatric Pub, 2013.
2 Maenner MJ, Shaw KA, Bakian AV, *et al*. Prevalence and characteristics of autism spectrum disorder among children aged 8 Years - autism and developmental disabilities monitoring network, 11 Sites, United States, 2018. *MMWR Surveill Summ* 2021;70:1–16.
3 McPheeters ML, Weitlauf A, Vehorn A. *U.S. preventive services Task force evidence syntheses, formerly systematic evidence reviews. screening for autism spectrum disorder in young children: a systematic evidence review for the US preventive services Task force*. Rockville (MD): Agency for Healthcare Research and Quality (US), 2016.
4 Lord C, Risi S, DiLavore PS, *et al*. Autism from 2 to 9 years of age. *Arch Gen Psychiatry* 2006;63:694–701.
5 Lipkin PH, Macias MM, Council on children with disabilities, section on developmental and behavioral pediatrics. Promoting optimal development: identifying infants and young children with developmental disorders through developmental surveillance and screening. *Pediatrics* 2020;145. doi:10.1542/peds.2019-3449. [Epub ahead of print: 16 Dec 2019].
6 Robins DL, Fein D, Barton ML, *et al*. The modified checklist for autism in toddlers: an initial study investigating the early detection of autism and pervasive developmental disorders. *J Autism Dev Disord* 2001;31:131–44.
7 Guthrie W, Wallis K, Bennett A, *et al*. Accuracy of autism screening in a large pediatric network. *Pediatrics* 2019;144.
8 Carbone PS, Campbell K, Wilkes J, *et al*. Primary care autism screening and later autism diagnosis. *Pediatrics* 2020;146. doi:10.1542/peds.2019-2314. [Epub ahead of print: 06 07 2020].
9 Chaidez V, Hansen RL, Hertz-Picciotto I. Gastrointestinal problems in children with autism, developmental delays or typical development. *J Autism Dev Disord* 2014;44:1117–27.
10 Rosen NJ, Yoshida CK, Croen LA. Infection in the first 2 years of life and autism spectrum disorders. *Pediatrics* 2007;119:e61–9.
11 Adams DJ, Susi A, Erdie-Lalena CR, *et al*. Otitis media and related complications among children with autism spectrum disorders. *J Autism Dev Disord* 2016;46:1636–42.
12 Ledford JR, Gast DL. Feeding problems in children with autism spectrum disorders. *Focus Autism Other Dev Disabl* 2006;21:153–66.
13 Sideris C, Alshurafa N, Pourhomayoun M, *et al*. A data-driven feature extraction framework for predicting the severity of condition of congestive heart failure patients. *Annu Int Conf IEEE Eng Med Biol Soc* 2015;2015:2534–7.
14 Nguyen BP, Pham HN, Tran H, *et al*. Predicting the onset of type 2 diabetes using wide and deep learning with electronic health records. *Comput Methods Programs Biomed* 2019;182:105055.
15 Park JH, Cho HE, Kim JH, *et al*. Machine learning prediction of incidence of Alzheimer's disease using large-scale administrative health data. *NPJ Digit Med* 2020;3:46.
16 Lingren T, Chen P, Bochenek J, *et al*. Electronic health record based algorithm to identify patients with autism spectrum disorder. *PLoS One* 2016;11:e0159621.
17 Vargason T, Frye RE, McGuinness DL, *et al*. Clustering of co-occurring conditions in autism spectrum disorder during early childhood: a retrospective analysis of medical claims data. *Autism Res* 2019;12:1272–85.
18 Downs J, Velupillai S, George G, *et al*. Detection of suicidality in adolescents with autism spectrum disorders: developing a natural language processing approach for use in electronic health records. *AMIA Annu Symp Proc* 2017;2017:641–9.
19 IBM MarketScan research databases, 2020. Available: https://www.ibm.com/products/marketscan-research-databases
20 Burke JP, Jain A, Yang W, *et al*. Does a claims diagnosis of autism mean a true case? *Autism* 2014;18:321–30.
21 Coleman KJ, Lutsky MA, Yau V, *et al*. Validation of autism spectrum disorder diagnoses in large healthcare systems with electronic medical records. *J Autism Dev Disord* 2015;45:1989–96.
22 Agency for Healthcare Research and Quality R, MD. HCUP clinical classifications software (CCS) for ICD-9-CM Healthcare Cost and Utilization Project (HCUP) 2006-2009; 2020. www.hcup-us.ahrq.gov/toolssoftware/ccs/ccs.jsp
23 Loomes R, Hull L, Mandy WPL. What is the male-to-female ratio in autism spectrum disorder? A systematic review and meta-analysis. *J Am Acad Child Adolesc Psychiatry* 2017;56:466–74.
24 He D, Mathews SC, Kalloo AN, *et al*. Mining high-dimensional administrative claims data to predict early hospital readmissions. *J Am Med Inform Assoc* 2014;21:272–9.
25 Hunink MGM, Weinstein MC, Wittenberg E. *Decision making in health and medicine : Integrating evidence and values*. 2nd ed. Cambridge University Press, 2014.
26 Jeni LA, Cohn JF, De La Torre F. Facing imbalanced data recommendations for the use of performance metrics. *Int Conf Affect Comput Intell Interact Workshops* 2013;2013:245–51.
27 Ozenne B, Subtil F, Maucort-Boulch D. The precision--recall curve overcame the optimism of the receiver operating characteristic curve in rare diseases. *J Clin Epidemiol* 2015;68:855–9.
28 Hyman SL, Levy SE, Myers SM, *et al*. Identification, evaluation, and management of children with autism spectrum disorder. *Pediatrics* 2020;145:e20193447.
29 Pottick K, Hansell S, Gutterman E, *et al*. Factors associated with inpatient and outpatient treatment for children and adolescents with serious mental illness. *J Am Acad Child Adolesc Psychiatry* 1995;34:425–33.
30 Siu AL, Bibbins-Domingo K, *et al*, US Preventive Services Task Force (USPSTF). Screening for autism spectrum disorder in young children: US preventive services task force recommendation statement. *JAMA* 2016;315:691–6.
31 Liu G, Pearl AM, Kong L, *et al*. Risk factors for emergency department utilization among adolescents with autism spectrum disorder. *J Autism Dev Disord* 2019;49:4455–67.

BMJ Health & Care Informatics

Predicting surgical department occupancy and patient length of stay in a paediatric hospital setting using machine learning: a pilot study

Yuval Barak Corren ,[1] Joshua Merrill,[2] Ronald Wilkinson,[2] Courtney Cannon,[2] Jonathan Bickel,[2] Ben Y Reis[2,3]

To cite: Barak Corren Y, Merrill J, Wilkinson R, *et al*. Predicting surgical department occupancy and patient length of stay in a paediatric hospital setting using machine learning: a pilot study. *BMJ Health Care Inform* 2022;**29**:e100498. doi:10.1136/bmjhci-2021-100498

► Additional supplemental material is published online only. To view, please visit the journal online (http://dx.doi.org/10.1136/bmjhci-2021-100498).

Received 14 October 2021
Accepted 05 July 2022

[1]Computational Health Informatics Program, Boston Children's Hospital, Boston, Massachusetts, USA
[2]Enterprise Analytics, Boston Children's Hospital, Boston, Massachusetts, USA
[3]The Department of Biomedical Informatics, Harvard Medical School, Boston, Massachusetts, USA

Correspondence to
Dr Yuval Barak Corren; yuval.barakcorren@childrens.harvard.edu

ABSTRACT

Objective Early and accurate prediction of hospital surgical-unit occupancy is critical for improving scheduling, staffing and resource planning. Previous studies on occupancy prediction have focused primarily on adult healthcare settings, we sought to develop occupancy prediction models specifically tailored to the needs and characteristics of paediatric surgical settings.

Materials and methods We conducted a single-centre retrospective cohort study at a surgical unit in a tertiary-care paediatric hospital in Boston, Massachusetts, USA. We developed a hierarchical modelling framework for predicting next-day census using multiple types of data—from bottom-up patient-specific orders and procedures to top-down temporal variables and departmental admission statistics.

Results The model predicted upcoming admissions and discharges with a median error of 17%–21% (2–3 patients per day), and next-day census with a median error of 7% (n=3). The primary factors driving these predictions included day of week and scheduled surgeries, as well as procedure duration, procedure type and days since admission. We found that paediatric surgical procedure duration was highly predictive of postoperative length of stay.

Discussion Our hierarchical modelling framework provides an overview of the factors driving capacity issues in the paediatric surgical unit, highlighting the importance of both top-down temporal features (eg, day of week) as well as bottom-up electronic health records (EHR)derived features (eg, orders for patient) for predicting next-day census. In the practice, this framework can be implemented stepwise, from top to bottom, making it easier to adopt.

Conclusion Modelling frameworks combining top-down and bottom-up features can provide accurate predictions of next-day census in a paediatric surgical setting.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ A shortage of surgical beds and frequent fluctuations in surgical-unit occupancy present major challenges for hospitals, and often result in last-minute surgical cancellations. Forecasting bed occupancy can mitigate this problem and has been shown to be feasible in adult settings.

WHAT THIS STUDY ADDS

⇒ This study demonstrates that accurate occupancy prediction is also feasible in paediatric settings and highlights the main features driving these predictions. We present a hierarchical modelling framework, combining top-down departmental variables with bottom-up patient-level data.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ This framework can improve resource utilisation in paediatric surgical units as well as guide the development of other practical clinical prediction models.

INTRODUCTION

Substantial fluctuations in surgical-unit occupancy present a major and ongoing challenge for healthcare providers. Many tasks that depend on future bed availability, including staffing, scheduling and patient transfer management, are directly impacted by these uncertainties. Failure to accurately predict surgical department census and the resulting lack of bed availability are among the leading causes of planned surgery cancellations.[1 2] Last-minute cancellations can lead to great frustration for the patient, the patient's family and the clinical team.[3] Lastly, delays in surgery can also result in grave medical complications and increased morbidity and mortality.[4]

In recent years, a growing body of work has focused on clinical occupancy prediction in adult healthcare settings,[5–13] yet very few studies have focused on predicting capacity in paediatric hospitals or paediatric surgical departments.[14 15] Yet while inpatient beds for adult patients are projected to increase over time, the number of paediatric inpatient beds has been dropping,[16] increasing the pressure on inpatient resources devoted to paediatric services and highlighting the importance of focusing specifically on this subpopulation.

In this study, we develop an occupancy prediction model for a surgical unit in a tertiary-care paediatric hospital. We apply an integrated hierarchical modelling framework that incorporates both top-down population-level data and bottom-up individual-level data. In addition to this main modelling framework, we also develop a separate smaller model focused on postsurgical length of stay (LOS), in order to study the key risk factors associated with long postoperative stays.

METHODS

Setting

We analysed data from a surgical unit in a tertiary-care paediatric hospital in Boston, Massachusetts, USA. The data included all clinical and administrative orders, surgical procedures and procedure durations for patients admitted to the surgical unit from September 2015 to March 2020 (4.5 years). The data also included the daily census of the top three admitting departments, comprising 90% of all admissions to the surgical unit: emergency department (ED), postanaesthesia care unit (PACU) and surgical intensive care unit (ICU). Data from the first 3.5 years of the study period (1 September 2015–28 February 2019) were used for model training, while data from the last year (1 March 2019–29 February 2020) were used for model validation.

Census prediction

The daily census was calculated based on the number of patients in the surgical unit each day at 8:00 hours. To predict next-day census, the following formula was used:

$$\textit{Predicted census} = \textit{Current census} + \textit{Expected admissions} - \textit{Expected discharges}$$

An overview of the modelling framework is presented in figure 1. This figure may serve as a useful guide to the description of the multipart model in the following paragraphs. We applied a hierarchical modelling approach that fused both population-level temporal trends (top-down) together with individual-patient-specific factors (bottom-up).

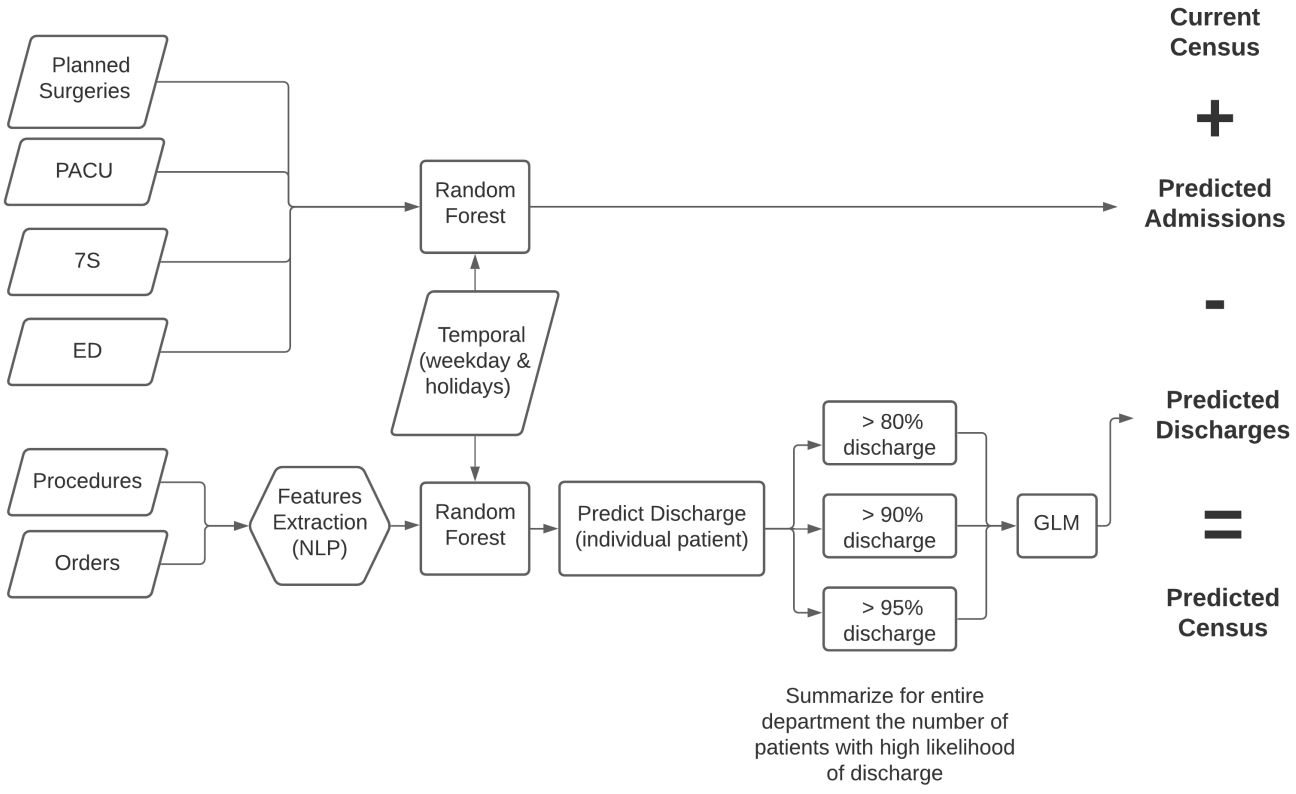

Figure 1 Overview of hierarchical modelling framework. Next-day surgical unit census is modelled as a function of the current census, the inflow of patients (ie, expected admissions) and the outflow of patients (ie, expected discharges). Patient inflow is predicted by a random forest model that uses the census data from the top three admitting departments (PACU, SICU, ED), as well as the schedule of planned surgeries, combined with temporal data on current weekday and nearby holidays. Discharge prediction for each individual patient is based on a random forest model that incorporates features extracted by natural language processing (NLP) tools from the orders and procedures datasets. These predictions also incorporate temporal effects related to the day of week and nearby holidays. All individual-patient discharge predictions are aggregated within a generalised linear model (GLM) to predict the overall number of expected discharges. ED, emergency department; PACU, post-anaesthesia care unit; SICU, surgical intensive care unit.

The top-down features included information on variations in occupancy associated with day of week, month of year and holidays, as well as the overall census in the top admitting departments. The bottom-up features included individual patient orders and procedures data.

Both prediction of expected admissions and prediction of expected discharges were based on a random forest model using the randomForest package in R.[17 18] The input features for the admission prediction model included the census in the top three admitting departments, the schedule of planned surgeries for the next day, as well as additional temporal information (month in year, day of week and nearby holidays). The input features for the discharge prediction model included the current day of admission, data on procedures and orders, and the temporal information described above. Further description of the features extracted from the procedures and orders data is provided in online supplemental appendix A.

Predictions for individual-patient discharges, based on both bottom-up and top-down features, were incorporated into a generalised linear model that took into account summary statistics on the total number of patients with a likelihood of discharge of over 80%, 90% and 95% within 24-hours to predict the total number of expected discharges (figure 1).

Model validation and performance evaluation

The model, developed using the training set, was validated on the testing set. Model performance was measured using median absolute error (MAE) and median absolute percentage error (MAPE). To evaluate model performance for binary predictions (such as the prediction of discharge within 24 hours for a given patient), we measured sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV). To measure the contribution of each of the three main features of the model (temporal features, number of predicted admissions and number of predicted discharges) to the overall prediction, a linear regression, with each variable separately and with all variables combined, was used. Then, the adjusted R^2 was measured for each variable and for all variables combined.

Secondary analysis: procedure duration and LOS

Alongside the main census prediction modelling framework described above, which aimed to predict census within the next 24 hours, we also developed a separate LOS prediction model in order to study the factors affecting a patient's LOS in the surgery unit. Specifically, we aimed to study the relationship between different surgical procedures and the LOS in the surgical unit following the procedure. For this analysis, we focused only on patients who underwent a surgical procedure prior to their admission.

First, we calculated the median LOS and IQR for each individual procedure, for all procedures of a given department and for each key-term extracted from the procedure's description. In addition, we calculated the Spearman's r correlation coefficient between procedure duration and LOS. We used the Spearman correlation since the distribution of the variables was not found to be normal according to the Shapiro-Wilk normality test. Finally, to assess the ability of these different factors to predict a patient's total LOS, a random forest model was developed. The model was based only on procedure related data (online supplemental appendix A), using the same training/validation sets as before. The accuracy of the prediction was measured in terms of MAPE and MAE.

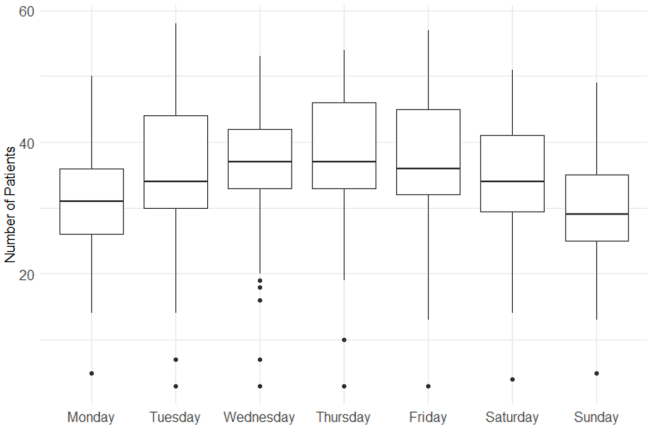

Figure 2 Occupancy of surgical department by day of week. This box-plot shows the median number of patients in the surgical unit at 8:00 hours each day, by day of week.

RESULTS

Variations in current census

There were 19 642 encounters in the surgical unit during the study period. Of these, 15 260 were used for training (78%) and 4382 used for validation (22%). The median number of patients in the surgical unit each day at 8:00 hours was 33 (IQR 28–40). The daily census varied from a minimum of 12 patients to a maximum of 58 patients per day, and the daily change in number of patients varied between a decrease of 20 patients to an increase of 17 patients.

Occupancy did not vary significantly from month to month (p=1), though weekday-to-weekend variations were more pronounced (p<0.001). On Sundays, the median number of patients was 27 while on Thursdays the median was 36 (figure 2). Variations in census were also noted around major US holidays. In all holidays but Columbus day, there was a decrease in the number of patients on both the day before and the day of the holiday. For most holidays, there was an overall decrease in the number of patients during the week surrounding the holiday (ie, no immediate compensatory increase in patients was observed after the holiday for the decrease in patients before the holiday). The holidays with the greatest decrease in the number of patients were Christmas Day (−8.4 patients), Thanksgiving Day (−8.2 patients) and Memorial Day (−6 patients). A detailed

summary of the effect of each holiday on occupancy is presented in online supplemental appendix B.

Predicting admissions

The median rate of admissions to the surgical unit was 12 patients per day (IQR 9–15). Ninety per cent of the patients came from three departments: PACU with 53% of admissions, ED with 31% of admissions and the surgical ICU with 6% of admissions. The admission prediction model was able to predict the number of new admissions to the surgical department with a MAPE of 16.7% and a MAE of two patients per day. The most significant factors identified by the model were the schedule of planned surgeries, the number of patients in the PACU and the day of the week.

Predicting discharges

The median rate of discharges from the surgical unit was 12 patients per day (IQR 9–15). The discharge prediction model was able to predict next-day discharge for individuals with an AUC of 0.84. At 90% specificity, the model was able to identify 57% of discharges with a PPV of 81% and NPV of 73% (online supplemental appendix C). The top factors associated with discharge timing included the number of days since admission (also representing the time since procedure), duration of the surgical procedure prior to admission and whether the procedure was planned or not.

As expected, 'discharge summary' orders were highly predictive of discharge with an OR of 13.4 for next-day discharge. Other less-intuitive orders associated with next-day discharge included those with descriptions containing the terms 'XR elbow' (OR of 5.6 (4.0–7.7)) and orders including the terms 'soft diet' (OR of 5.9 (4.9–7.0)). In contrast, patients receiving total parenteral nutrition, or any intravenous medications were unlikely to be discharged within 24 hours (OR <0.01). A summary of the top 40 orders to predict discharge (or lack of) can be found in online supplemental appendix C. Incorporating these individual-patient discharge predictions into a prediction of the overall discharges from the surgical unit, the generalised linear model was able to provide predictions with an error rate of 21.4% (MAPE) or a median of three patients per day (MAE).

Predicting next-day census

Integrating all of the above information into the census prediction formula (next-day census = current census + expected admissions − expected discharges), we were able to predict the next-day census in the validation set with a median error rate of 7% or three patients per day (IQR 1–5 patients per day) (figure 3). A prediction of an increase in the census was correct 85% of the time, while a prediction of a decrease in the census was only correct 60% of the time. The model was able to explain about 60% of the daily variability in census (R^2=0.58), with most of the variability explained by the day of week and nearby holidays (R^2=0.35), followed by the number of predicted admissions (R^2=0.26) and the number of predicted discharges (R^2=0.08).

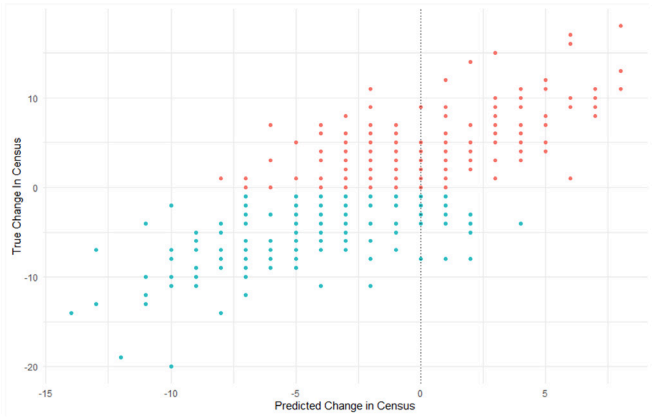

Figure 3 Predicted versus actual daily change in census during the validation time period. Observations to the left of the vertical dotted line (x=0) are days in which the model predicted a decrease in census. Observations to the right of the line are days in which the model predicted an increase in census. Days in which the actual census decreased are shown in blue (below y=0), while days in which the actual census increased are shown in red (above y=0).

Secondary analysis: surgery duration and LOS

As described in the methods section, alongside the main analysis of census prediction, we also developed a separate model in order to examine the factors affecting the overall LOS among patients who underwent a surgical procedure prior to their admission. Of all included surgical-unit patients, 68% (13 439/19 642) underwent a surgical procedure prior to their hospitalisation. Of these, 47% were for orthopaedics, 21% were for general surgery, 17% orofacial surgery, 9% plastic surgery, 2% gastrointestinal (GI) surgery and 2% genitourinary procedures (GU). Sixteen patients underwent more than one procedure. The 6229 admissions (32%) that were not preceded by a surgical procedure were either followed by a surgical procedure later during the hospitalisation (6.5%) or were without any documentation of a surgical procedure during the specific hospital encounter (25.5%).

Patients stayed in the surgical unit for a median of 1.7 days (IQR 0.9–3.3). LOS varied by the type and duration of the preadmission procedure. For example, the median LOS for patients undergoing 'medical' procedures (eg, placement of a central line) was 5.8 days, compared with only 1 day for those undergoing ophthalmology procedures (p<0.001). Patients undergoing spine halo application stayed for a median of 21 days (IQR 10–34), while patients undergoing tympanomastoidectomy stayed for a median 17 hours (IQR 12–22 hours, p<0.001). Procedures with descriptions that included terms such as 'exploratory' (eg, 'exploratory laparotomy') were associated with longer stays (median 6.1 days, IQR 2–10) while those that included terms such as 'percutaneous' (eg, 'elbow closed reduction with percutaneous pinning') were associated with shorter stays (22 hours

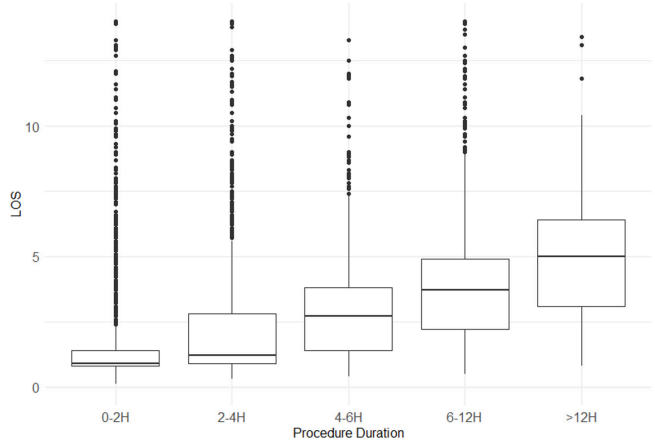

Figure 4 Length of stay (LOS) in the surgical unit by duration of surgical procedure. Boxplot showing the median LOS in days by the procedure duration in hours.

on median, IQR 19–38, p<0.001). Further details can be found in online supplemental appendix D.

Procedure duration was more predictive of total LOS than procedure type or procedure description text: Patients who underwent a procedure that took over 12 hours stayed on average 6 days longer than patients who underwent a procedure that took less than 2 hours (LOS of 8.4 days (95% CI 6.4 to 10.5) vs 2.3 days (95% CI 2.1 to 2.4) (figure 4)). Similarly, within a specific procedure type (eg, appendectomy), patients undergoing shorter procedures were likely to stay for a shorter time than those who underwent longer procedures (21 hours stay when the appendectomy took less than 1 hour vs 2 days when the appendectomy took longer than 1 hour, p<0.001). A detailed review of the correlation between procedure duration and the LOS can be found in online supplemental appendix D.

In online supplemental appendix E, we provide summary statistics for this separate model that predicts overall LOS based on procedure data. This model, separate from the main modelling framework, was able to predict LOS with a MAPE of 36% and MAE of 0.8 days, with the procedure duration serving by far as the most important factor for the prediction.

DISCUSSION

In this study, we developed and validated a model to predict next-day census in the surgical unit of a large tertiary paediatric hospital. We used a hierarchical modelling framework in order to isolate and evaluate the importance of individual prediction components (temporal, current census, discharges, admissions) and to evaluate which predictive factors best predict each of the components. This model was able to predict admissions with a median error of 16.7% (two patients per day), predict discharges with a median error of 21% (three patients per day) and predict next-day census with a median error of 7% (three patients per day). The factors found to be of most significance in predicting next-day census were the day of the week, followed by the number of predicted admissions, and finally the number of predicted discharges.

In addition to the prediction of next-day census, a particular focus was given for the prediction and analysis of the LOS after surgery. Based only on procedure-related data, a separate random forest model was able to predict LOS with a median error of 36% or 0.8 days per patient. By far the most significant factor for this prediction of LOS was the procedure duration. Longer procedures were associated with longer hospital stays, both when comparing all procedures only by their duration and when comparing the same procedure for different durations. These results are consistent with prior studies that have shown an association between procedure duration and total LOS,[10 19 20] yet these other studies were all carried out in an adult population where different factors drive procedure duration.[21] The association between procedure duration and LOS is not surprising, as major procedures are expected to take longer than minor procedures and to require longer in-hospital recovery time. Studies also suggest that prolonged procedures are associated with an increased risk for perioperative surgical-site infections, thus contributing even more to the overall LOS.[10]

The hierarchical modelling framework used in this study enables better understanding of the factors driving capacity issues and prediction accuracy. This approach also lends itself well to stepwise model implementation, from basic top-down temporal features such as day-of-week and holidays to bottom-up features extracted from detailed order sets using advanced natural language processing tools. In this study, we show the contribution of these different 'building blocks' and highlight the importance of easy-to-obtain temporal information to the overall prediction. In addition, we provide a systematic investigation of paediatric surgical procedures with respect to postoperative hospital LOS. We show that the LOS can be predicted with good accuracy based on procedures data alone, and that procedure duration is highly predictive of the total LOS for most, but not for all paediatric procedures. Since paediatric surgical procedures differ greatly from adult procedures, and since previous studies have focused on the adult population, our results can provide an important contribution to the field of paediatric capacity planning.

This study has several limitations. First, it is a single-centre retrospective study of a tertiary paediatric hospital, and thus, the findings may not be applicable to other settings. Second, while the present model was developed to predict next-day census, other settings may require different timescales for prediction—from the prediction of hourly changes in census, to long-term prediction of the census in the next week, month or year. These different timescales will require retraining of the model according to the desired outcomes. Third, only procedures completed prior to the surgical-unit admission were included in the models, possibly omitting valuable information from the prediction. Nevertheless, we would

expect the models to only improve if this information would be added in future implementations. Fourth, as in any prediction model, the census prediction model is not 100% accurate. When used in clinical practice, clinicians and administrators can incorporate its predictions as one of several inputs in their decision-making processes. Lastly, our data included only data recorded prior to the COVID-19 pandemic, and thus may not be representative of data collected during the pandemic.

As the cost of building inpatient bed spaces continues to rise and financial pressure on paediatric hospitals increases, efficient utilisation of existing inpatient spaces becomes increasingly vital for healthcare sustainability. Predictive tools make the smoothing of elective procedures feasible and enable proactive planning to align staffing and other expensive resources. Similarly, in facilities with excess capacity, predictive tools help minimise waste of resources during periods of low occupancy. Facility planning and long-range staffing strategies will be more accurate with the help of high-performance system-wide prediction models. As healthcare resources are not stretched too much, quality and patient safety can be improved greatly.

We have shown that surgery occupancy prediction in a paediatric setting is plausible. We have further shown that surgical unit occupancy is dependent on both top-down temporal factors as well as bottom-up individual-patient information, including data on surgical procedures planned and performed, occupancy in other related departments, and clinical and administrative orders given to admitted patients. A hierarchical modelling framework that combines both types of factors has the potential to be better suited for predicting future surgical-unit occupancy, supporting decision-makers in their quest for improved scheduling, staffing and resource planning, reducing overcrowding and cancellations of surgeries in paediatric healthcare settings.

Contributors study conception and design: YBC, JM, RW, CC, JB and BYR; data collection: JB, JM and RW; analysis and interpretation of results: YBC, JM, RW; draft manuscript preparation: YBC, CC, JB and BYR. All authors reviewed the results and approved the final version of the manuscript. BYR acts as the guarantor of the study.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient consent for publication Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available on reasonable request. Anonymised data are available on reasonable request.

ORCID iD

Yuval Barak Corren http://orcid.org/0000-0002-9603-4728

REFERENCES

1 Wong DJN, Harris SK, Moonesinghe SR, *et al*. Cancelled operations: a 7-day cohort study of planned adult inpatient surgery in 245 UK National health service hospitals. *Br J Anaesth* 2018;121:730–8.
2 Gillies MA, Wijeysundera DN, Harrison EM. Counting the cost of cancelled surgery: a system wide approach is needed. *Br J Anaesth* 2018;121:691–4.
3 Herrod PJJ, Adiamah A, Boyd-Carson H, *et al*. Winter cancellations of elective surgical procedures in the UK: a questionnaire survey of patients on the economic and psychological impact. *BMJ Open* 2019;9:e028753.
4 Hanna TP, King WD, Thibodeau S, *et al*. Mortality due to cancer treatment delay: systematic review and meta-analysis. *BMJ* 2020;371:m4087.
5 Kortbeek N, Braaksma A, Burger CAJ, *et al*. Flexible nurse staffing based on hourly bed census predictions. *Int J Prod Econ* 2015;161:167–80.
6 Weissman C, Scemama J, Weiss YG. The ratio of PACU length-of-stay to surgical duration: practical observations. *Acta Anaesthesiol Scand* 2019;63:1143–51.
7 Baek H, Cho M, Kim S, *et al*. Analysis of length of hospital stay using electronic health records: a statistical and data mining approach. *PLoS One* 2018;13:e0195901.
8 Kasotakis G, Schmidt U, Perry D, *et al*. The surgical intensive care unit optimal mobility score predicts mortality and length of stay. *Crit Care Med* 2012;40:1122–8.
9 Schneider EB, Hyder O, Wolfgang CL, *et al*. Provider versus patient factors impacting Hospital length of stay after pancreaticoduodenectomy. *Surgery* 2013;154:152–61.
10 Tan T-W, Kalish JA, Hamburg NM, *et al*. Shorter duration of femoral-popliteal bypass is associated with decreased surgical site infection and shorter hospital length of stay. *J Am Coll Surg* 2012;215:512–8.
11 Clague JE, Craddock E, Andrew G, *et al*. Predictors of outcome following hip fracture. admission time predicts length of stay and in-hospital mortality. *Injury* 2002;33:1–6.
12 Djurasovic M, Kiskaddon E, Bratcher KR, *et al*. Determinants of hospital length of stay in patients undergoing lumbar spinal fusion. *The Spine Journal* 2014;14:S103–4.
13 Zheng F, Cammisa FP, Sandhu HS, *et al*. Factors predicting hospital stay, operative time, blood loss, and transfusion in patients undergoing revision posterior lumbar spine decompression, fusion, and segmental instrumentation. *Spine* 2002;27:818–24.
14 Dexter F, Epstein RH, Shi P. Forecasting the probability that each surgical case will either be ambulatory or the patient will remain in the hospital overnight versus having a length of stay of two or more days. *Cureus* 2020;12:e10847.
15 Shapiro NL, Bhattacharyya N. Complications and determinants of length of stay for inpatient pediatric otolaryngologic procedures. *Arch Otolaryngol Head Neck Surg* 2003;129:169–72.
16 Khare M, Rauch DA. Trends in national pediatric bed census. *Pediatrics* 2017;140:20.
17 Liaw A, Wiener M. Classification and regression by randomForest. *R News* 2002;2:18–22.
18 R Core Team. R: a language and environment for statistical computing. Vienna, Austria R Foundation for Statistical Computing; 2020. https://www.R-project.org/
19 Garbarino LJ, Gold PA, Sodhi N, *et al*. The effect of operative time on in-hospital length of stay in revision total knee arthroplasty. *Ann Transl Med* 2019;7:66.
20 Collins TC, Daley J, Henderson WH, *et al*. Risk factors for prolonged length of stay after major elective surgery. *Ann Surg* 1999;230:251.
21 Bravo F, Levi R, Ferrari LR, *et al*. The nature and sources of variability in pediatric surgical case duration. *Paediatr Anaesth* 2015;25:999–1006.