臺北醫學大學
TAIPEI MEDICAL UNIVERSITY

Research using TMUCRD Data

Jason C. Hsu

Chair, Local Host Committee of the 2022 OHDSI APAC Symposium
Associate Professor, Taipei Medical University, Taiwan
Director, Clinical Data Center, Office of Data Science, Taipei Medical University

Outline

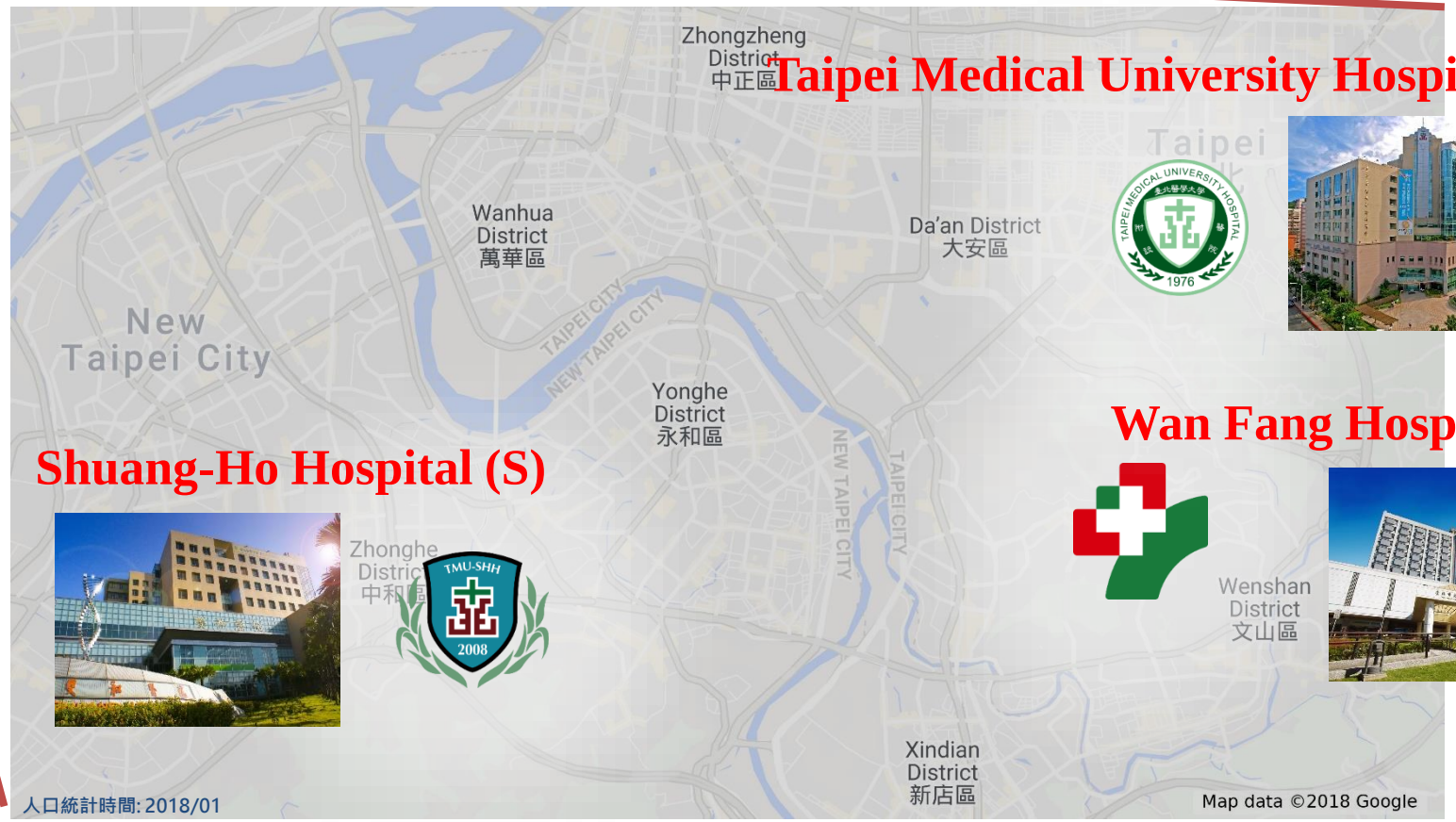
1. TMUCRD: Introduction

2. TMUCRD: Development

3. TMUCRD: Applications

1. TMUCRD: Introduction

TMU's Three Affiliated Hospitals

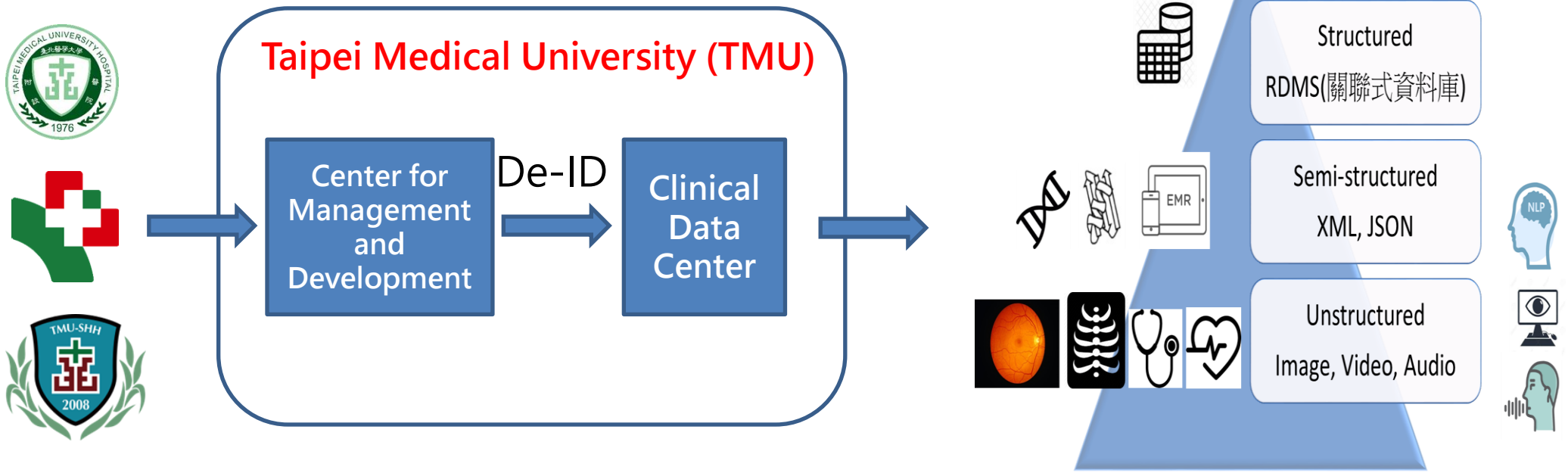


Taipei Medical University Hospital (T)

Shuang-Ho Hospital (S)

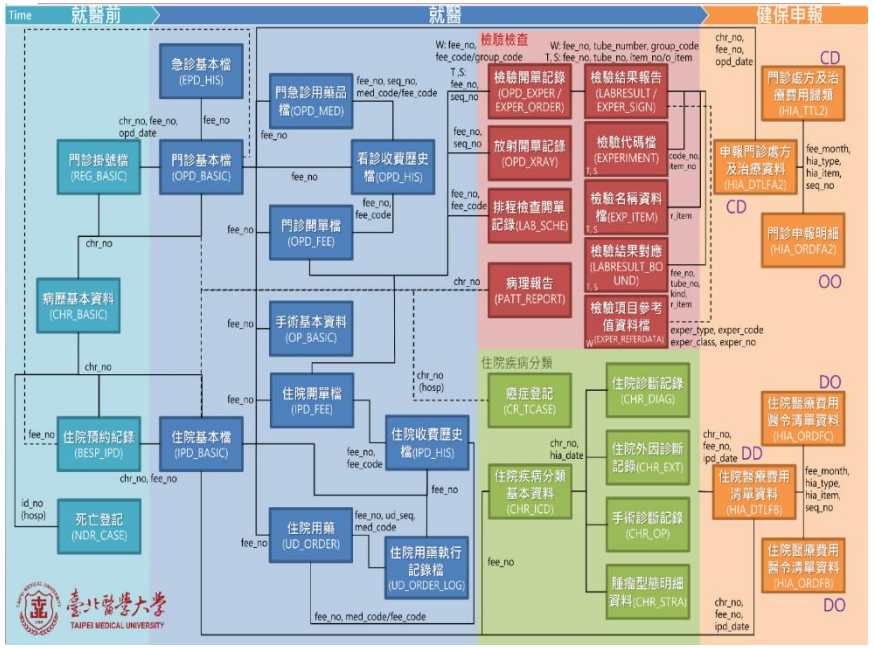
Wan Fang Hospital (W)

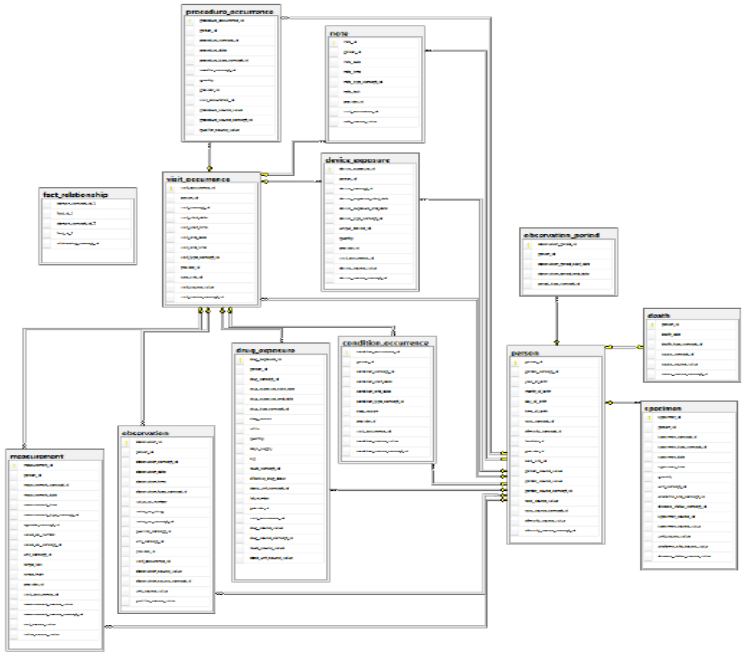
Taipei Medical University Clinical Research Database (TMUCRD)



TMUCRD

Composition of TMUCRD





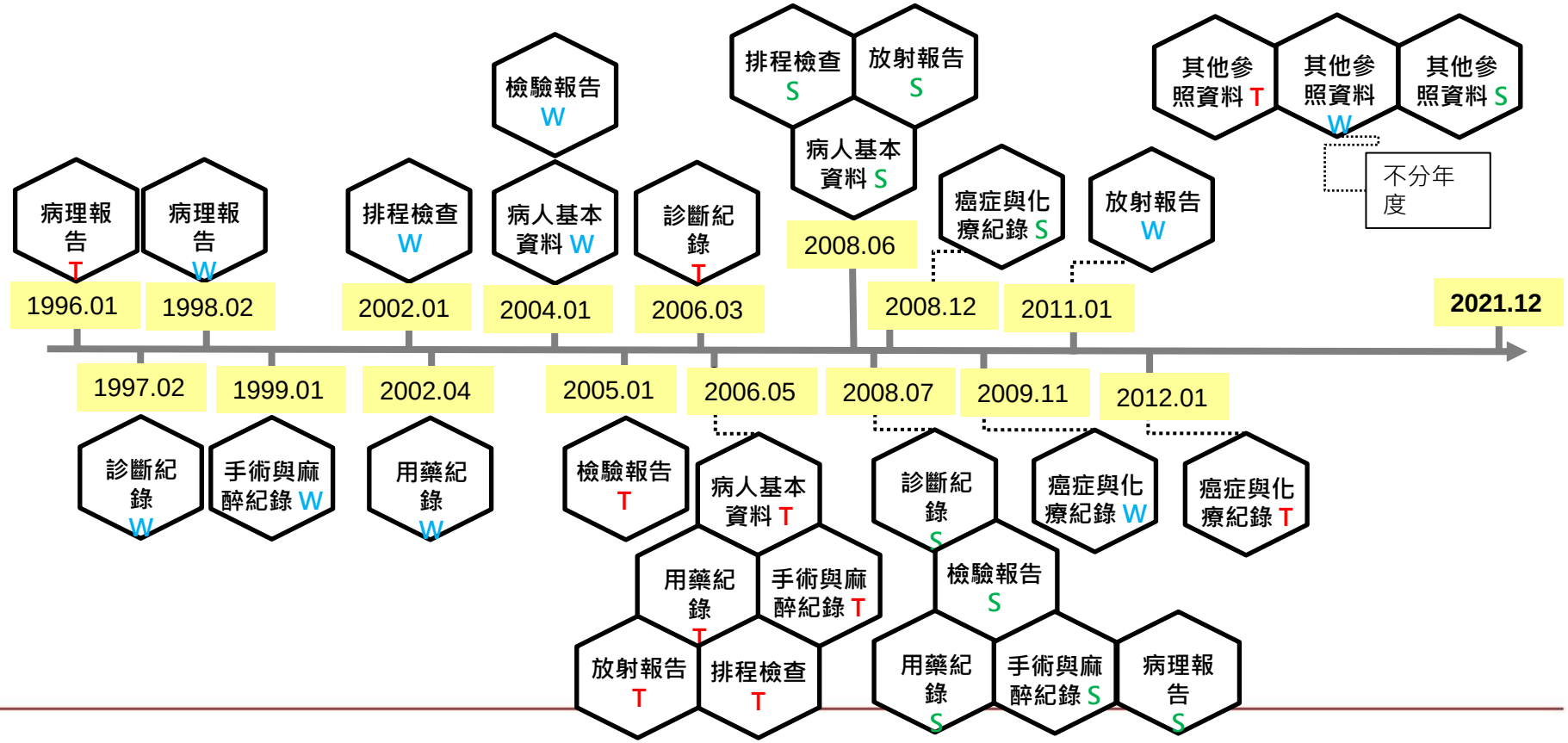
PERSON
OBSERVATION_PERIOD
SPECIMEN
DEATH
VISIT_OCCURRENCE
VISIT_DETAIL
PROCEDURE_OCCURRENCE
DRUG_EXPOSURE
DEVICE_EXPOSURE
CONDITION_OCCURRENCE
MEASUREMENT
NOTE
NOTE_NLP
OBSERVATION
FACT_RELATIONSHIP

TMUCRD **OHDSI OMOP CDM**

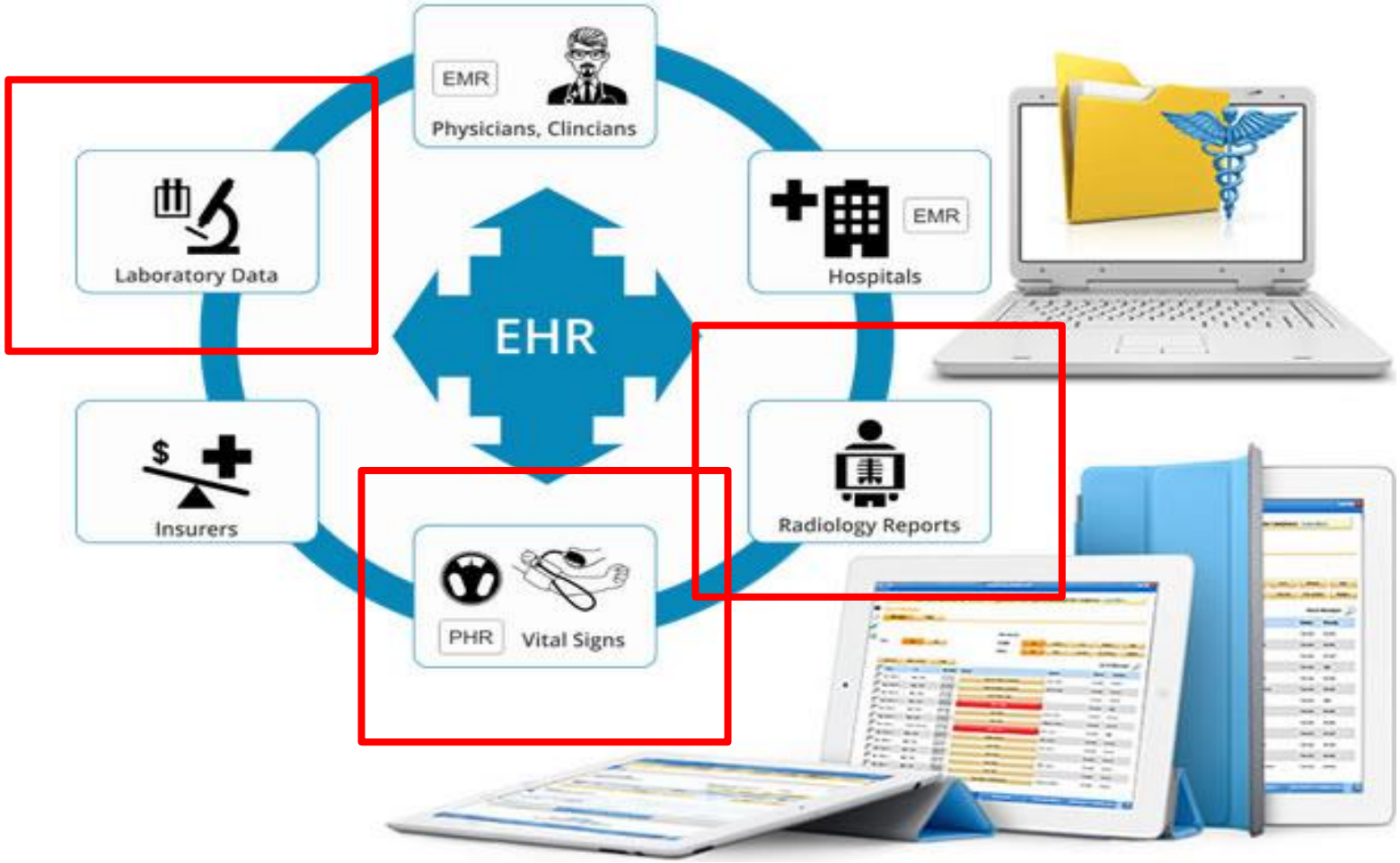
Time Period and Case Number

1996-2021 (25 years)

N=4,125,097 (18.7% of Taiwan's population)



Features of TMUCRD (compared to NHIRD)



Health Examination

New Drugs

TMUCRD was linked to cancer registry and death registry databases

Cancer Patients' ID

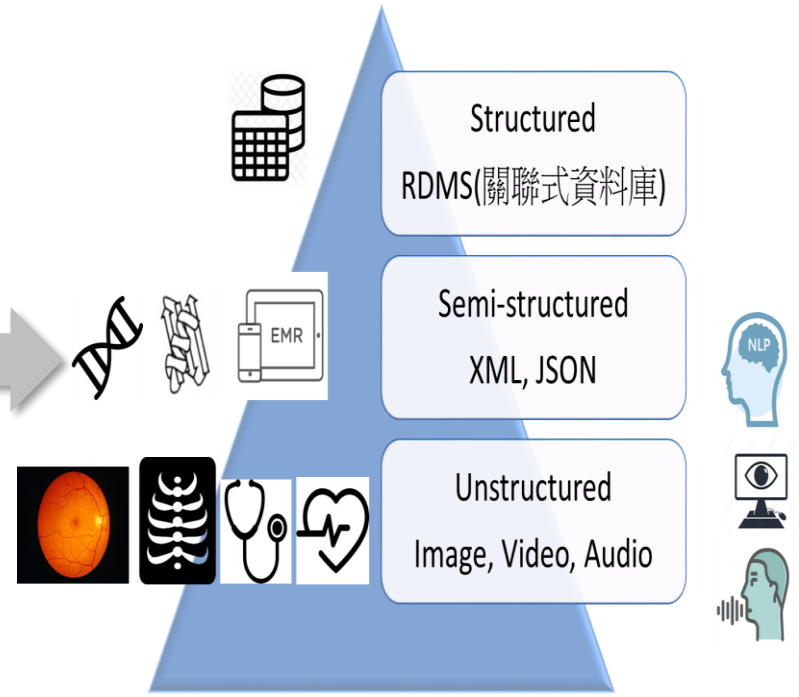
All Patients' ID

Cancer Registry

Death Registry

1. De-ID
2. Matching

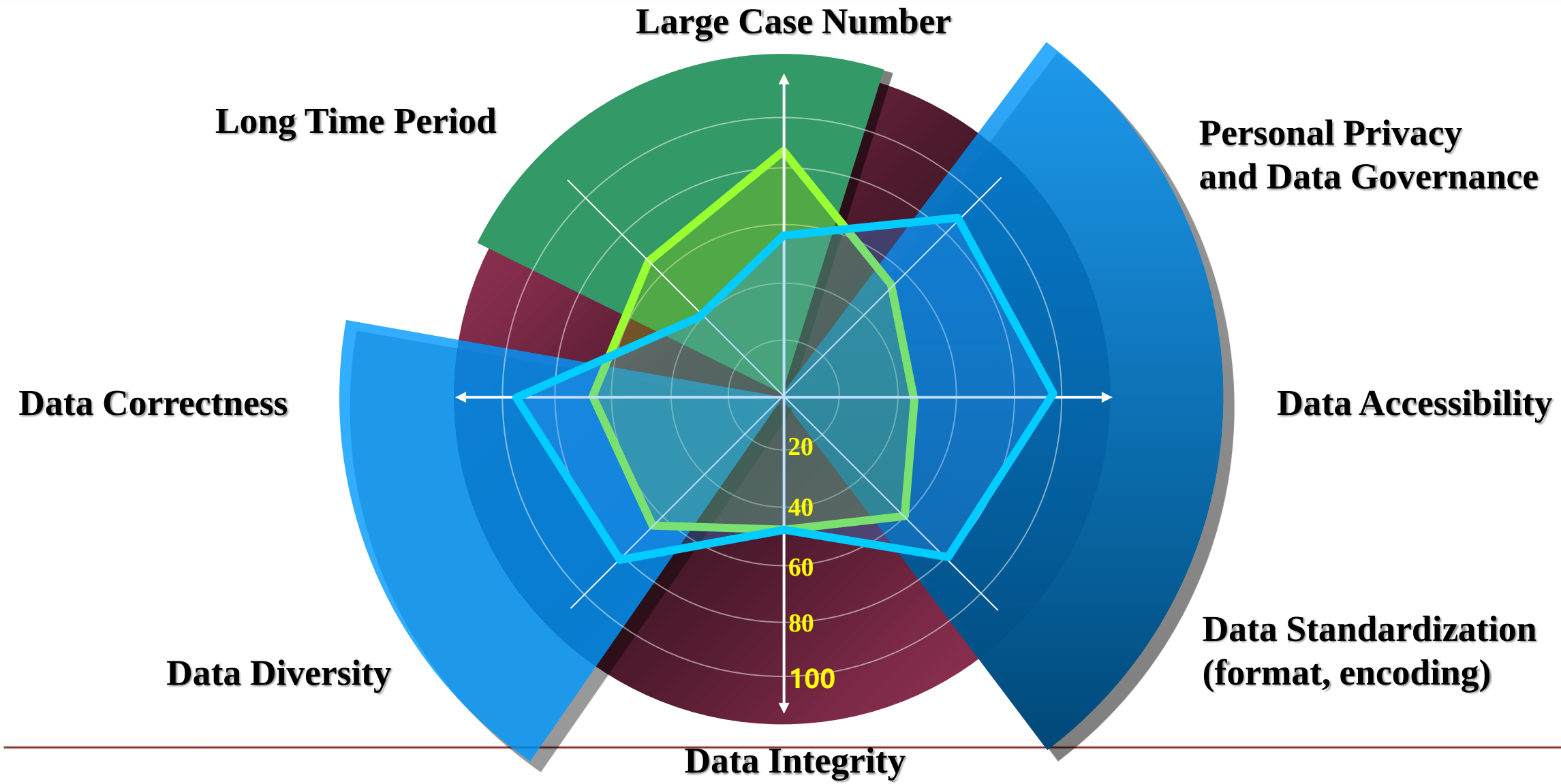
Center for Management and Development



TMUCRD


2. TMUCRD: Development

High Quality Database



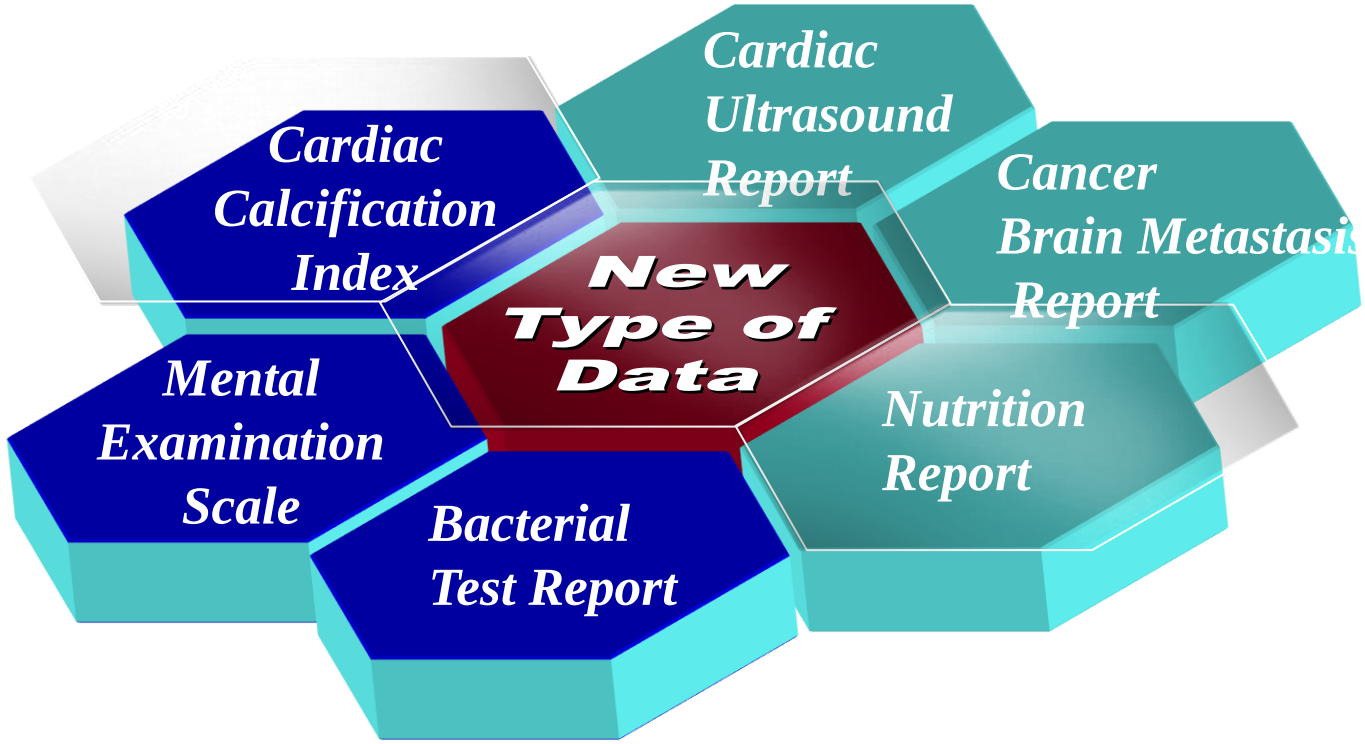
From unstructured data to structured data

Structure unstructured text reports and move towards automated reporting

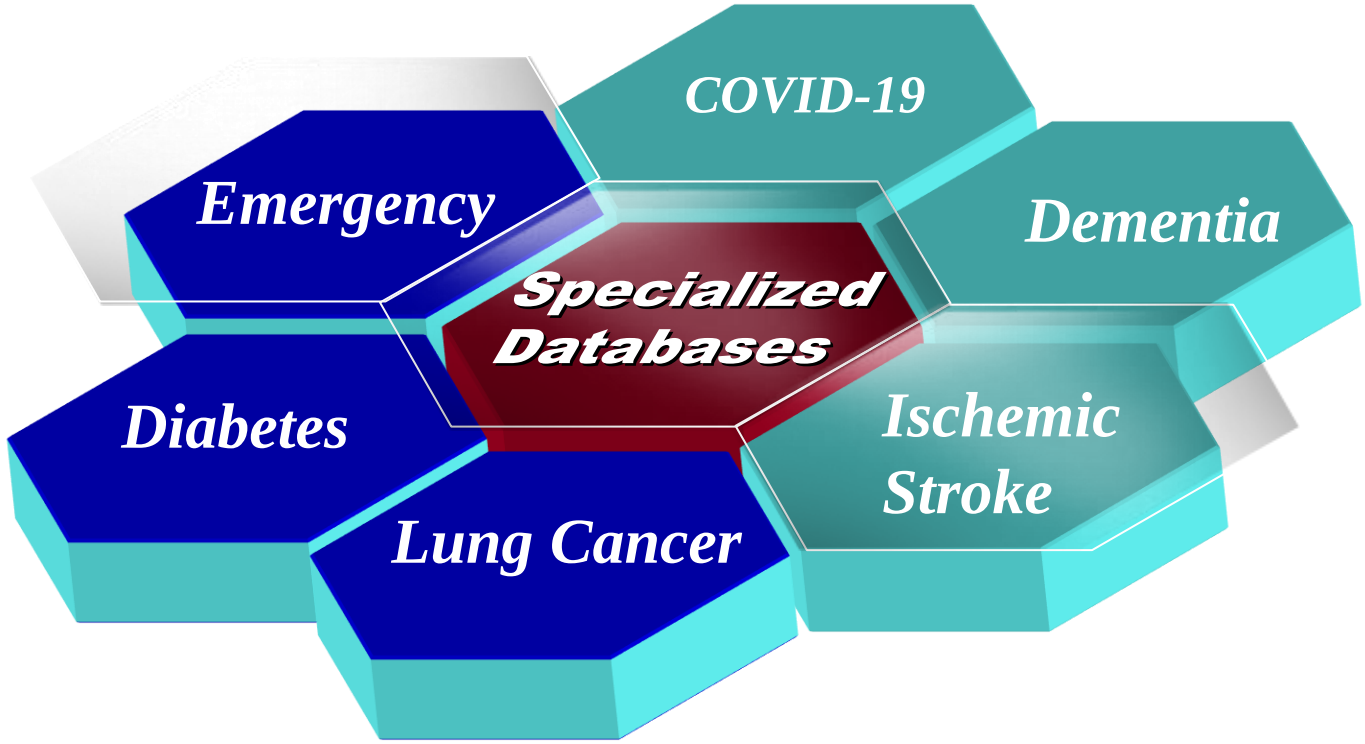
ORGH SOAP	**organ**	**locat**	**operation**	**diagnosis**
	Breast	left	partial mastectomy	Invasive ductal carcinoma
ORGH Report	**Histologic type**		**Histologic grade**	**pT**
	invasive ductal carcinoma		Grade 3 (scores of 8 or 9)	pT1c: Tumor >10 mm but < or =20 mm in greatest dimension.
	pN		**pM**	**pTNM**
	pN(sn)0 (i-): No sentinel lymph node metastasis identified histologically, negative IHC		Not applicable.	T1 N0 M0 IA
	ER	**PR**	**Her2**	**Ki-67**
	negative	positive	positive (Score 3+)	30%

【Unstructured Data】 【Structured Data】

Create New Type of Data



Build Specialized Databases



Cross-Institutional Data Research Cooperation

Diabetes 台大醫院 NTUH

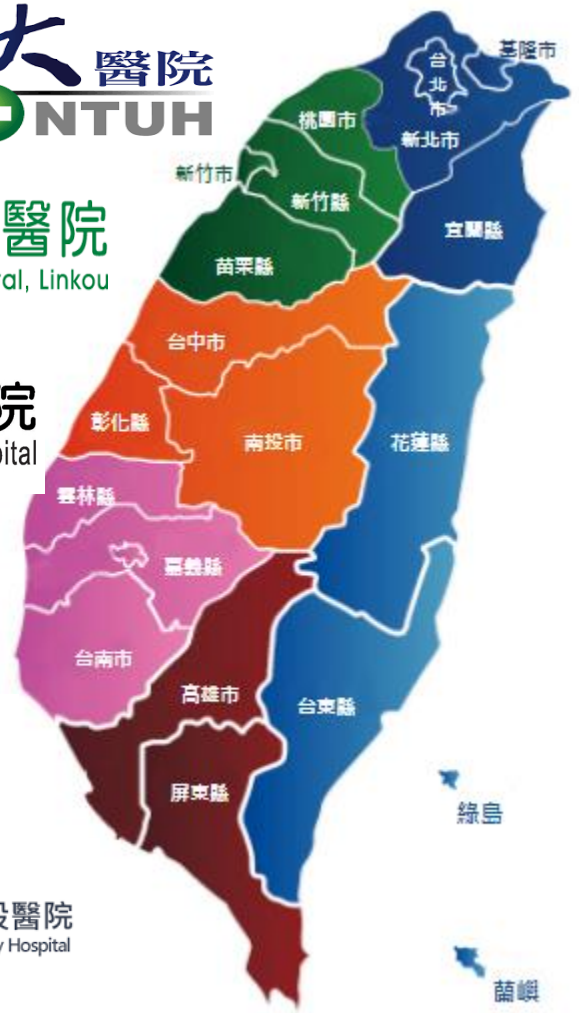
CKD 林口長庚紀念醫院 Chang Gung Memorial Hospital, Linkou

童綜合醫療社團法人 童綜合醫院 Tungs' Taichung MetroHarbor Hospital

戴德森醫療財團法人 嘉義基督教醫院 CHIA-YI CHRISTIAN HOSPITAL

Emergency 奇美醫療財團法人 奇美醫院 Chi Mei Medical Center

Dementia 國立成功大學附設醫院 National Cheng Kung University Hospital



Academic-Industry Collaboration

VIATRIS™

GCSF and Febrile Neutropenia

IQVIA

Inflammatory Bowel Disease (IBD)

NOVARTIS

Renal Dysfunction

3. TMUCRD: Applications

Rich research performance from the use of TMUCRD

Large Research Projects (9)

Theses and Dissertations (23)

Academic Publications (25)

Observational Study (Estimation)

Journal of *Clinical Medicine*

MDPI

Article

Adverse Outcomes after Major Surgeries in Patients with Diabetes: A Multicenter Matched Study

Chao-Shun Lin [1,2,3], Chuen-Chau Chang [1,2,3], Yuan-Wen Lee [1,2,3], Chih-Chung Liu [1,2,3], Chun-Chieh Yeh [4,5], Yi-Cheng Chang [6], Ming-Tsang Chuang [7,8,9], Tzu-Hao Chang [7,8,9,†], Ta-Liang Chen [1,2,3,‡] and Chien-Chang Liao [1,2,3,10,11,*]

1 Department of Anesthesiology, School of Medicine, College of Medicine, Taipei Medical University, Taipei 110, Taiwan; lin.soon@gmail.com (C.-S.L.); nekota@tmu.edu.tw (C.-C.C.); yuarn438@yahoo.com.tw (Y.-W.L.); aneslcc@gmail.com (C.-C.L.); tlc@tmu.edu.tw (T.-L.C.)
2 Department of Anesthesiology, Taipei Medical University Hospital, Taipei 110, Taiwan
3 Anesthesiology and Health Policy Research Center, Taipei Medical University Hospital, Taipei 110, Taiwan
4 Department of Surgery, China Medical University Hospital, Taichung 404, Taiwan; b8202034@gmail.com
5 Department of Surgery, University of Illinois, Chicago, IL 60637, USA
6 Department of Internal Medicine, National Taiwan University Hospital, Taipei 100, Taiwan;
10 School of Chinese Medicine, College of Chinese Medicine, China Medical University, Taichung 404, Taiwan

To assess postoperative complications and mortality in patients with diabetes.

Diabetic patients undergoing surgery have higher risk of infectious complications and hospitalization and mortality compared with non-diabetic patients undergoing similar major surgery

Observational Study (Estimation)

biomedicines

MDPI

Article

Concurrent Blockade of Endothelial EGFR and VEGF Signaling on Malignant Associated Pleural Fluid Induced Angiogenesis: From Clinic to Bench

Wei-Teing Chen [1,2,†], Yu-Huei Lin [3,†], Chih-Ying Changchien [2,4,†], Ying Chen [4], Hsin-Han Chang [4], Wen-Chiuan Tsai [5], Hao-Chung Tsai [6], Chieh-Yung Wang [7], Ming-Sheng Shen [8], Li-Ting Cheng [7] and Chen-Liang Tsai [7,*]

[1] Division of Chest Medicine, Department of Medicine, Cheng-Hsin General Hospital, Taipei 112, Taiwan; stigma712@yahoo.com.tw

[2] Department of Internal Medicine, Tri-Service General Hospital, National Defense Medical Center

[5] Department of Pathology, Tri-Service General Hospital, National Defense Medical Center, Taipei 114, Taiwan; ab95057@hotmail.com

check for

Malignant-associated pleural fluid (MAPF) represented an unsolved problem in advanced lung cancer.

The results indicated that the addition of bevacizumab on gefitinib treatment could suppress MAPF-induced angiogenesis in lung adenocarcinoma patients.

Observational Study (Estimation)

PLOS ONE

RESEARCH ARTICLE

Investigation of dual antiplatelet therapy after coronary stenting in patients with chronic kidney disease

Chih-Chin Kao[1,2,3], Mai-Szu Wu[1,3,4], Ming-Tsang Chuang[5], Yi-Cheng Lin[6,7], Chun-Yao Huang[8,9,10], Wei-Chiao Chang[4,11,12], Chih-Wei Chen[9,10]*, Tzu-Hao Chang[13,14]*

1 Division of Nephrology, Department of Internal Medicine, School of Medicine, College of Medicine, Taipei Medical University, Taipei, Taiwan, 2 Division of Nephrology, Department of Internal Medicine, Taipei Medical University Hospital, Taipei, Taiwan, 3 TMU Research Center of Urology and Kidney (TMU-RCUK), Taipei Medical University, Taipei, Taiwan, 4 Division of Nephrology, Department of Internal Medicine, Shuang-Ho Hospital, Taipei Medical University, New Taipei City, Taiwan, 5 Clinical Data Center, Office of Pharmacy, Taipei Medical University, Taipei, Taiwan, 12 Master Program for Clinical Pharmacogenomics

To compare the efficacy and safety of long-term and short-term dual antiplatelet therapy (DAPT) after coronary stenting in patients with CKD

Long-term DAPT was associated with similar risk of MACE (HR: 1.05, 95% CI: 0.65–1.70, *P* = 0.83) compare with short-term DAPT. Different CKD risk did not modify the risk of MACE. There was also no significant difference in all-cause mortality (HR: 1.10, 95% CI: 0.75–1.61, *P* = 0.63) and TIMI bleeding (HR 1.19, 95% CI: 0.86–1.63, *P* = 0.30) between groups.

Observational Study (Estimation)

PLOS ONE

RESEARCH ARTICLE

Association between intracytoplasmic sperm injection and neurodevelopmental outcomes among offspring

Cheng-Wei Wang[1☯], Tzu-Hao Chang[2,3☯], Nai-Chen Chuang[4], Heng-Kien Au[5,6], Chi-Huang Chen[1,5], Sung-Hui Tseng[7,8]*

1 Division of Reproduction Medicine, Department of Obstetrics and Gynecology, Taipei Medical University Hospital, Taipei, Taiwan, **2** Graduate Institute of Biomedical Informatics, College of Medical Science and Technology, Taipei Medical University, Taipei, Taiwan, **3** Clinical Big Data Research Center, Taipei Medical University Hospital, Taipei, Taiwan, **4** Clinical Data Center, Office of Data Science, Taipei Medical University, Taipei, Taiwan, **5** Department of Obstetrics and Gynecology, School of Medicine, College of Medicine, Taipei Medical University, Taipei, Taiwan, **6** Department of Obstetrics and Gynecology, Taipei Medical University Hospital, Taipei, Taiwan, **7** Department of Physical Medicine and Rehabilitation, Taipei Medical University Hospital, Taipei, Taiwan, **8** Department of Physical Medicine and Rehabilitation, School of Medicine, College of Medicine, Taipei Medical University, Taipei, Taiwan

Our study indicated that the use of ICSI does not associated with higher risk of neurodevelopmental disorders in the offspring. But male sex, and ICU admission do have increased risk of neurodevelopmental disorders.

neurodevelopmental outcomes among offspring. PLoS ONE 16(9): e0257268. https://doi.org/10.1371/journal.pone.0257268

Purpose

To compare the risk of neurodevelopmental disorders in children conceived via intracytoplasmic sperm injection (ICSI) and those conceived naturally.

Observational Study (Estimation)

ARTICLE IN PRESS

CLINICAL INVESTIGATION

Relationship between metformin use and lactic acidosis in advanced chronic kidney disease: The REMIND-TMU Study

Chien-Chou Chen, MD[1,2,3], Yu Ko, PhD[4,5], Chin-Hua Chen, PhD[6], Yi-Jen Hung, MD, PhD[7], Ting-En Wei, MD[3,8], Tzu-Hao Chang, PhD[9,10],

Hospital Songshan Branch, Taipei, Taiwan; [2] Graduate Institute of Clinical Medicine,

The goal of the study is to analyze the incidence and associated factors of lactic acidosis between metformin user and non-user with advanced CKD.

Conclusions: Metformin was associated with a significant increased risk of laboratory-defined lactic acidosis (p=0.0204) even after adjusting confounder such as age, sex and underlying comorbidities. This "REMIND" study reminds us that metformin- associated lactic acidosis is mainly caused by decreased drug renal elimination other than underlying comorbidities in advanced CKD patients.

Observational Study (Intervention)

Journal of *Clinical Medicine*

MDPI

Article

Effects of the COVID-19 Pandemic on Treatment Efficiency for Traumatic Brain Injury in the Emergency Department: A Multicenter Study in Taiwan

Carlos Lam [1,2], Ju-Chuan Yen [3,4], Chia-Chieh Wu [1,2], Heng-Yu Lin [5] and Min-Huei Hsu [6,*]

1 Emergency Department, Wan Fang Hospital, Taipei Medical University, Taipei 11696, Taiwan; lsk@w.tmu.edu.tw (C.L.); setfreej@gmail.com (C.-C.W.)
2 Department of Emergency, School of Medicine, College of Medicine, Taipei Medical University, Taipei 11030, Taiwan
3 Department of Ophthalmology, Taipei City Hospital, Renai Branch, Taipei 10629, Taiwan; m701061@tmu.edu.tw
4 Graduate Institute of Biomedical Informatics, College of Medical Technology, Taipei Medical University, Taipei 11030, Taiwan
5 School of Medicine, College of Medicine, Taipei Medical University, Taipei 11030, Taiwan; b101105091@tmu.edu.tw
6 Graduate Institute of Data Science, College of Management, Taipei Medical University, Taipei 11030, Taiwan

The time interval between ED arrival and brain CT was significantly shortened during P1 and P2 compared with the pre-pandemic interval, and no significant delay between ED arrival and surgical management was found, indicating increased treatment efficiency for TBI in the ED during the COVID-19 pandemic.

Observational Study (Intervention)

Open access Original research

BMJ Open

Early Chronic Kidney Disease Care Programme delays kidney function deterioration in patients with stage I–IIIa chronic kidney disease: an observational cohort study in Taiwan

Shu-Fen Niu,[1,2,3] Chung-Kuan Wu ,[4,5] Nai-Chen Chuang,[6] Ya-Bei Yang,[7] Tzu-Hao Chang[8,9]

To cite: Niu S-F, Wu C-K, Chuang N-C, *et al*. Early

org/10.1136/bmjopen-2020-041210).

ABSTRACT

Objectives To investigate the effect of the Early

propensity score to reduce bias between two groups.

Outcome measures The risks of CKD stage I–IIIa

Strengths and limitations of this study

early CKD, rather than major adverse cardiac events or mortality.

Compared with the control group, the case group demonstrated more comorbidities and higher proportions of hypertension, diabetes mellitus, gout, dyslipidaemia, heart disease and cerebrovascular disease, but had lower risk of progression to CKD stage IIIb before and (HR 0.72; 95% CI 0.61 to 0.85) and after (adjusted HR (aHR) 0.67; 95% CI 0.55 to 0.81) adjustments.

Machine Learning Study (Prediction)

Original Paper

Machine-Learning Monitoring System for Predicting Mortality Among Patients With Noncancer End-Stage Liver Disease: Retrospective Study

Yu-Jiun Lin[1,2], MD; Ray-Jade Chen[3,4], MD, MSc; Jui-Hsiang Tang[5], MD; Cheng-Sheng Yu[1,2*], PhD; Jenny L Wu[1,2], BSc; Li-Chuan Chen[6,7], MSN; Shy-Shin Chang[1,2,6*], MD, PhD

[1]Department of Family Medicine, School of Medicine, College of Medicine, Taipei Medical University, Taipei, Taiwan
[2]Department of Family Medicine, Taipei Medical University Hospital, Taipei, Taiwan
[3]Department of Surgery, School of Medicine, College of Medicine, Taipei Medical University, Taipei, Taiwan
[4]Division of General Surgery, Department of Surgery, Taipei Medical University Hospital, Taipei, Taiwan
[5]Division of Gastroenterology and Hepatology, Department of Internal Medicine, Taipei Medical University Hospital, Taipei, Taiwan
[6]Department of Community and Preventive Medicine, Taipei Medical University Hospital, Taipei, Taiwan
[7]School of Gerontology Health Management, College of Nursing, Taipei Medical University, Taipei, Taiwan
*these authors contributed equally

Corresponding Author:
Shy-Shin Chang, MD, PhD
Department of Family Medicine
School of Medicine, College of Medicine
Taipei Medical University
250 Wuxing Street
Taipei, 11031

AUCROC = 0.852 (Random Forest)

Machine Learning Study (Prediction)

JOURNAL OF MEDICAL INTERNET RESEARCH Yu et al

Original Paper

Development of an Online Health Care Assessment for Preventive Medicine: A Machine Learning Approach

Cheng-Sheng Yu[1,2], PhD; Yu-Jiun Lin[1,2], MD; Chang-Hsien Lin[1,2], MD; Shiyng-Yu Lin[1,2], MD; Jenny L Wu[1,2], BSc; Shy-Shin Chang[1,2], MD, PhD

[1]Department of Family Medicine, Taipei Medical University Hospital, Taipei, Taiwan

To construct a medical database system from electronic medical records (EMRs) of subjects who have undergone health examination.

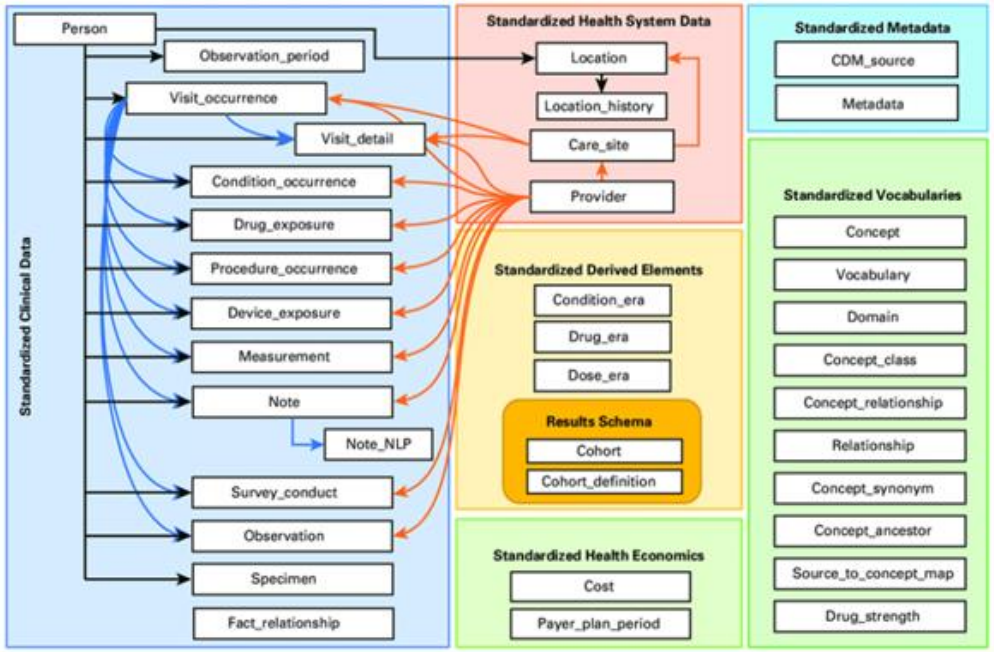
This system aims to provide online self-health evaluation to clinicians and patients, enabling personalized health and preventive health.

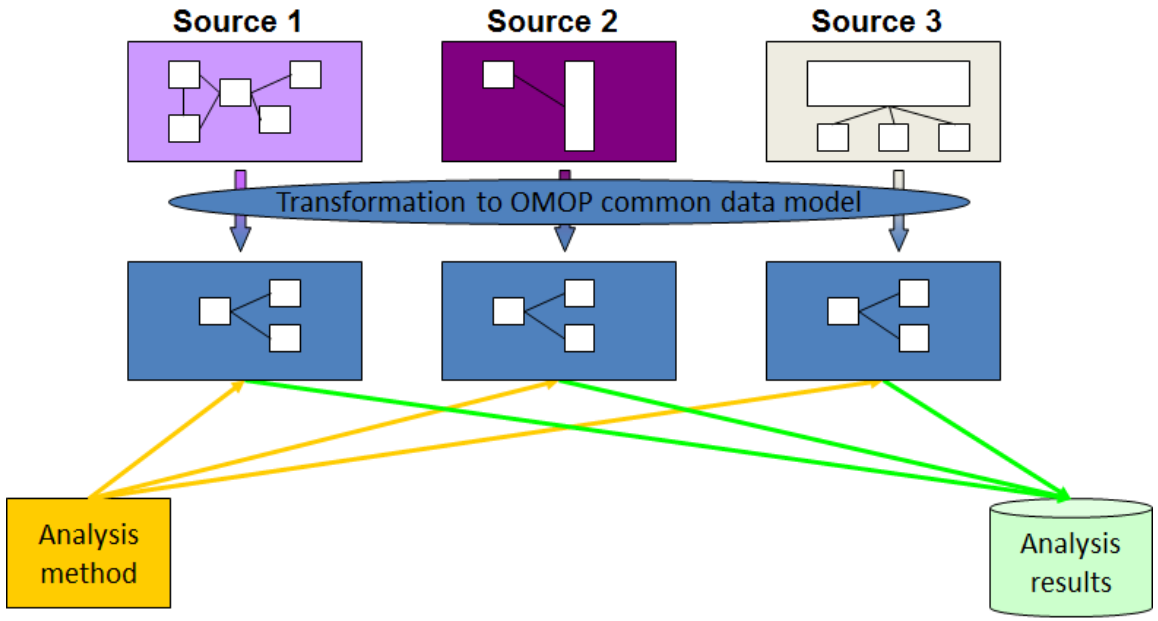
Phone: 886 2 23565926
Email: sschang0529@gmail.com

Related Article:
This is a corrected version. See correction statement in: h

AUCROC = 0.904 (predict metabolic syndrome)
AUCROC = 0.982 (predict CKD)

From TMUCRD study to OHDSI study





JAMA Network Open

(IF=13.353)

Original Investigation | Cardiology

Analysis of Dual Combination Therapies Used in Treatment of Hypertension in a Multinational Cohort

Yuan Lu, ScD; Mui Van Zandt, BS; Yun Liu, PhD; Jing Li, MS; Xialin Wang, MS; Yong Chen, PhD; Zhengfeng Chen, MBBS, MMed; Jaehyeong Cho, PhD; Sreemanee Raaj Dorajoo, PhD; Mengling Feng, PhD; Min-Huei Hsu, MD, PhD; Jason C. Hsu, PhD; Usman Iqbal, PharmD, MBA, PhD; Jitendra Jonnagaddala, PhD; Yu-Chuan Li, MD, PhD; Siaw-Teng Liaw, MBBS, PhD; Hong-Seok Lim, MD, PhD; Kee Yuan Ngiam, MBBS, MMed; Phung-Anh Nguyen, PhD; Rae Woong Park, MD, PhD; Nicole Pratt, PhD; Christian Reich, MD, PhD; Sang Youl Rhee, MD; Selva Muthu Kumaran Sathappan, MSc; Seo Jeong Shin, PhD; Hui Xing Tan, MTech; Seng Chan You, MD, PhD; Xin Zhang, MS; Harlan M. Krumholz, MD, SM; Marc A. Suchard, MD, PhD; Hua Xu, PhD

Abstract

IMPORTANCE More than 1 billion adults have hypertension globally, of whom 70% cannot achieve their hypertension control goal with monotherapy alone. Data are lacking on clinical use patterns of dual combination therapies prescribed to patients who escalate from monotherapy.

OBJECTIVE To investigate the most common dual combinations prescribed for treatment escalation in different countries and how treatment use varies by age, sex, and history of cardiovascular disease.

DESIGN, SETTING, AND PARTICIPANTS This cohort study used data from 11 electronic health
illion patients across 8 countries and regions between January
ded participants were adult patients (ages ≥18 years) who newly

Key Points

Question What are the most common antihypertensive dual combinations prescribed to patients who escalate from monotherapy in clinical practice, and how do the combinations differ by country and patient demographic subgroup?

Findings In this cohort study of 970 335 individuals from 11 large databases, 12 dual combinations of

Observational Study (Estimation)

Development of Lung Cancer Survival Prediction Models Based on Real-world Data and Machine Learning

Jason C. Hsu; Phung-Anh Nguyen; Phan Thanh Phuc; Tsai-Chih Lo; Min-Huei Hsu; Chi-Tsun Cheng; Tzu-Hao Chang; Cheng-Yu Chen
Taipei Medical University, Taiwan


Jason C. Hsu


Phung-Anh Nguyen


Phan Thanh Phuc


Chi-Tsun Cheng

Abstract

Background
The development of disease risk and prognosis prediction models using machine learning or deep learning algorithms with big data is a major area of academic research based on AI in the medical field. Various researchers have used machine learning or deep learning algorithms to develop lung cancer risk and prognosis prediction models.1-6

Objectives
The purpose of this study was to use clinical real-world data with multiple attributes and multiple machine learning algorithms to establish a prediction model for the survival of lung cancer patients and to determine the key factors that affect overall survival.

Methods
This study used Taipei Medical University Clinical Research Database (TMUCRD) with data from 3 hospitals as the data source, the data were mapped to OHDSI OMOP CDM. We selected non-small-cell lung cancer patients from a retrospective development dataset of TMUCRD and Taiwan Cancer Registry between January 2008 and December 2018. All patients were monitored from the index date of cancer diagnosis until the event of death or the last visit to hospitals. Variables including demographics, comorbidities, medications, laboratories, and gene tests of patients were retrieved and used to develop the machine learning models. Nine machine learning algorithms with various modes (e.g., integrating different variables) were used to develop the predicted models. The performance of the algorithms was measured by the area under the receiver operating characteristic curve (AUC), accuracy, sensitivity (Recall), specificity, positive predictive value (Precision), and F1-score.

Results
In total, 3,714 patients were included (2,280 for the training dataset and 1,434 for the testing dataset). The artificial neural network (ANN) AUC values of different modes were observed with the highest score of 89%. The best performance of the ANN model was achieved when integrating all variables with the AUC, accuracy, precision, recall, and F1-score of 0.89, 0.82, 0.91, 0.75, and 0.65, respectively. The most important features were the cancer stage, cancer size, diagnosed age, smoking, drinking status, EGFR gene, and body mass index.

Conclusion
In this evaluation of lung cancer survival, the ANN model led to a better predictive performance with high AUC, precision, and recall when integrating different data types. Further research is necessary to determine the feasibility of applying the algorithm in the clinical setting and explore whether using this tool could improve care and outcomes. This study is expected to be developed into a multinational cooperative research using OHDSI tools and OMOP CDM in the future.

Methods

Study Design and Data Source
We conducted a retrospective study in which we obtained the data from the Taipei Medical University Clinical Research Database (TMUCRD), which were mapped to OHDSI OMOP CDM. The TMUCRD retrieved data from various electronic medical records (EHR) of three hospitals, Taipei Medical University Hospital (TMUH), Wan-Fang Hospital (WFH), and Shuang-Ho Hospital (SHH). The database contains the electronic medical record data of 3.8 million people accumulated from 1998 to 2020. This study has been approved by the Joint Institute Review Board of Taipei Medical University (TMU-JIRB), Taipei, Taiwan (approved number, N202101080).

Cohort Selection
This study selected patients with lung cancer (ICD-O-3 code: C33, C34) from 2008 to 2018 in the TCR database. Exclusion criteria included individuals with ages under 20, SCLC patients, and patients who did not have any medical history in the three hospitals (TMUH, WFH, SHH). These 3,714 patients were included in this study, including 960 patients from TMUH, 1,320 from WFH, and 1,434 from SHH.

Machine Learning Study (Prediction)

Outcome Measurement
We ascertained the study outcomes using TMUCRD EHR and vital status data from the Taiwan Death Registry (TDR). We used the diagnosis date of NSCLC as the index date, and the outcome of this study was death within two years following diagnosis. Data were censored at the date of death or loss to follow-up, insurance termination, or the study's end on December 31, 2018.

Methods

Feature Selection
Based on a literature review and consultation with clinicians, we selected features that may lead to the mortality of NSCLC patients to build prediction models. Those features consisted of: (1) Demographic information, (2) Cancer conditions, (3) Comorbidities, (4) Medications, (5) Laboratory tests, (6) Genomic tests.

Development of the Algorithms
This study established prediction models based on four modes and different algorithms.
(1) The primary mode (e.g., mode 1) included demographic information, cancer conditions, comorbidities, and medications.
(2) The second mode (mode 2) included the data of mode one and the laboratory tests.
(3) The third mode (mode 3) included the data of mode one and genomic tests.
(4) The fourth mode (mode 4) considered all the above features.
The study aims to predict the survival of lung cancer patients; therefore, the problem can be formulated as a classification model as it could occur in the same patients. We used those possible machine learning techniques such as logistic regression (LR), boosting machine (LGBM), gradient boosting machi random forest (RF), AdaBoost, support vector machin methods were briefly introduced as follows.

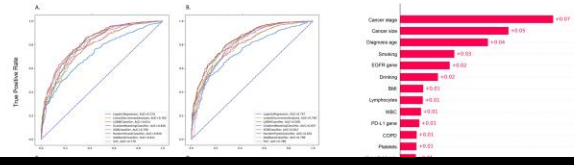
Evaluating the Algorithms
The training dataset contained the data of patients validation was applied in the training set to assess th and general errors. In words, patients in the traini repeatedly as the internal validation set. We recruite for generalizing the model.
The performance of the algorithms was measur characteristic curve (AUC), accuracy, sensitivity (Re Precision), negative predictive value (NPV), and F1-sc AUC by comparing various models based on the ex feature's contribution (i.e., features importance) of th exPlanations).

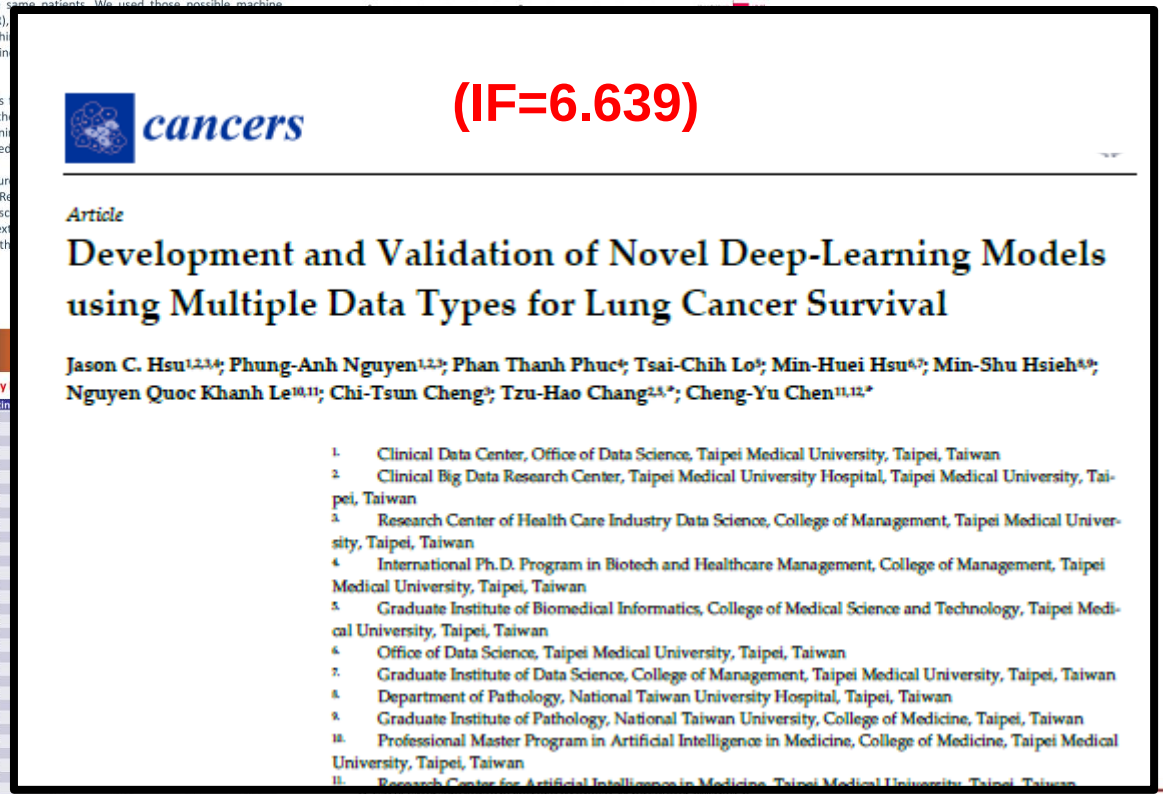
Results

Table 1. Performance of various Prediction Models by

Modes	Models	AUC Training	AUC Testin				
Mode 1	LR	0.70	0.72				
	LDA	0.78	0.78				
	LGBM	0.98	0.81				
	GBM	0.96	0.83				
	XGBoost	0.99	0.80				
	RF	0.90	0.82				
	AdaBoost	0.94	0.81				
	SVC	0.78	0.78				
	ANN*	0.89	0.88				
Mode 2	LR	0.74	0.75				
	LDA	0.81	0.79				
	LGBM	0.99	0.83				
	GBM	0.96	0.84				
	XGBoost	1.00	0.81				
	RF	0.92	0.83				
	AdaBoost	0.95	0.80				
	SVC	0.81	0.79				
	ANN*	0.89	0.89				
Mode 3	LR	0.70	0.73				
	LDA	0.80	0.81				
	LGBM	0.98	0.85				
	GBM	0.96	0.85				
	XGBoost	1.00	0.83				
	RF	0.91	0.84				
	AdaBoost	0.95	0.83				
	SVC	0.80	0.81				
	ANN*	0.89	0.89				
Mode 4	LR	0.74	0.75				
	LDA	0.83	0.82				
	LGBM	0.99	0.86				
	GBM	0.97	0.85				
	XGBoost	1.00	0.84				
	RF	0.93	0.85	0.75	0.93	0.73	0.82
	AdaBoost	0.96	0.83	0.76	0.92	0.75	0.83
	SVC	0.83	0.81	0.75	0.90	0.76	0.84
	ANN*	0.89	0.89	0.82	0.91	0.75	0.65

Results





cancers

(IF=6.639)

Article

Development and Validation of Novel Deep-Learning Models using Multiple Data Types for Lung Cancer Survival

Jason C. Hsu[1,2,3,4]; Phung-Anh Nguyen[1,2,3]; Phan Thanh Phuc[4]; Tsai-Chih Lo[5]; Min-Huei Hsu[6,7]; Min-Shu Hsieh[8,9]; Nguyen Quoc Khanh Le[10,11]; Chi-Tsun Cheng[3]; Tzu-Hao Chang[2,5,*]; Cheng-Yu Chen[11,12,*]

1. Clinical Data Center, Office of Data Science, Taipei Medical University, Taipei, Taiwan
2. Clinical Big Data Research Center, Taipei Medical University Hospital, Taipei Medical University, Taipei, Taiwan
3. Research Center of Health Care Industry Data Science, College of Management, Taipei Medical University, Taipei, Taiwan
4. International Ph.D. Program in Biotech and Healthcare Management, College of Management, Taipei Medical University, Taipei, Taiwan
5. Graduate Institute of Biomedical Informatics, College of Medical Science and Technology, Taipei Medical University, Taipei, Taiwan
6. Office of Data Science, Taipei Medical University, Taipei, Taiwan
7. Graduate Institute of Data Science, College of Management, Taipei Medical University, Taipei, Taiwan
8. Department of Pathology, National Taiwan University Hospital, Taipei, Taiwan
9. Graduate Institute of Pathology, National Taiwan University, College of Medicine, Taipei, Taiwan
10. Professional Master Program in Artificial Intelligence in Medicine, College of Medicine, Taipei Medical University, Taipei, Taiwan
11. Research Center for Artificial Intelligence in Medicine, Taipei Medical University, Taipei, Taiwan

14. Gunn SR. Support vector machines for classification and regression. *ISIS technical report*. 1998;14(1):5-16.
15. Agatonovic-Kustrin S, Beresford R. Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research. *Journal of pharmaceutical and biomedical analysis*. 2000;22(5):717-727.

Development of Machine Learning models for Cancer Survival among Lung cancer patients with Tyrosine Kinase Inhibitors (TKIs) treatment

Alex PA. Nguyen[1], Phuc T. Phan[2], Min-Huei Hsu[1], Jason C. Hsu[1,2*],
[1] Clinical Data Center, Office of Data Science, Taipei Medical University, Taipei, Taiwan;
[2] International PhD Program in Biotech and Healthcare Management, College of Management, Taipei Medical University, Taipei, Taiwan;



Background

Lung cancer is the most common cause of cancer death worldwide, including in Taiwan. Mutation in the EGFR gene is a driver in lung adenocarcinoma, as this gene is overexpressed in more than 50% of non-small cell lung cancer (NSCLC) in Asia. Most patients benefited from TKI therapies, but 5%-10% of patients did not achieve disease control when administered EGFR-TKIs and therefore acquired drug resistance within 10-12 months.

In this study, we aimed to develop prediction models for lung cancer survival among patients with TKI treatment using a larger number of samples, different data types, and various machine learning algorithms.

Methods

Study Design and Data Source

We conducted a retrospective study in which we obtained the data from the Taiwan Cancer Registry (TCR) database and the Taipei Medical University Clinical Research Database (TMUCRD).

Cohort Selection

Patients with lung cancer (ICD-O-3 code: C33, C34.1) from 2008 to 2018 in the TCR database. Exclusion criteria included individuals under 20, small cell lung cancer (SCLC) patients, and patients who did not receive lung cancer treatment in the three hospitals. Following that, only cancer patients who were undergoing TKIs (i.e., patients using EGFR-TKIs, ATC codes L01EB) were included in our study cohorts.

Outcome Measurement

The outcome of this study was death within two years following diagnosis. Data were censored at the date of death or loss to follow-up, insurance termination, or the study's end on December 31, 2020.

Feature Selection

The selected features were as follows: (1) Demographic information; (2) Cancer condition; (3) Comorbidities; (4) Current medications use; and (5) Laboratory test results. All the features were defined before the time patients were prescribed TKI drugs.

Developing the Machine Learning models

Six machine learning algorithms were used including Logistic Regression (LR), bootstrap aggregation (bagging), gradient boosting machine (GBM), AdaBoost, random forest (RF), and extreme gradient boosting (XGBoost)., to develop the prediction models.

The training set, containing the data of Taipei Medical University Hospital and Wang Fang Hospital. The testing set, including the data of Shuang Ho Hospital, was used to validate the models. The 5-fold cross-validate was applied.

ristic curve (AUC), accuracy, sensitivity, specificity, and F1-
the performance of all prediction models.

Machine Learning Study (Prediction)

Results

Table 1. Baseline demographic of cohort patients in the study

Feature	Training cohort (n=731)	Testing cohort (n=454)
Demographic		
Gender, No. (%)		
Female	327 (44.7%)	206 (45.4%)
Male	404 (55.3%)	248 (54.6%)
Age, Mean (SD), y	67.8 (13.2)	67.3 (12.7)
BMI		
Mean (SD)	23.4 (3.85)	23.2 (4.00)
Median [Min, Max]	23.1 [13.0, 61.3]	22.9 [13.2, 38.1]
Missing	238 (32.6%)	94 (20.7%)
Smoking, No. (%)		
No	356 (48.7%)	222 (48.9%)
Yes	156 (21.3%)	139 (30.6%)
Unknown	219 (30.0%)	93 (20.5%)
Drinking, No. (%)		
No	425 (58.1%)	316 (69.6%)
Yes	85 (11.6%)	45 (9.9%)

Feature	Training cohort (n=731)	Testing cohort (n=454)
Cancer Condition		
Tumor size, No. (%)		
T=<3cm	212 (29.0%)	103 (22.7%)
3=<T=<7cm	321 (43.9%)	210 (46.3%)
T>7cm	63 (8.6%)	48 (10.6%)
Missing	135 (18.5%)	93 (20.5%)
Cancer stage, No. (%)		
stage = 0	52 (7.1%)	28 (6.2%)
stage = 1	17 (2.3%)	9 (2.0%)
stage = 2	81 (11.1%)	33 (7.3%)
stage = 3	547 (74.8%)	368 (81.1%)
stage = 4	34 (4.7%)	16 (3.5%)
Unknown	52 (7.1%)	28 (6.2%)
Mortality, No. (%)	609 (83.3%)	368 (81.1%)

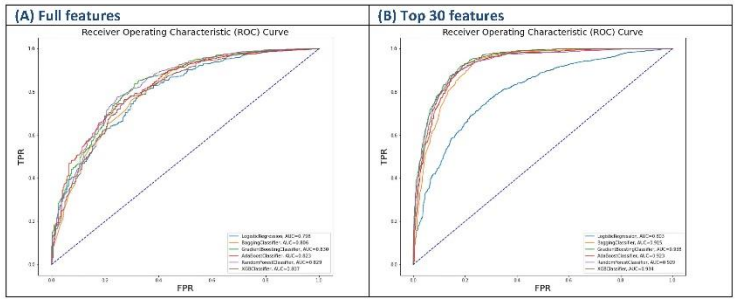


Figure 1. Receiver Operating Characteristic (ROC) Curve of various models

Conclusions

Random forest was observed as the best model when using all features. Moreover, while choosing the top 30 features, Gradient Boosting Classifier was found with the highest AUC of 0.94.

In summary, the model developed using the Gradient Boosting Classifier algorithm had the highest AUC regardless of the mode and was the most suitable tool for NSCLC survival prediction among patients who underwent TKI treatment. In addition, using more types of data (especially laboratory and genomic test results) led to better predictive performance. Cancer stage, cancer size, gender, diagnosis age, and body mass index were the essential features for NSCLC survival prediction.

Contact: **Jason C. Hsu**, International Ph.D. program in Biotech and Health Management, College of Management, Taipei Medical University, Taipei, Taiwan; 11F., No.172-1, Sec. 2, Keelung Rd., Daan Dist., Taipei City 106, Taiwan (R.O.C.); E-mail: jasohsu@tmu.edu.tw

Development of Breast Cancer Survival Prediction Models Based on Real-world Data and Machine Learning

Chun-Jung Wang[1], Min-Huei Hsu[2], Ruo-Kai Lin[3], Chin-Sheng Hung[4,5], Nei-Hui Kuo[6], Yu-Wen Cheng[1,7], Phung-Anh Nguyen[8], Phan Thanh Phuc[9], Chi-Tsun Cheng[10], Jason C. Hsu[11,12]*


Chun-Jung Wang


Jason C. Hsu


Min-Huei Hsu

1. School of Pharmacy, Taipei Medical University, Taipei, Taiwan.
2. Graduate Institute of Data Science, College of Management, Taipei Medical University, Taipei, Taiwan.
3. Graduate Institute of Pharmacognosy, College of Pharmacy, Taipei Medical University, Taipei, Taiwan.
4. Department of Surgery, School of Medicine, College of Medicine, Taipei Medical University, Taipei, Taiwan.
5. Division of General Surgery, Department of Surgery, Taipei Medical University Hospital, Taipei, Taiwan.
6. Cancer Center Division Head, Taipei Medical University Hospital, Taipei, Taiwan
7. Ph.D. Program in Drug Discovery and Development Industry, College of Pharmacy, Taipei Medical University, Taipei, Taiwan.
8. Clinical Data Center, Office of Data Science, Taipei Medical University, Taipei, Taiwan
9. International Ph.D. Program in Biotech and Healthcare Management, College of Management, Taipei Medical University, Taipei, Taiwan
10. Research Center of Health Care Industry Data Science, College of Management, Taipei Medical University, Taipei, Taiwan
11. International PhD Program in Biotech and Healthcare Management, College of Management, Taipei Medical University, Taipei, Taiwan.
12. Clinical Big Data Research Center, Taipei Medical University Hospital, Taipei Medical University, Taipei, Taiwan.

Background

Background

Breast cancer is the cancer with the highest incidence and mortality among women in most countries[1]. There are approximately 2.3 million newly diagnosed cases worldwide each year, and approximately 680,000 deaths annually. With the rise of artificial intelligence, in the past, many researchers used various clinical big data and machine learning algorithms to establish prediction models for breast cancer diagnosis[2] and prognosis[3], respectively, to assist medical decision-making and improve treatment outcomes. However, the parameters and accuracy of such prediction models may vary due to differences in race, geographic location, or other ethnic or individual factors, so it is necessary to use various data sources to develop various prediction models.

Objectives

This study aims to use clinical real-world data with multiple attributes and multiple machine learning algorithms to determine the key factors that affect overall survival when the patient be diagnosed with breast cancer and establish a prediction model which can act as a supporting decision aid for physician by modifying the magnitude of treatment.

Methods

Taipei Medical University Clinical Research Database (TMUCRD) was the data source of this study, which contains the electronic medical records of three hospitals in Taiwan, including Taipei Medical University Hospital (TMUH), Wan-Fang Hospital (WFH), and Shuang-Ho Hospital (SHH). All the data was mapped to OHDSI OMOP CDM. We selected breast cancer female patients whose ICD-O-3 code was C50.0-C50.9 from 2000 to 2019 as the study cohort, and non-primary breast cancer cases or cases with insufficient information on personal medical background and treatment were excluded. Neither do the patient whose follow-up period less than one year. Patients from TMUH and WFH were the training dataset, and patients from SHH were used for external testing. The percentage of alive and death is around 87% and 13% in training data, and 84% and 16% in external testing data, respectively. The date of diagnosis of breast cancer for each patient was used as the index date, and death within five years after diagnosis was used as the outcome. All the information could be gathered on the index data. Totally, there were 45 features involved, including the patient's basic demographic information, cancer condition, comorbidity, current medication, laboratory test result, were selected and used in the model generation. The comorbidities which occurred prior to breast cancer diagnosis were collected. And the lab values which recording within one year from diagnostic date were remained. Missing value of categorical data were classified as a new [illegible]th mean value. If the percentage of missing value more [illegible]ded features are BUN, CA153, CEA, creatine kinase, Ki67, [illegible]els or vascular invasion. Finally, there are 37 features be [illegible]logistic regression (LR), support vector machine (SVM), [illegible]om forest (RF), and artificial neural network (ANN) were applied to build prediction modules. Based on the external test results, the model with the largest area under the receiver operating characteristic curve (AUC) is the best model.

Machine Learning Study (Prediction)

Results

A total of 5,503 patients were included (4,071 for the training dataset and 1,432 for the testing dataset). Based on the external test results, neural network model had the highest AUC (0.880), following by GB (AUC=0.864), RF (AUC=0.863), SVM (AUC=0.804), LR (AUC=0.798) and DT (AUC=0.675). The accuracy of all models is above 85%. In addition, this study also found that, according to the results of the best model (NN), tumor clinical stage, clinical lymph node stage, primary site, Charlson-Deyo Comorbidity Index, and dementia played the most important role in predicting the five-year survival of breast cancer.

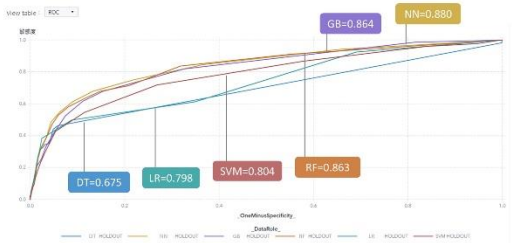


Figure1. ROC Curves of five years survival prediction model when diagnosed with breast cancer.

Model	AUC	Accuracy	Sensitivity	Specificity	Precision	F1-score
Logistic Regression	0.798	0.881	0.383	0.975	0.744	0.506
SVM	0.804	0.862	0.300	0.968	0.636	0.407
Decision Tree	0.675	0.865	0.348	0.963	0.637	0.450
Gradient Boosting	0.864	0.854	0.123	0.992	0.737	0.211
Random Forest	0.863	0.873	0.344	0.973	0.703	0.462
Neural Network	0.880	0.871	0.273	0.983	0.756	0.401

Table1. Performance of Survival Prediction Models.

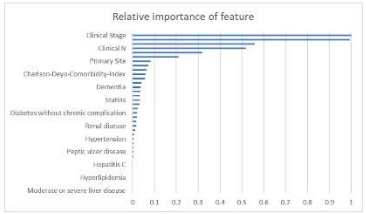


Figure2. Relative importance of feature in neural network model.

Conclusions

This study successfully established an accurate 5-year survival predictive model for breast cancer patients. Furthermore, this study also found many key factors that may affect the survival of breast cancer patients in Taiwanese patients. The results of the study can be used as a reference for clinical practice of breast treatment.

References

1. Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. 2021;71(3):209-249.
2. He T, Puppala M, Ezeana CF, et al. A Deep Learning–Based Decision Support Tool for Precision Risk Assessment of Breast Cancer. JCO Clin Cancer Inform. 2019(3):1-12.
3. Hernandez RK, Wade SW, Reich A, Pirolli M, Liede A, Lyman GH. Incidence of bone metastases in patients with solid tumors: analysis of oncology electronic medical records in the United States. BMC Cancer. 2018;18(1):44-44.

Machine Learning to Predict the Ischemic Stroke among Type 2 Diabetes Mellitus Patients using Taipei Medical University Clinical Research Database

PHAN THANH PHUC [1], PHUNG ANH NGUYEN [2], JASON C. HSU [1,2,*]
[1] International PhD Program in Biotech and Healthcare Management, College of Management, Taipei Medical University, Taipei, Taiwan;
[2] Clinical Data Center, Office of Data Science, Taipei Medical University, Taipei, Taiwan;



Background

Ischemic stroke has been recognized as a clinically important complication of type 2 diabetes (T2DM) patients. Risk prediction models for DM complications/comorbidities have substantial capacity to support the decision-making process regarding the patient's clinical management.
This study aims to develop machine learning algorithms to predict the risk of ischemic stroke among T2DM patients using various predictors such as patients' characteristics, disease history, laboratory tests, and medication.

Methods

1. Data source and study population
The dataset was collected from the Taipei Medical University Clinical Research Database (TMUCRD) in this study. Index 2008 data as wash-out-period—newly diagnosed T2DM patients from 2009 to 2019 as our cohort study with [ICD9-CM] codes 250.xx, and [ICD10-CM] codes E11 .xx.
2. Outcome
All patients were monitored from the date of taking antidiabetic drugs to the date the patients were admitted to hospitals with ischemic stroke (ICD9-CM codes 433, 434, 436, and ICD10-CM codes I60, I61, I62) during a one-year follow-up.
3. Features
The features were collected, including (i) patient characteristics (i.e., age, sex), (ii) comorbidities (i.e., any diagnoses before the date of taking antidiabetic drugs), (iii) other medication uses, and (iv) laboratory exams (i.e., Glucose, HbA1C, etc.)..
4. Statistical analysis and Model development
The training set, containing the data of Taipei Medical University Hospital and Wang Fang Hospital. The testing set, including the data of Shuang Ho Hospital, was used to validate the models. The stratified 10-fold cross-validate was applied in the training set to assess different machine learning models' performance and general errors.
Machine learning techniques, such as Logistic Regression (LR), Linear Discriminant Analysis (DT), Gradient Boosting Machine (GBM) and Random Forest (RF), to develop the prediction models. The performance of the algorithms was measured by Area Under the Curve (AUC), sensitivity, specificity, and F1-score.

Results

Table 1. Patient baseline and characteristic

Feature	Overall (n=4,697)	Training cohort (n=4,697)	Testing cohort (n=4,582)
	217 (2.34)	89 (1.9%)	128 (2.8%)
	61.8 (4)	60.7 (15.0)	62.9 (13.0)
	305 (46.4)	2,182 (46.5%)	2,123 (46.3%)
	974 (53.6)	2,515 (53.5%)	2,459 (53.7%)
	217 (2.34)	24.1 (0.34)	23.3 (0.35)
Laboratory test, Mean (SD)	6.3 (0.08)	6.03 (0.07)	6.7 (0.09)
Medication, Mean (SD)	25 (0.2)	23 (0.2)	27 (0.2)

Table 2. Model performance evaluation

Model	AUC (CV)	AUC (Testing)	Sensitivity	Specificity	F1-score
Logistic Regression	0.88	0.85	0.819	0.748	0.16
LDA	0.88	0.85	0.721	0.802	0.161
GBM	0.91	0.85	0.744	0.813	0.164
Random Forest	0.93	0.84	0.811	0.692	0.133

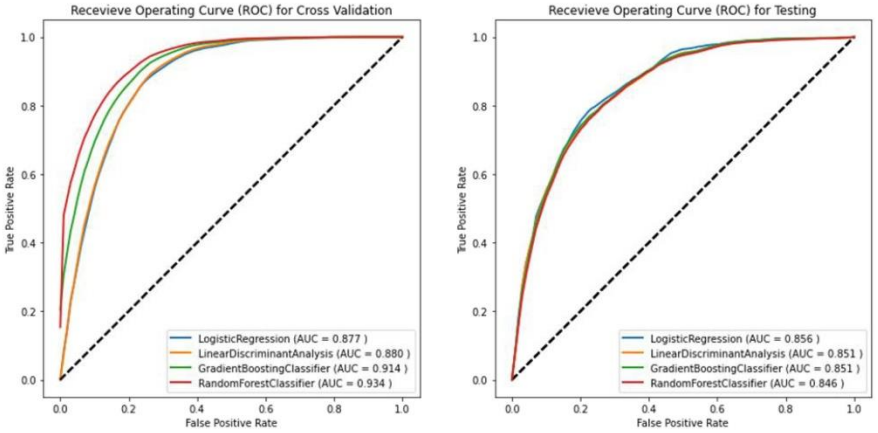


Figure 1. Receiver Operating Characteristic (ROC) Curve to evaluate the model performance

Conclusions

We successfully developed machine learning models to predict the risk of ischemic stroke among T2DM. Our model performance improved from Random Forest to Gradient Boosting Machine. The top three important features executed from our best model are antiplatelet agent, age, and prior stroke.

The strong association of diabetes with stroke has long been appreciated. To the best of our knowledge, there are limited studies in classifying and predicting ischemic stroke in the T2DM cohort by developing machine learning-based models. Therefore, our findings were essential to improve the accuracy of early detection, diagnosis, and prognosis of ischemic stroke to manage the risk of diabetes complications.

Machine Learning Study (Prediction)

Contact: * **Jason C. Hsu**, International Ph.D. program in Biotech and Health Management, College of Management, Taipei Medical University, Taipei, Taiwan; 11F., No.172-1, Sec. 2, Keelung Rd., Daan Dist., Taipei City 106, Taiwan (R.O.C.);
E-mail: *jasohsu@tmu.edu.tw*

Post-Stroke Prediction on Cognitive Impairment Development: A Machine Learning Approach

Muhammad Solihuddin Muhtar[1], Faizul Hasan[2], Alex P.A. Nguyen[1], Jason C. Hsu[1], Hsiao-Yean Chiu[2], Min-Huei Hsu[1]

[1]Graduate Institute of Data Science, College of Management, Taipei Medical University, Taipei, Taiwan
[2]School of Nursing, College of Nursing, Taipei Medical University, Taipei, Taiwan.

Background

Cognitive impairment following stroke has wide prevalence ranging from 25% to 81%. Further, stroke and the subtypes, including ischemic stroke, transient ischemic attack and intracerebral hemorrhage, significantly increase the long-term risk of dementia after 5 and 10 years. The incidence rate of post-stroke dementia increases yearly, though the relative risk gradually decreases. The study aims to predict the dementia development one year after stroke diagnose (index date).

Methods

The study conducted on TMUCRD from January 2004 to September 2017. The inclusion, exclusion and outcome criteria are selected based on ICD9 and ICD10 codes. We include all patient with history of stroke, insomnia, cognitive impairment and other codes related with the diseases (362.3, 433.x1, 434.x1, 436, 431.x, 430.x, 435.x, H34.1, I63.x, I64.x, I61.x, I60.x, G45.x). We exclude psychiatric disorder, sleep apnea, traumatic brain injury, cancer, Parkinson's disease, and cognitive impairment from the outcome (300.4, 296.2-296.3, 300, 293.84, 296.4-296.7, 295, 327.23, 800-804, 850.0, 850.1, 850.5, 850.9, 854.0, 959.01, 199.1, 332.0, F34.1, F32.9, F41. 9, F31, F31.x, F32, F32.x, F33, F33.x, F20. 9, G47. 33, S02.0, S02.1, S02.8, S02.91, S04.02, S04.03, S04.04, S06, S07.1, T74.4, S09.90, C80.1, G20). The outcome are mild cognitive impairment, Senile dementia, uncomplicated, Senile dementia with delusional or depressive features, Senile dementia with delirium, Dementia in conditions classified elsewhere, Alzheimer's disease, Frontotemporal dementia, and Senile degeneration of brain (331.83, 290.0, 290.1, 290.2, 290.3, 294.1, 331.0, 331.1, 331.2, G31.84, R41.89, R41.84, R41.83, R41.82, R41.81, F03.90, F03). The patient with outcome at least one year after stroke index date are labelled by 1, and the rest without outcome labelled by 0. We use pycaret library to compare the performance of many different machine learning algorithms.

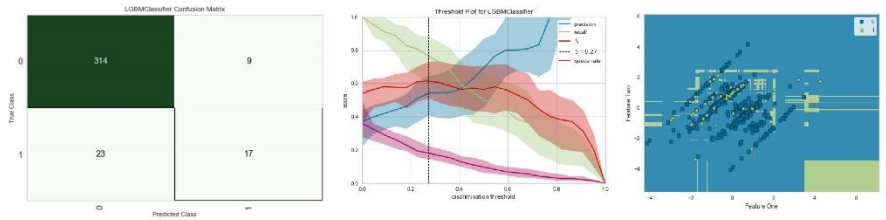
Results

oldout data (453 out of 4935 patients). LightGBM
he 10 fold cross validation training. The scores are
ecision, 0.32 for recall, and 0.21 for F1. In the current
the features.

Machine Learning Study (Prediction)

Results

We can see the performance in Figure 1. The true positive is still less than the false negative. This is to be expected since we use common features between two labels. The Figure 2 shows the optimal threshold is pretty low, 0.27, much lower than default 0.5 for binary classification. Figure 3 shows the distribution of the data based on labels. We can see that some positive labels are overlapping with the negative ones (have the same features).



Conclusions

The current model is able to determine whether a patient will develop cognitive impairment in the next year or not, though the probability is still very low, by using only gender, age and ICD code. Further features engineering will be conducted to improve the performance, such as adding medication or demographic features, especially to increase the true positive and to reduce the false negative numbers. Some hyperparameters may need to be adjusted to obtain better metrics, since the current model still uses the default pycaret parameters. We plan to run it on the CDM once the final result is reasonable.

References

1. del Ser, T., Barba, R., Morin, M. M., Domingo, J., Cemillan, C., Pondal, M., & Vivancos, J. (2005). Evolution of cognitive impairment after stroke and risk factors for delayed progression. Stroke, 36(12), 2670-2675

2. Li, C. H., Chang, Y. H., Chou, M. C., Chen, C. H., Ho, B. L., Hsieh, S. W., & Yang, Y. H. (2019). Factors of post-stroke dementia: A nationwide cohort study in Taiwan. Geriatr Gerontol Int, 19(8),815-822. doi:10.1111/ggi.13725

Contact: jasonhsu@tmu.edu.tw

Acknowledgement
We appreciate the technical assistance provided by the staff of Taipei Medical University's Office of Information Technology.

Analysis of Influencing Factors of Mortality in COVID-19 Patients: A Retrospective Cohort Study

Do Duy Khang; Phung-Anh Nguyen; Chang-I Chen; Chung-Chien Huang; Carlos Shu-Kei Lam; Noi Yar; Christine Y. Lu; Chi-Tsun Cheng; Jason C. Hsu*
Taipei Medical University, Taiwan

Abstract

Background
Coronavirus Disease (COVID-19) has spread rapidly around the world since the end of 2019. Because of its high incidence and high mortality, it is currently the most concerned health issue in the world. Clinically, avoiding mortality or severe illness is the main goal of Covid-19 treatment. Previous studies of factors influencing death of COVID-19 patients have shown that older age or certain comorbidities may increase the risk of severe illness in people with COVID-19, and some of these conditions may be fatal.1,2 In particular, cancer patients are particularly vulnerable to health consequences after infection, including increased risk of life-threatening infections and interruption of cancer or normal treatment.3 A comprehensive understanding of the factors affecting the mortality of Covid-19 cases and timely implementation of appropriate improvement strategies is one of the most important issues in clinical disease treatment.

Objectives
The purpose of this study was to explore the main influencing factors leading to Covid-19-related mortality and to provide clinical treatment recommendations based on the findings. This study used Taipei Medical University Clinical Research Database (TMUCRD) with data from 3 hospitals in Taiwan as the data source, the data were mapped to OHDSI OMOP CDM. It is expected to be developed into a multinational cooperative research using OHDSI tools and OMOP CDM in the future as well.

Methods
This study is a retrospective observational study. We obtained data from the TMUCRD, which collects three hospital electronic medical records in northern Taiwan. This study obtained 2021.01.01-2021.09.30 inpatients infected by Covid-19 from TMUCRD as the main study cohort. Patients who have not visited three hospitals in the past or who were younger than 20 years were excluded. The patient's first day of hospitalization was the index date, and the mortality was the main outcome. Covariates include demographic characteristics, health status, selected comorbidities and selected medications. Logistic regression with univariate and multivariate analysis method was used to estimate the association of each influencing factor with outcome. In addition, we also used the Cox regression model to conducted a further overall and stratified analysis about the associations between specific influencing factors and mortality among COVID patients.

Results
Totally 713 inpatient patients were included in this study. Uni-variable analysis showed that males, elderly, high CCI scores, co-morbidities such as congestive heart failure, cerebrovascular disease, dementia, chronic obstructive pulmonary disease, diabetes, renal disease, cancer, hypertension, hyperlipidemia, anemia, parkinson's disease, osteoporosis, etc., as well as the use of clonazepam sleeping pills, have a higher risk of mortality. However, after adjustment for other factors, only the following were statistically significant: older age and use of the clonazepam sleeping drug (OR= 4.358; 95% CI: 1.693-11.221; p-value=0.002). The results of the Cox regression regarding clonazepam use found that, overall, clonazepam sleeping pills significantly increase the mortality risk of COVID patients (HR=1.995; 95% CI: 1.007-3.954; p-value=0.048). Especially, when the patient's age was less than 65 years old, 0≤CCI<3, and no depression, the patients who used clonazepam sleeping pills had a higher risk of mortality than those who did not use clonazepam sleeping pills.

Methods

This study is a retrospective observational study. We obtained data from the Taipei Medical University Clinical Research Database (TMUCRD), which collects three hospital electronic medical records in northern Taiwan. This study obtained 2021.01.01-2021.09.30 inpatients infected by Covid-19 from TMUCRD as the main study cohort. Exclusion criteria included patients admitted to the ICU immediately after admission, deaths within 24 hours of admission, and cases under the age of 20. The patient's first day of hospitalization was the index date, and the mortality was the main outcome. Covariates include demographic characteristics, health status, selected comorbidities and selected medications. Logistic regression with univariate and multivariate analysis method was used to estimate the association of each influencing factor with outcome. In addition, we also used the Cox regression model to conducted a further overall and stratified analysis about the associations between clonazepam use and mortality among COVID patients by overall and stratified analysis.

Observational Study (Estimation)

Results

Table 1. Results of Uni-variable and Multi-variable Logistics Regression

Variables	Uni-variable Logistics Regression				Multi-variable Logistics Regression			
	OR	Lower 95% CI	Upper 95% CI	*P*-value	OR	Lower 95% CI	Upper 95% CI	*P*-value
Age					1.059	1.041	1.078	**<0.001**
Age <65								
65=<AGE<=85	1.959	1.161	3.305	**0.012**				
AGE >85	2.754	2.574	7.36	**<0.001**				
CCI score					1.177	0.858	1.614	0.313
CCI=0								
0=<CCI<3	5.605	3.271	9.603	**<0.001**				
CC>=3	11.22	6.06	20.804	**<0.001**				
Benzodiazepine derivatives (all)	1.436	0.774	2.2665	0.252	1.221	0.626	2.381	0.559
Benzodiazepine derivatives (clonazepam)	3.969	1.822	8.646	**0.001**	4.358	1.693	11.221	**0.002**
Benzodiazepine derivatives (N05BA)	1.409	0.617	3.219	0.416	0.954	0.352	2.588	0.926
Benzodiazepine derivatives (N05CD)	1.464	0.668	3.209	0.341	0.578	0.229	1.462	0.247

Table 2 Results of the associations between clonazepam use and mortality among COVID patients by overall and stratified analysis

Variables	N	Multi-variable COX Regression			
		HR	Lower 95% CI	Upper 95% CI	*p*-value
Overall Analysis	718	1.995	1.007	3.951	**0.048**
Stratified Analysis					
Age					
Age <65	300	11.340	2.179	59.005	**0.004**
CCI score					
0 =<CCI<3	238	7.171	1.218	42.216	**0.029**
Comorbidity					
Congestive heart failure (CHF)					
No	75	3.193	1.449	7.036	**0.004**
Diabetes mellitus (DM)					
No	147	3.203	1.355	7.573	**0.008**
Cancer					
Yes	4	11.267	1.264	100.443	**0.030**
Hypertension					
No	255	3.655	1.5	8.906	**0.004**
Hyperlipidemia					
No	163	3.584	1.469	8.742	**0.005**
Depression					
No	33	2.213	1.097	4.46	**0.026**
Parkinson's disease					
No	33	2.313	1.09	4.909	**0.029**
Osteoporosis					
No	47	2.196	1.088	4.434	**0.028**

Results

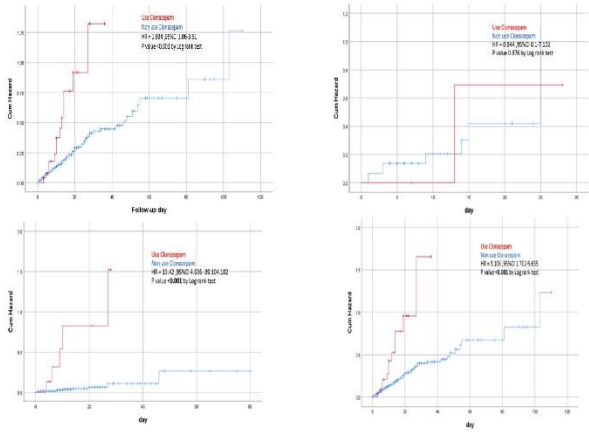


Figure 1. Kaplan-Meier curve for the mortality in COVID-19 patients between clonazepam users and non-users: (a) Overall population; (b) Age <65; (c) patients with depression; (d) patients without depression.

Conclusions

Our results suggest that male, elderly COVID-19 inpatients are at higher risk of mortality. Clonazepam sleeping pills users have significantly higher risk of mortality than non-users significantly. Patients younger than 65 years old, 0≤CCI<3, and without depression should avoid the use of clonazepam sleeping pills to reduce the risk of mortality.

References
1.Murthy S, Gomersall CD, Fowler RA. Care for critically ill patients with COVID-19. Jama. 2020;323(15):1499-1500.
2.Wu Z, McGoogan JM. Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: summary of a report of 72 314 cases from the Chinese Center for Disease Control and Prevention. Jama. 2020;323(13):1239-1242.
3.Desai A, Sachdeva S, Parekh T, Desai R. COVID-19 and cancer: lessons from a pooled meta-analysis. JCO global oncology. 2020;6.

Contact: jasonhsu@tmu.edu.tw

Future trends and prospects

Estimation + Prediction

(Data Science)

Technology

Research

Data

Team Work

EMR + Claims

Multicenter Data Collaboration
(Domestic/International)

臺北醫學大學
TAIPEI MEDICAL UNIVERSITY

THANKS

Jason C. Hsu

Chair, Local Host Committee of the 2022 OHDSI APAC Symposium
Associate Professor, Taipei Medical University, Taiwan
Director, Clinical Data Center, Office of Data Science, Taipei Medical University